# README for “The Returns to Nursing: Evidence from a Parental-Leave Program”

Authors: Benjamin Friedrich, [benjamin.friedrich@kellogg.northwestern.edu](mailto:benjamin.friedrich@kellogg.northwestern.edu)

Martin Hackmann, [mbhackmann@gmail.com](mailto:mbhackmann@gmail.com)

## Data Availability Statement

All main results in this paper use administrative microdata from Denmark, maintained by **Statistics Denmark** on a secure server.

According to the Act on Processing of Personal Data, (Act No. 429 of 31 May 2000 with amendments, which can be found at <http://www.datatilsynet.dk/english>), working with individual level data is required to guarantee confidentiality and preserve full anonymity for the data subjects in every respect. For this reason, individual level data from Statistics Denmark is not delivered to any external firm, institution, or person. Instead, datasets and linkages between datasets constructed at Statistics Denmark are stored at Statistics Denmark.

However, researchers employed at specific authorized environments can establish remote online access to these datasets stored at Statistics Denmark. Although researcher may get access to rather detailed individual level data, they are only allowed to publish statistical analyses and results at an aggregate level where no single person or enterprise may be identified. For security reasons, only researchers employed at authorized research institutions can get access to individual level data at Statistics Denmark. And only permanent research institutions with a responsible leader and several researchers can be authorized. Statistics Denmark does not authorize foreign research institutions, but foreign researchers may get access to relevant data if they are affiliated to an authorized Danish research institution. In this case, foreign researchers will get remote access to data at Statistics Denmark, directly from their own workplaces wherever they work or live.

Specifically, researchers need to apply to get access to these confidential data. Statistics Denmark provides guidelines for data access here: <https://www.dst.dk/en/TilSalg/Forskningsservice>

When applying for data access, you must request the following datasets in your proposal:

**Relevant Data Registers Maintained by Statistics Denmark:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Register** | **Description** | **Link (last accessed October 12th, 2020)** | **References:** |
| **Integrated Database for Labor Market Research (IDA)** | **Employment data and Firm information** |  | **See Timmermans (2017) for a description and Hummels et al. (2014) for an application** |
| IDAN | Employment spells with pay and person-establishment link | https://www.dst.dk/extranet/ForskningVariabellister/IDAN%20-%20IDA%20ans%C3%A6ttelser.html |  |
| IDAS | Establishment ID, workplace characteristics | https://www.dst.dk/extranet/ForskningVariabellister/IDAS%20-%20IDA%20arbejdssteder.html |  |
| IDAP | Personal work characteristics, labor market experience | https://www.dst.dk/extranet/ForskningVariabellister/IDAP%20-%20IDA%20persondata.html |  |
| IND | Personal income data | https://www.dst.dk/extranet/ForskningVariabellister/IND%20-%20Indkomst.html |  |
|  |  |  |  |
| **National Population Registers** | **Population data** |  | **See Helgertz et al. (2014) and Østergaard et al. (2016) for applications.** |
| FAFA | Family ID, number and age of children | https://www.dst.dk/extranet/ForskningVariabellister/FAFA%20-%20C-Familier.html |  |
| BEF | Person ID, Family ID, household characteristics | https://www.dst.dk/extranet/ForskningVariabellister/BEF%20-%20Befolkningen.html |  |
| FAIN | Person ID, Family ID, household characteristics, residential address | https://www.dst.dk/extranet/ForskningVariabellister/FAIN%20-%20Husstande%20og%20familier.html |  |
|  |  |  |  |
| **Social Statistics Register** | **Benefit data** |  | **See Druedahl et al. (2019) for an application.** |
| SHSS | Receipt and duration of social benefits | https://www.dst.dk/extranet/ForskningVariabellister/SHSS%20-%20Sammenh%C3%A6ngende%20socialstatistik.html |  |
|  |  |  |  |
| **Education Registers** | **Education data** |  | **See Jensen et al. (2011) for a description and Østergaard et al. (2016) for an application.** |
| INST | Educational institutions | https://www.dst.dk/extranet/ForskningVariabellister/INST%20-%20Institutionsregistret.html |  |
| UDDA | Personal education data: enrollment and completion | https://www.dst.dk/extranet/ForskningVariabellister/UDDA%20-%20Uddannelser%20(BUE).html |  |
|  |  |  |  |
| **Cause of Death Register** | **Mortality data** |  | **See Helweg-Larsen (2011) for a description and Nielsen et al. (2012) for an application.** |
| DODSAARS | Until 2001 | https://www.dst.dk/extranet/ForskningVariabellister/DODSAARS%20-%20D%C3%B8ds%C3%A5rssagsregistret.html |  |
| DODSAASG | 2002-2018 | https://www.dst.dk/extranet/ForskningVariabellister/DODSAASG%20-%20D%C3%B8ds%C3%A5rsagsregister.html |  |
|  |  |  |  |
| **National Patient registry (NPR)** | **Hospitalization data** |  | **See Schmidt et al. (2015) for a description and Nielsen et al. (2017) for an application** |
| LPRADM | Hospital discharge records | https://www.dst.dk/extranet/ForskningVariabellister/LPR\_ADM%20-%20Landspatientregistret%20-%20administrative%20oplysninger.html |  |

