READ ME file for the replication package to:

"The Causal Effect of Scaling up Access to Psychotherapy"

By Benjamin Ly Serena, email: blse@rff.dk, last edited: 11/04/2025.

This file provides a guide to replicating the data and results of the paper, *"The Causal Effect of Scaling up Access to Psychotherapy".* Since the data used in the analysis is confidential, I cannot share the data set. However, below I describe how to obtain access to the data and provide all the code necessary to reproduce the analysis.

Data Access

The paper uses a combination of several Danish administrative registers, which are merged using anonymized personal identifiers assigned at birth.

The use of this data is subject to the European Union's General Data Protection Regulation (GDPR). The data are stored on secure servers at Statistics Denmark and are not allowed to be transferred to computers outside Statistics Denmark. Researchers seeking access to the register data used in this paper must submit a written application for approval by Statistics Denmark. In the case of the prescription drug register, approval from the Danish Health Data Authority is also required. Access may be granted to researchers affiliated with an accredited Danish research institution, or to international researchers collaborating with accredited Danish institutions. Once approval has been obtained, the data can be accessed remotely through Statistics Denmark's secure server. The complete list of registers required and who to contact to get access to them is provided in Table 1. For more information see: https://www.dst.dk/en/TilSalg/Forskningsservice.

Software

All data generation and analysis was either performed using SAS (version 9.4 for Windows), Stata (version 16.0 for Windows (64-bit)), or RStudio (version 2023.03.2 for Windows).

Replication Procedure

The following lists all steps (to be executed in the specified order) from data collection to analysis needed to reproduce the results in the main paper and appendix.

A. Building blocks used in generation of main data

1. SAS file "PriceSuiFinal.sas" to estimate costs associated with suicide attempts.
2. SAS file "CollectData_HealthCareUse_OddsRatio_IdentifyRelevantDiseases.sas" to identify somatic diseases related to mental illness.
3. SAS file "CollectData_HealthCareUse_Prices.sas" to estimate costs associated with somatic health care services.
4. SAS file "ChildhospCutoff.sas" to collect data used to estimate municipality-specific age cutoffs in psychiatric hospitals (used in RD-DID model).
5. Stata file "IdentChildhospCutoff.dta" to identify municipality-specific age cutoffs in psychiatric hospitals (used in RD-DID model).

B. Build main data

1. SAS file "CollectData_Psyc_AllAges.sas" to collect main data.
2. Stata file "Psyc_HighFreq_data_AllAges.do" to create monthly data set.
3. Stata file "Psyc_HighFreq_data_AllAges_YearCollapse.do" to create yearly data set.
4. SAS file "CollectData_HealthCareUse.sas" to collect data on physical health outcomes.
5. SAS file "CollectData_LaborMarket.sas" to collect monthly labor market outcomes.
6. Stata file "ImputeEmp.sas" to collect data to predict monthly employment before 2008.
7. R file "ImputeEmpPred.R" to predict monthly employment before 2008.
8. Stata file "PsykTestLongRunRD.do" to create main event study data set.

C. Build auxiliary data sets

Severity prediction used in Table 1

1. Stata file "MLdataSev_Pre_Sample.do" to create samples for prediction.
2. SAS file "CollectData_MLData_Sev.sas" to collect additional data for prediction.
3. Stata file "MLdataSev_Pre.do" to calculate pre-shock covariates already in data.
4. Stata file "MLdataSev_Pre_MergeDat.do" to merge prediction covariates.
5. R file "ML_R_SevPrediction.R" to run machine learning prediction.

Event-time-specific outcome predictions used in Appendix Table A.19

1. Stata file "MLOutPred_Data.do" to collect prediction data.
2. R file "MLOutPred.R" to run machine learning prediction.

Other data

- Stata file "CollapseLabor.do" to calculate yearly labor market transfers used in Table 1.
- SAS file "CollectData_Inc.sas" to collect socioeconomic status used for heterogeneity and yearly labor market analyses.
- Stata file "TherapyLongRunRD.do" to create event data set based on PCP mental health consultations rather than psychometric tests, used in Appendix Table A.20.

D. Analysis

1. Stata file "Final_FiguresAndTables.do" to run main paper results.
2. Stata file "Final_FiguresAndTables_Appendix.do" to run appendix results.

Table 1: Overview of administrative registers and relevant contacts

Register	Description	Years	Contact
BEF	Population	1995-2019	Statistics Denmark
DODSAASG & DODSAARS	Cause of death	1970-2019	Statistics Denmark
LMDB	Danish National Prescription Register	1995-2019	Danish Health Data Authority
LPR	National Patient Register	1995-2019	Statistics Denmark
LPSYPOP2014 & LPSYBES2014	Psychiatric Hospital Register	1995-2014	Statistics Denmark
SSSY & SYSI	The National Health Insurance Register	1995-2019	Statistics Denmark
UDDA	Educations	1980-2019	Statistics Denmark
IND	Income	1980-2019	Statistics Denmark
SGDP	Sickness benefits	1980-2019	Statistics Denmark
DREAM	Labor market attachment	1999-2019	Statistics Denmark
LON	Wage register	1999-2019	Statistics Denmark
IDAP	Occupations	1999-2019	Statistics Denmark