README File for Data and Programs
"Can Female Doctors Cure the Gender STEMM Gap? Evidence from Exogenously-assigned General Practitioners"

Review of Economics and Statistics
RESTAT MS#23454
Julie Riise, Barton Willage, Alexander Willén

OVERVEIW OF DATA USED

Here we describe the types of administrative data sets used from Norway. Note that all data are restricted, and are not available for public posting; this paper has been exempted from the data requirement. See below for information on the process used to obtain this data.

All data sets are linked using the person of interests unique ID (lopenr), or father's or mother's unique ID (lopenr_far/ lopenr_mor), or GP's unique ID (BehandlerID, renamed to GPID).

DETAILS ABOUT OBTAINING DATA

The data used in the analysis are drawn from Norwegian administrative registers. Researchers can gain access to the data by submitting a written application to Statistics Norway. 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 in order to ensure that data are processed in a manner that protects the personal integrity of individuals surveyed. For the health registers, applications must also be sent to and certified by the register owners and the regional committee for medical and health service research (REK).

Conditional on such approvals, Statistics Norway will then determine which data one may obtain in accordance with the research plan; all data are processed by Statistics Norway. A list of all individuals who will have access to the data must be provided along with a 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.

We use the following datasets:

1. Population register (month/year of birth, family links, municipality)
2. GP register (GP patient lists, GP age and gender, municipality of GP office, swaps between GPs)
3. Norway Control and Payment of Health Reimbursement Database - KUHR (dates of visits, related diagnoses, actions taken)
4. Tax register (earnings)
5. Register on educational attainment (years of education)

LIST OF FILES AND DESCRIPTIONS

* GP_Swap_Cleaning.do – imports general practice (GP) data set, and cleans and determines primary treatment variables and doctor characteristics. Also cleans and merges parent characteristics (such as education, income, municipality). The resulting data sets are **GP_Chars.dta**, **PrevGP_Chars.dta**, and **finalset_clean.dta** . **MainRHS.dta.**

* Health_Cleaning.do – imports, cleans, and merges several health-related data sets. Health outcomes include fertility, mental health diagnoses, and number of diagnoses. Health outcomes are determined for the mother and the daughter. The resulting data sets are **HealthOutcomes.dta** and **MomHealthOutcomes.dta**.

* Education_Cleaning.do – imports, cleans, and merges several education-related data sets. Education outcomes include grade point average for lower secondary (compulsory) school and upper secondary (high school) school; educations track/concentration/major in secondary and post-secondary school; and education level achieved. The resulting data set is **EdOutcomes.dta**.

*Analysis.do – performs analysis for ALL tables and figures in manuscript and online appendix. ALL cleaning files must be run first. The code for each table/figure is labeled. Data in opened for each table/figure, so only the CD (directory) set at the top and the relevant code needs to be run once all cleaning files are run.

SOFTWARE VERSION AND OPERATING SYSTEM

OS version: Windows Server 2016 Standard (x64)

Stata version: Stata MP 16 (64-bit)

ORIGINAL FILE LOCATIONS ON SERVER

* S:\Project\DemoSos2\data\DAR
* S:\Project\DemoSos2\data\KUHR
* S:\Project\DemoSos2\data\NPR
* S:\Project\DemoSos2\data\Utd

CODEBOOK

For codebook describing variables used in the analysis, see PDF titled “Analytic Files Codebook.pdf”.