

# **Data Replication README for Twenty Year Economic Impacts of Deworming**

March 11, 2021

These files produce the results presented in Hamory, Joan, Edward Miguel, Michael Walker, Michael Kremer, and Sarah Baird (2021) “Twenty Year Economic Impacts of Deworming”.

To replicate the results in the paper, unzip the folder but maintain the file structure within the zipped folder. There are five sub-folders, containing the data, do files, ado files, temp, and output presented in the paper and online appendices. Running the do files (described in detail below) in the "do" folder will utilize the datasets in the "data" folder, and produce the output in the "output" folder.

In what follows, we describe each file in turn, and provide any additional instructions on using that file.

## **DATA FOLDER:**

There are five datasets used in the analysis:

1. **Worms20\_Analysis.dta** is the main analysis dataset, and contains individual-round level outcome measures and controls for all 6,424 PSDP respondents of the KLPS Rounds 2-4 (excluding those in the Technical and Vocational Vouchers Program (VocEd) treatment group for KLPS-3 and KLPS-4 respondents, and excluding those in the Start-up Capital for Youth (SCY) treatment group for KLPS-4 respondents. This dataset is used to produce most results in the paper and appendices.
2. **Worms20\_Analysis\_scyvoked.dta** is the analysis dataset which contains individual-round level outcome measures and controls for 1,527 individuals in the VocEd or SCY treatment groups that were excluded in Worms20\_Analysis.dta. This dataset is used to produce Appendix Tables S5 and S6.
3. **Worms20\_Attrition.dta** contains information on the full sample of 7,527 PSDP individuals sought at any time during KLPS Rounds 2-4. This dataset is used in the analysis to analyze respondent attrition and tracking (Appendix Tables S1 and S2).
4. **Worms20\_Heterogeneity.dta** contains individual-round level outcome measures and controls for all of the outcomes presented in Appendix Tables S10 and S11.
5. **Worm\_Infection\_Panel.dta** contains individual-year level information on helminth infections collected for a subset of 2,695 PSDP participants in 1999 and 2001. For more information on this subsample, see the publicly available PSDP data files available on Harvard's Dataverse website. This dataset is used in the analysis to study changes in infection rates during the PSDP, in Appendix Tables S10 and S11.

## **CODE FILES:**

1. **Worms20\_MasterDo.do** is used to produce the tables, figures, and in-text results presented in Hamory et al. (2021). Before running this do file, you will need to update the global directory (line 13) for your primary directory. This do file calls upon Worms20\_Globals.do, Worms20\_master\_main.do, and Worms20\_master\_paper\_appendix.do described below.

- A. **Worms20\_Globals.do** is used to set conversion rates and set up controls of main cross-sectional and pooled regression specifications.
- B. **Worms20\_master\_main.do** is used to produce the main tables and figures presented in Hamory et al. (2021). It calls multiple files within the do/main/ folder. Note that Figure 1 needs to be run separately in R (Cost\_benefit\_pooled.R); see 2 below.
- C. **Worms20\_do\_appendix.do** is used to produce the tables, figures, and in text results presented in the online appendices of Hamory et al. (2021) (excluding Appendix Figure S1, S2, and S4 and Appendix Table S12). It calls multiple files within the do/appendix/ folder. Note that Appendix Figure S4 and Appendix Table S12 need to be run separately in R. Appendix Figure S1 is produced separately in a Word document and uses results from Appendix Table S1. Files used to produce Appendix Figure S2 are omitted due to data privacy purposes.

2. **main/Cost\_benefit\_pooled.R** is used to produce Figure 1 in Hamory et al. (2021). Before running this do file, you will need to update the directory (line 20) for your primary directory.

3. **appendix/Consumption\_graph.R** is used to produce the round-by-round consumption results of Figure S4 in Hamory et al. (2021). Before running this do file, you will need to update the directory (line 22) for your primary directory. In addition, before running this do file, you need to run Consumption\_gender\_results\_for\_R.do (called up in Worms20\_do\_appendix.do) which produces the estimates displayed in the graph.

4. **appendix/Earnings\_graph.R** is used to produce the round-by-round earnings results of Figure S4 in Hamory et al. (2021). Before running this do file, you will need to update the directory (line 22) for your primary directory. In addition, before running this do file, you need to run Earnings\_gender\_results\_for\_R.do (called up in Worms20\_do\_appendix.do) which produces the estimates displayed in the graph.

5. **appendix/Cost\_benefit\_table\_pooled.Rmd** is used to produce Appendix Table S12 using R-Markdown.

Stata and R code files have been run using Stata Version 15.1 and R Version 3.6.1.

## **ADO FILES:**

1. **fdr\_sharpened\_qvalues.do** is used to produce the adjusted q-values presented in Table 1 and Appendix Tables S3 and S5. This file is taken from Michael Anderson's website at [http://are.berkeley.edu/~mlanderson/ARE\\_Website/Research.html](http://are.berkeley.edu/~mlanderson/ARE_Website/Research.html). Note that this ado file is called upon in the do files used to produce those tables.

2. We make use of additional ado files from Stata's SSC, which can be installed from within Stata. These include: keeporder, estout, lincom, lincomest.