The only supplementary data that we have collected outside of Statistics Denmark is information from wage bargaining agreements in Denmark. We have received scanned copies of the wage bargaining agreements for nurses 1991-1999 from the Danish Nurses’ Council (Dansk Sygeplejerad, [www.dsr.dk](http://www.dsr.dk)), and these pdf documents are included with this replication package in the data folder. From these scanned wage agreements, we have hand-collected contractual wages for nurses in the public sector to create the graph on wage bargaining.

Data dictionary for wage bargaining data (nurse\_wage\_tables\_short.csv):

|  |  |
| --- | --- |
| **Variable** | **Description** |
| time | Date when the collective bargaining agreement takes effect |
| salary\_group | pay grade in the compensation system |
| c\_ii, c\_iii, c\_iv, c\_v, c\_vi | Hourly wages applicable in municipalities belonging to areas II, III, IV, V, VI, respectively |
|  |  |

**Data References:**

Druedahl, Jeppe, Mette Ejrnæs, and Thomas H. Jørgensen. "Earmarked paternity leave and the relative income within couples." *Economics Letters* 180 (2019): 85-88.

Helgertz, Jonas, Pieter Bevelander, and Anna Tegunimataka. "Naturalization and earnings: a Denmark–Sweden comparison." *European Journal of Population* 30, no. 3 (2014): 337-359.

Helweg-Larsen, Karin. "The Danish register of causes of death." *Scandinavian journal of public health* 39, no. 7\_suppl (2011): 26-29.

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang. "The wage effects of offshoring: Evidence from Danish matched worker-firm data." *American Economic Review* 104, no. 6 (2014): 1597-1629.

Jensen, Vibeke M., and Astrid W. Rasmussen. "Danish education registers." *Scandinavian journal of public health* 39, no. 7\_suppl (2011): 91-94.

Nielsen, Sune F., Børge G. Nordestgaard, and Stig E. Bojesen. "Statin use and reduced cancer-related mortality." *New England Journal of Medicine* 367, no. 19 (2012): 1792-1802.

Nielsen, Peter Brønnum, Flemming Skjøth, Mette Søgaard, Jette Nordstrøm Kjældgaard, Gregory YH Lip, and Torben Bjerregaard Larsen. "Effectiveness and safety of reduced dose non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study." *Bmj* 356 (2017).

Østergaard, Søren D., Janne T. Larsen, Søren Dalsgaard, Timothy E. Wilens, Preben B. Mortensen, Esben Agerbo, Ole Mors, and Liselotte Petersen. "Predicting ADHD by assessment of Rutter’s indicators of adversity in infancy." *PloS one* 11, no. 6 (2016): e0157352.

Schmidt, Morten, Sigrun Alba Johannesdottir Schmidt, Jakob Lynge Sandegaard, Vera Ehrenstein, Lars Pedersen, and Henrik Toft Sørensen. "The Danish National Patient Registry: a review of content, data quality, and research potential." *Clinical epidemiology* 7 (2015): 449.

Timmermans, Bram. "The Danish integrated database for labor market research: towards demystification for the English speaking audience." *Aalborg: Aalborg University* (2010). <http://webdoc.sub.gwdg.de/ebook/serien/lm/DRUIDwp/10-16.pdf>, last accessed October 12th, 2020.

