Readme file for the replication of

"Patient-Physician Race Concordance, Physician Decisions, and Patient Outcomes"

Authors: Han Ye and Junjian Yi

Environment

- System environment: Windows 10
- Software environment: Stata/MP 16.0. All the commands are either a part of the standard package or can be downloaded using the "ssc" command.

Data

We use hospital administrative data obtained from a proprietary source in Southeast Asia. To strictly observe the confidentiality of the data, as per specified by the non-disclosure agreement, we would not be able to post the data on public domain. Nonetheless, we understand that ensuring replicability is essential to any scientific endeavor. We extract variables used to replicate all the results reported in the paper and the online appendix (EDdata.dta in the package), and publish all computer code used for the analysis. We will submit all data to a Research Data Repository (RDR) located at the National School of Department (NSD), Peking University. The RDR is to safeguard researchers' academic integrity and is immune to subsequent data alteration after submission. If research misconduct is legitimately suspected, NSD can decide to set up an investigative committee, who would be granted access to the data stored in the RDR. Interested researchers can contact the author at junjian.yi@gmail.com for further information on how to access the original data.

Included Files

The replication package includes the stata data file and do files that generate all the results reported in the paper and online appendix.

Data file

– EDdata.dta (data used to replicate all the results reported in the paper and the online appendix)

Code files

– 0_master.do (run all the do files)
– 1_data clean.do (data clean)
– 2_variable construction.do (generate variables)
– 3_data extraction.do (extract data for analysis)
– 4_main tables.do (replicate all the tables in the paper)
– 5_appendix tables.do (replicate all the tables in the online appendix)

Data dictionary for EDdata.dta

obs: 254,327			
vars: 29			
variable name	storage type	display format	variable label
Admit	double	%10.0g	Inpatient admission (dummy)
ADI	float	%9.0g	Advance diagnostic imaging (dummy)
Consulttime	double	%10.0g	Length of consultation (minutes)
lnConsulttime	double	%10.0g	Log (length of consultation+1)
Revisit	double	%10.0g	7-day revisit after ED discharge (dummy)
Death	double	%10.0g	Death in the ED (dummy)
detailedICD	float	%9.0g	detailed diagnosis (dummy)
Male	double	%10.0g	Patient gender(dummy)
age	double	%10.0g	Patient age (years)
agegroup	float	%12.0g	Patient age groups
severitylevel	double	%10.0g	triage severity level
patracegroup	float	%9.0g	patient race group
phyracegroup	float	%9.0g	physician race group
SameRace	float	%9.0g	Patient-physician race concordance
consultstartdow	double	%10.0g	consultation start day of week
consultstartmonth	double	%tm	consultation start month by year
consultstarthour	double	%10.0g	consultation start hour of day
doctorid	float	%9.0g	doctor id
experienced	float	%9.0g	Above-median experience (dummy)
severecase	double	%10.0g	severe case (dummy)
ImageUse	float	%9.0g	Imaging utilization (dummy)
LabUse	float	%9.0g	Lab test utilization (dummy)
TestCount	float	%9.0g	Number of diagnostic tests

BrDx	byte	%81.0g	Broad diagnostic category
Proportion_Samerace	float	%9.0g	% same-race physicians
SpDx	float	%34.0g	Specific diagnosis
priorcase	float	%9.0g	number of patients treated
workhour	float	%12.0g	number of hours worked
irelto	float	%18.0g	hours relative to EOS