README

Kowalski, Amanda. "Behavior within a Clinical Trial and Implications for Mammography Guidelines," Review of Economic Studies, forthcoming.

Data Availability Statement

This paper uses confidential, deidentified data from the Canadian National Breast Screening Study (CNBSS) obtained through an individual data sharing agreement supported by the National Bureau of Economic Research and the University of Toronto, the host institution for the CNBSS data. Per the terms of this agreement, I am not permitted to share these data. Interested researchers may submit their own data request application to Dr. Anthony Miller (ab.miller@utoronto.ca), the CNBSS PI, at the University of Toronto. My analysis code is available in the Zenodo digital repository: https://doi.org/10.5281/zenodo.6210284.

Dataset

- The confidential data from the CNBSS are not available for sharing. At the individual level, these data include information about randomization status, demographics, and outcomes at multiple follow-up lengths.

Software Requirements

- Stata (code was last run with version 14)
- SAS (code was last run with version 9.4)
- LaTeX

Description of Programs

- Master Program: The "master.do" program defines the file path of the project folder in the beginning of the file. It also describes and executes the programs used to extract, clean, and analyze the datasets, and to generate the figures and tables presented in the paper. It is suggested to run the programs in the order listed below as some programs use outputs from the preceding programs as input. The code requires approximately 24 hours to run. Run time can be reduced by reducing the number of bootstrap replications, but this will affect standard error estimates. To run the project, follow the steps in the "Instructions to Replicators" section below after receiving the data.
- Programs Used in Data Preparation: These programs will extract and reformat the CNBSS datasets used in the analyses. The master.do file will run them all in the following order:
    - data_prep_SAS.do: This file calls the following SAS programs in order from the "SAS/Programs" directory:
        a. "0.1 Importing CPORT files.sas"
        b. "1.0 Assigning formats - N05.sas"
        c. "1.0 Assigning formats - N08.sas"
        d. "1.0 Assigning formats - N09.sas"
        e. "1.0 Assigning formats - N10.sas"

   f. "1.0 Assigning formats - N11.sas"
   g. "1.0 Assigning formats - N12.sas"
   h. "1.0 Assigning formats - N14.sas"
   i. "1.0 Assigning formats - N15.sas"
   j. "1.0 Assigning formats - N18.sas"
   k. "1.0 Assigning formats - Other.sas"
- import_datasets.do
- patients_information.do
- identification_cancers.do
- patients_year_information.do
- basic_cleaning_covariates.do
- define_covariates_miller_definition.do
- create_endogenous.do
- cnbss_data_setup.do

- **Programs Used in Data Analysis:** These programs include the bootstrap and regression codes used in the paper. The master.do file will run them all after the data preparation programs:
  - linmte_no_covars_treat_eff_1000boot.do
  - linmte_no_covars_treat_eff_by_years_1000boot.do
  - bootstrap_all_covar_1000boot.do
  - mastectomy_bootstrap_1000boot.do
  - bootstrap_invasive_tumor_year_brca20_1000boot.do
  - treat_eff_bsample_compute_ci_pval_1000boot.do
  - treat_eff_by_years_bsample_compute_ci_pval_1000boot.do

- **Programs Used to Generate Exhibits:** These programs write the LaTeX files that are used to generate the figures and tables in the paper. The section named "List of Exhibits and Programs" matches each exhibit in the paper to one of the following do files.
  - write_udline.do
  - write_brcapaperderive.do
  - write_mortalityfigpaper.do
  - write_ybrca20figUOtestpaper_pvals_1000boot.do
  - write_sumstatsgrouped_1000boot.do
  - write_ybrca20boundfigpaper_pvals_1000boot.do
  - write_yr_2_5_formatted_1000boot.do
  - write_tableRB_pvals_4cols_1000boot.do
  - write_mortalityboundfigpaper_pvals_1000boot.do
  - write_tableovertimeshort_ybrca20_pvals_1000boot.do

- **Programs Used to Generate In-Text Numbers:** These programs generate numbers which appear in the paper but not any figures or tables. The section named "In-Text Numbers" matches each number to the relevant program and output files.
  - write_base_health_compare.do
  - create_outcome_comparison_data.do
  - asymptotic_did_equivalence_tehtest_bmw_pvals.do