## Computational requirements

The raw data are loaded and prepared in SAS. The analysis has been conducted in Stata/SE 15.1 (Windows).

Part of the Stata code uses the boottest procedure. This module should be installed from within Stata by typing "ssc install boottest".

### Memory and Runtime Requirements

The replication package was last tested in October 2020. Running the entire Stata code sequentially took about 3 days on a 4-core Intel server (Windows Server 2012 R2 Standard) with 512 GB RAM.

### File Directory

Code, data, and output results are saved in the following folders

* SAS code is located in “programs/sas”
* Stata code is located in “programs/stata”
* Stata data sets and the wage bargaining data are located in “data”
* Output files are located in “results”

### Description of programs

* Programs in programs/sas will conduct data preparation and cleaning of raw data files.
  + The file dataprep.sas will extract and prepare the raw datasets, loading pre-merged files by ECONAU at Aarhus University. Register PERSONER merges BEF and FAIN; register IDAPALL contains relevant merged sources BEF, FAIN, IDAN, IDAP, IND, UDDA; register FAMILIE contains relevant merged sources FAFA, BEF, FAIN.
* Programs in programs/stata will conduct data preparation and further cleaning of the stata files prepared in the previous step and conduct the main analysis. The relevant stata files are aggregated in programs/stata/Master\_dataprepandanalysis.do, which first defines the directories. Next it executes three main files, which we describe in further detail below, in the following order:
  + The file programs/stata/Master\_dataprep.do will run all Stata files in the correct order to construct the main analysis files.
  + The program programs/stata/Master\_labormarket.do will run all results related to the labor market response to the reform.
  + The program programs/stata/Master\_healthoutcomes.do will run all results related to quality of care outcomes as a consequence of the reform.

### Details

The following list provides more details on the file structure to construct the main analysis files. These codes need to be run in order because saved data files are inputs for subsequent data steps.

The data preparation master file, Master\_dataprep.do, executes the following files in the respective order.

Hospital data prep:

Prepare\_hospitalsample\_PartA.do

This code prepares hospital discharge, population, and mortality data for the empirical analysis of the quality of hospital care.

The file loads population data from demographics\_long.dta and deathregister.dta and stores the combined information under demo\_death.dta. The second part loads admissions\_10\_2019\_Stata.dta and the data with demo\_death.dta. Diagnosis codes are mapped into acute conditions together with diagnoses\_to\_acute.do, which is executed in Prepare\_hospitalsample\_PartA.do. The information is aggregated to the county-year level and saved in hosp\_regs\_countyyear\_extra.dta, which contains the key outcome measures for AMI patients.

Prepare\_ county\_controls.do

This code prepares key control variables for the main regression analysis.

The file loads admissions\_10\_2019\_Stata.dta and demo\_death.dta, which contains relevant population data, see Prepare\_hospitalsample\_PartA.do. Diagnosis codes are mapped into acute conditions together with diagnoses\_to\_acute.do, which is executed in Construct\_ county\_controls. The information is aggregated to the county-year level and saved in county\_demographics.dta. This file contains population counts and patient demographics.

Nursing home prep:

Prepare\_elderly\_data.do

This file merges residential addresses (addresses.dta) with the death register to find locations where individuals die in a nursing home contemporaneously. We check if children also live at these addresses and only use those locations with no children.

The final data set is saved under elderly\_data.dta, which contains demographics for age 60 plus including an indicator for nursing home residency and death register match.

Prepare\_mortalitysample\_PartA.do

This file prepares elderly mortality data.

The file loads admissions\_10\_2019\_Stata.dta, demographics\_long.dta, deathregister.dta and saves the combined data set under mortality\_sample\_long.dta.

Prepare\_mortalitysample\_PartB.do

This file continues the preparation of elderly mortality data and starts by loading mortality\_sample\_long.dta. The cause of death codes are mapped into relevant disease categories using causes\_of\_death\_nursinghome\_who.do. Nursing home and overall unconditional morality outcomes are risk-adjusted and saved under mortality\_residuals\_pre\_long.dta. A subset of variables that summarize the person’s hospital history are saved separately under nhsample\_prehosps.dta.

