This document provides detailed information on the steps needed to replicate the full analysis in the paper **Does your doctor matter?** by Rita Ginja, Julie Riise, Barton Willage and Alexander Willén.

Data access

The data employed in the analysis are drawn from Norwegian administrative registers. Researchers can gain access to the data by submitting a written application to Statistics Norway (SSB). The application should include a detailed research proposal describing the goals and methods of the project, a detailed list of variables, the selection criteria to be used, and how the research will be funded. Once received, applications must be certified by in order to ensure that data are processed in a manner that protects the personal integrity of individuals surveyed. Conditional on this approval, Statistics Norway will then determine which data one may obtain in accordance with the research plan; all data are processed by Statistics Norway. Individuals must provide a list of all individuals who will have access to the data along with description of a secure way in which the data will be stored for the period of the research project. Inquiries about access should be addressed to: Statistics Norway Postboks 2633 St. Hanshaugen, NO-0131 Oslo.

Researchers wishing to replicate our work should refer **project number 607173** "*Demografiske og sosiale endringer – utfordringer for sysselsetting og velferd*" at SSB. This project refers to the extracts made available to the Department of Economics - University of Bergen. This should be done after approval of the data from owners and the Regional Committees for Medical and Health Research Ethics. The sample includes everyone who resided in Norway in the period 1992-2015 and their parents (approx. 6.5 million people).

Steps for replication

All the data preparation and analysis for this paper was performed using `Stata` version 16.1. The full replication package and raw data are stored at the servers of the University of Bergen. We provide programs that create a simulated data set that allow researchers to replicate the results of our paper, along with the original do files used to manipulate the raw data extracts.

Memory and runtime requirements

The code was run on a 20-core (2 processors) Intel Xeon Silver 4114 CPU @ 2.20GHz server with 256 GB of RAM, Windows Server 2016 Standard.

Do-file structure

Data preparation

The data preparation is easiest to replicate by following the order of the files in `master.do`. The research should start by preparing the raw data for analysis following the file `master.do`:

1. `clean_npr.do`. Uses information from the *National Patient Register* ("Norsk pasientregister", 2008-2020, Helsedirektoratet (2017b)) to construct the health indicators. These data are available from the Norwegian Health Directorate; see https://www.helsedirektoratet.no/tema/statistikk-registre-og-rapporter/helsedata-og-helseregistre/norsk-pasientregister-npr. In order to access these datasets, approvals from the Regional Committees for Medical and Health Research Ethics are also required besides approval from the primary data owners.

2. `clean_health.do`. Uses information from the *Control and Payment of Health Refunds* ("Kontroll og utbetaling av helserefusjoner", 2006-2020, Helsedirektoratet (2017a)) to construct the health indicators. These data are available from the Norwegian Health Directorate; see https://www.helsedirektoratet.no/tema/statistikk-registre-og-rapporter/helsedata-og-helseregistre/kuhr. In order to access these datasets, approvals from the Regional Committees for Medical and Health Research Ethics are also required besides approval from the primary data owners.

3. `clean_labormarket.do`. Uses information from the *Social Security Database* and information on employment is obtained from several register-based employment databases, and in particular we rely on information from the Employer and Employee Register from the Norwegian Labour and Welfare Administration that is maintained by Statistics Norway (Statistisk Sentralbyrå, 2017c,d), and the documentation about these data can be found at https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/arbeidsmarked and https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/fd-trygd.

4. `Cleaning_PreFinalSet.do`. This file gathers and cleans information from several data sets. In particular, it obtains several income measures for each individual from the *Tax Records*. Data on individuals' income is collected by the Tax Authority and kept by the Statistics Norway. The documentation about these data can be found at https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/inntekt. Itgathers information on sick leave for all individuals, demographic information, education, income and Social Security Records. Information about the data containing demographic characteristics (Statistisk Sentralbyrå, 2017a) can be found at https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/befolkning. Information about each individual's education is available from the register for the Demographic information and highest education ("Demografiske opplysninger og høyeste utdanning") (Statistisk Sentralbyrå, 2017b).

5. `Cleaning_PostFinalSetR2.do`. Prepares the final samples for the various analyses using the data setts prepared from the previous do files.

TABLE 1.
Data Sources

Data source	Main description	Variables	Years available
Control and Payment of Health Reimbursement (KUHR)	Individual primary care visits (to GPs or emergency rooms)	Dates (year) of visits, with related diagnoses and symptoms (ICPC2)	2006-2020
Norwegian Patient Registry (NPR)	Individual inpatient and outpatient visits in specialist care.	Entry & discharge dates (year), diagnoses (ICD10)	2008-2020
Mortality Records (DAR)	Individual death event	Date (year) & cause of death (ICD10)	up to 2020
Education Records	Individual level information	Highest Completed Degree	1970-2020
Tax Authority Records	Annual information for eligible individuals (those formally employed)	Labor earnings and income from other sources	1992-2020
Population Records	Annual demographic information about all individuals	Marital status, municipality of residence, gender, age, nationality	1967-2020

Note: This table presents the time frame covered by the administrative records used in the analysis, the variables and years of data used.

Analysis

The table below lists each set of results in the paper (tables, figures), along with the do-file which generates it. See also the do-file `master.do`:

Exhibit	**Stata do file**
Figure 1, Figure A7, Table A4	`ValueAdded.do`
Figure A8, Tables A3, A5 (col5, 6), Table A8	`ValueAdded_all.do`
Figure A9, Table A14	`ValueAdded_het.do`
Table A12, Table A13, Figure A6	`ValueAdded_robust.do`
Tables 1, 2, 4 to 7, A1, A2, A7, A9, A10, A15, A17, A18 Figures A2, A4, A5	`fe_analysis.do`
Table 3, Tables A3, A5 (col 1-4)	`fe_descriptives.do`
Tables 8, 9, 10, A6, A19	`endog_va.do`
Figure A1	`map.do`
Figure A3	`mortality_byconditions.do`
Table A11	`external_validity.do`
Table A16	`lasso_final.do`

References

Helsedirektoratet (2017a). 'Kontroll og utbetaling av helserefusjoner–databasen, 2006-2016', https://www.helsedirektoratet.no/tema/statistikk-registre-og-rapporter/helsedata-og-helseregistre/kuhr, (last accessed: 11 July 2022).

Helsedirektoratet (2017b). 'Norsk pasientregister, 2008-2014', https://www.helsedirektoratet.no/tema/statistikk-registre-og-rapporter/helsedata-og-helseregistre/norsk-pasientregister-npr, (last accessed: 11 July 2022).

Statistisk Sentralbyrå (2017a). 'Befolkning Registret', https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/befolkning, (last accessed: 11 July 2022).

Statistisk Sentralbyrå (2017b). 'Nasjonal utdanningsdatabase, 1970-2014', https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/utdanning, (last accessed: 11 July 2022).

Statistisk Sentralbyrå (2017c). 'Registerbaserte sysselsettingsstatistikken', https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/arbeidsmarked, (last accessed: 11 July 2022).

Statistisk Sentralbyrå (2017d). 'System for persondata', https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/arbeidsmarked, (last accessed: 11 July 2022).