- Custom Functions: These programs are called in several bootstrap .do files. They are called in the first few lines of master.do (the "SET UP" section).
  - der_MTO.ado
  - der_MUO.ado
  - treatment_effects.ado
  - bootsrap_statistics.ado

Outside Packages:

This code uses the following public Stata packages. They are also installed in the "SET UP" section of master.do, but they can be manually installed with "ssc install <package name>":

- confirmdir
- personage

Confidential CNBSS Materials:

I received the following confidential materials from the CNBSS team:

- Raw data files in SAS:
  - ybmi.sas7bdat
  - ydeaths.sas7bdat
  - yelig_alloc.sas7bdat
  - ysingle.sas7bdat
  - yfx_n05.sas7bdat
  - yfx_n08.sas7bdat
  - yfx_n09.sas7bdat
  - yfx_n10.sas7bdat
  - yfx_n11.sas7bdat
  - yfx_n12.sas7bdat
  - yfx_n14.sas7bdat
  - yfx_n15.sas7bdat
  - yfx_n18.sas7bdat
- SAS format files:
  - cat05.sas7bcat
  - cat08.sas7bcat
  - cat09.sas7bcat
  - cat10.sas7bcat
  - cat11.sas7bcat
  - cat12.sas7bcat
  - cat14.sas7bcat
  - cat15.sas7bcat
  - cat18.sas7bcat
- Codebook with documentation on surveys and data:
  - COMBINED.pdf

Instructions to Replicators (for those who receive data from the CNBSS):

Future researchers who receive the above materials can replicate my results with the following steps:

- Download the replication files with the existing directory structure.
- Place all raw data files in "SAS/Data".
- Using a Windows version of SAS, save all SAS format files as transport files (.cport).
- Place all transport files in "SAS/Data/CPORT_FILES"
- Run master.do.

List of Exhibits and Programs

Figure/ Table	Program File (.do)	Output File(s) --- Corresponding Tag in the Manuscript (.tex/.pdf)
Figure 1	write_udline	udline
Figure 2	write_brcapaperderive	brcapaperderive
Figure 3	write_mortalityfigpaper	mortalityfigpaperbrca20_1000boot
Figure 4	write_ybrca20figUOtestpaper_pvals_1000boot	ybrca20figUOtestpaper_pvals_1000boot
Table 1	write_sumstatsgrouped_1000boot	sumstatsmainspec40s
Figure 5	write_ybrca20boundfigpaper_pvals_1000boot	ybrca20boundfigpaper_pvals_1000boot
Table 2	write_yr_2_5_formatted_1000boot	yr_2_5_formatted_1000boot
Table 3	write_tableRB_pvals_4cols_1000boot	tableRB_pvals_4cols_1000boot
Figure 6	write_mortalityboundfigpaper_pvals_1000boot	mortalityboundfigpaper_pvals_1000boot
Table 4	write_tableovertimeshort_ybrca20_pvals_1000boot	tableRBovertimeshort_ybrca20_pvals_1000boot

In-Text Numbers

The following sentence contains a **bolded** in-text number created by the file *asymptotic_did_equivalence_tehtest_bmw_pvals.do* and can be found in its output *DID_treat_eff_D_yany_asym_exclrisk_Y_brca_20_40s.dta*: *"Accordingly, the test rejects the null hypothesis of treatment effect homogeneity at the **4.3%** level under the Brinch et al. (2017) approach and at the 2.3% level under my proposed approach."*

The following sentences contain **bolded** in-text numbers created by the file *create_outcome_comparison_data.do* and can be found in its output *avg_outcome_comparison.tex:* *"Relative to excluded women, women in the main analysis sample have lower breast cancer incidence: **442 versus 651 per 10,000**. They also have lower mortality: **456 versus 551 per 10,000**."*

The following sentence contains **bolded** in-text numbers created by the file *write_base_health_compare.do* and can be found in its output *base_health_compare.tex*: *"They are less likely to have had any mammograms prior to enrollment: **15% versus 35%**; and they are less likely to have practiced breast self-examination: **45% versus 54%**."*

All other in-text numbers are based on statistics reported in the tables and figures.

References

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond LATE with a discrete instrument. Journal of Political Economy 125 (4), 985—1039.

CNBSS (2015). Canadian National Breast Screening Study: Cumulative Data 1980-2005. University of Toronto.