Labor market prep:

Prepare\_microdata\_healthworkers\_by\_industry.do

This file creates a micro panel of health care professionals with information about industry, experience and pay.

This file starts from the raw IDAPALL matched employer-employee micro data and defines health care workers based on education (doctors, nurses, nursing assistants). We then add data from IDAS (establishment panel) to define health care providers based on industry information of establishments. We conduct data cleaning to correct classification changes and to reduce measurement error and we save the combined panel as main\_nursing\_timeseries\_extended.dta

Prepare\_estimates\_individual\_takeup.do

This file estimates individual takeup to the leave reform among health care workers.

The file first uses monthly social benefits receipts data (SHSS) to define an indicator for maternity leave in November (when employment status in IDA is defined). We save this information in pnr\_maternityleave.dta.

Next, we merge these leave spells onto main\_nursing\_timeseries\_extended.dta. We define absence in November (excluding maternity leave) and estimate difference-in-difference models for individual takeup by health care worker type for 1993-1995. We save estimated coefficients based on worker characteristics to reflect exposure to the reform in file exposure\_fullleave.dta.

Prepare analysis data for employment:

Prepare\_aggregate \_county\_exposure.do

This file computes exposure to the leave program at county and health sector level and saves the main analysis file at county level. This file combines timeseries data (main\_nursing\_timeseries\_extended.dta) with exposure estimates (exposure\_fullleave.dta), then aggregates at the county and health care sector level. We add county-level demographics (county\_demographics.dta) and generate additional control variables for the main regressions.

We save the main analysis file county\_regression\_dataset\_prerun\_allgroups\_fullleave\_final.dta

Prepare analysis data for hospitals:

Prepare\_hospitalsample\_PartB.do

This file finalizes the preparation of the hospital outcome data, that are used for the analysis of AMI readmission and mortality rates.

The file loads hosp\_regs\_countyyear\_extra.dta and merges the data with county\_regression\_dataset\_prerun\_allgroups\_fullleave\_final, which contains relevant control variables for the regression analysis. The final data set is saved under hosp\_regs\_countyyear\_oldsimple.dta.

Prepare analysis data for nursing homes:

Prepare\_mortalitysample\_PartC.do

This file finalizes the preparation of the preparation of elderly mortality data, that are used for the analysis of unconditional nursing home and total mortality.

The file loads mortality\_residuals\_pre\_long.dta, aggregates the data to the county-year level, and merges the data with county\_regression\_dataset\_prerun\_allgroups\_fullleave\_final , which contains relevant control variables for the regression analysis. The final data set is saved under mortality\_residuals\_pre\_county\_wild.dta

Prepare\_ Nursing\_home\_mortalityCONDmort.do

This file prepares the analysis of conditional nursing home morality.

The file loads elderly\_data.dta and merges the data with information on the patient’s hospital history contained in nhsample\_prehosps.dta. The data are then aggregated to the county-year level and merged with county\_regression\_dataset\_prerun\_allgroups\_fullleave\_final, which contains relevant control variables for the regression analysis. The final data set is saved under mortality\_residuals\_COND\_county.dta.

Prepare\_Nursing\_home\_mortalityCONDresidents.do

This file prepares the analysis the elderly share residing in nursing homes.

The file loads elderly\_data.dta and merges that data with information on the patient’s hospital history contained in nhsample\_prehosps.dta. The data are then aggregated to the county-year level and merged with county\_regression\_dataset\_prerun\_allgroups\_fullleave\_final, which contains relevant control variables for the regression analysis. The final data set is saved under mortality\_residuals\_CONDres\_county.dta.

Prepare\_ Nursing\_home\_mortalityuncond\_causes.do

This file prepares the analysis the causes of nursing home deaths.

The file loads mortality\_residuals\_pre\_long.dta, aggregates the data to the county-year level, and merges the data with county\_regression\_dataset\_prerun\_allgroups\_fullleave\_final, which contains relevant control variables for the regression analysis. The final data set is saved under mortality\_residuals\_pre\_county\_causes.dta.

Prepare\_Nursing\_home\_mechanisms.do

This file prepares the analysis the mechanisms behind nursing home deaths.

The file loads elderly\_data.dta and admissions\_10\_2019\_Stata.dta, aggregates the merged data to the county-year level and merges with county\_regression\_dataset\_prerun\_allgroups\_fullleave\_final, which contains relevant control variables for the regression analysis. The final data set is saved under nhome\_causedeath\_mechanism.dta.

Analysis files:

The following list explains the high-level file structure to generate the main results of the paper. In the table below, we then list the subfiles that generate each figure and table respectively, together with the file name used for each output.

Master\_labormarket.do

* This file generates Figures 1-5 and Table 1. This file also estimates the IV results for Tables 2 and 3, respectively.
* Subfiles can be run independently to generate specific Figures.

Master\_healthoutcomes.do

* This file first loads a series of do-files, that each produce the difference-in-difference and the event-study regression results along with the bootstrapped confidence intervals for one outcome at the time. This part needs to be run first.
* Building on these results, the code executes Figures\_wild\_main.do, which produces the graphs contained in Figures 6 and 7.
* Building on the regression results the code executes Tables\_wild\_main\_plusIV.do, which produces Tables 2-5.

## 

## List of tables and programs

|  |  |  |
| --- | --- | --- |
| Figure/Table # | Generated in program: | Output file name: |
| Table 1 | Results\_firststage\_nh.do  Results\_firststage\_hosp.do | Tab1\_employment\_main.tex |
| Table 2 | Tables\_wild\_main\_plusIV.do | Tab2\_hospquality\_main.tex |
| Table 3 | Tables\_wild\_main\_plusIV.do | Tab3\_nhquality\_main.tex |
| Table 4 | Tables\_wild\_main\_plusIV.do | Tab4\_causeofdeath.tex |
| Table 5 | Tables\_wild\_main\_plusIV.do | Tab5\_causeofdeath\_mechanisms.tex |
| Figure 1 | estimate\_takeup\_fulleave.do | Fig1\_rd\_allleave\_skilled\_CI.eps  Fig1\_rd\_allleave\_unskilled\_CI.eps  Fig1\_rd\_allleave\_docs\_CI.eps |
| Figure 2 | Results\_firststage\_timeseries.do | Fig2\_scatter\_timeseries\_nurses\_hosp.eps  Fig2\_scatter\_timeseries\_nurses\_nh.eps  Fig2\_scatter\_timeseries\_assistants\_hospnh.eps  Fig2\_scatter\_timeseries\_docs\_hospnh.eps |
| Figure 3 | Results\_institutional\_educ.do | Fig3\_education\_nurses\_newenrollment\_educdata.eps  Fig3\_education\_assistants\_newenrollment\_educdata.eps |
|  | Results\_institutional\_wagebargain.do | Fig3\_wages\_combined\_1y10y.eps |
|  | Results\_institutional\_immigration.do | Fig3\_immigrant\_skillednurse\_newarrival\_hospnh.eps |
|  | Results\_institutional\_leavespells.do | Fig3\_hist\_childleave\_skilledunskilled\_9599.eps |
| Figure 4 | Results\_firststage\_scatterplots\_9196\_final.do | Fig4\_scatter9691\_nurseemplhosp.eps  Fig4\_scatter9691\_nurseemplnh.eps |
|  | Results\_firststage\_hosp.do | Fig4\_lambdas\_skilled\_hosp.eps |
|  | Results\_firststage\_nh.do | Fig4\_lambdas\_skilled\_nh.eps |
| Figure 5 | Results\_firststage\_experience\_short.do | Fig5\_hist\_fullexp\_1993\_skilled\_hosp.eps  Fig5\_hist\_fullexp\_1993\_skilled\_nh.eps |
| Figure 6 | Figures\_wild\_main.do | Fig6\_ scatter9691\_hospmort.eps  Fig6\_ scatter9691\_hospread.eps  Fig6\_lambdas\_hospmort.eps  Fig6\_lambdas\_hospread.eps |
| Figure 7 | Figures\_wild\_main.do | Fig7\_scatter9691\_nhdeath.eps  Fig7\_scatter9691\_death.eps  Fig7\_lambdas\_nhuncond.eps  Fig7\_lambdas\_totuncond.eps |