Code and data for Stoye, "The Distribution of Doctor Quality: Evidence from Cardiologists in England"

This file contains information about all data and do-files used in Stoye (2025).

All files run on Stata 16.1.

Data

Please note that the analysis is derived from the Hospital Episode Statistics (Admitted Patient Care datasets for the financial years 2003-04 to 2017-18, Accident and Emergency datasets for the financial years 2007-08 to 2017-18, and the Outpatient Care datasets for the financial years 2005-06 to 2017-18), the Office for National Statistics (ONS) UK mortality records and the NHS Electronic Staff Record (financial year 2017-18 only). **The data sharing agreement does not allow us to post the original data, or any variables derived from these data.**

Researchers can apply for the Hospital Episodes Statistics and ONS mortality data from NHS Digital through the Data Access Request Service (DARS). The application process, and the security requirements for holding the data, are described here: https://digital.nhs.uk/services/data-access-request-service-dars/data-access-request-service-dars-process. Note that as part of this process, researchers must describe how their research using these data will benefit the UK health and social care system, and a route to achieving these benefits.

The NHS Electronic Staff Records contains an extract of NHS monthly payroll records, made available to researchers at the Institute for Fiscal Studies for workforce-related research purposes. This project uses an anonymised extract of cardiologists employed by NHS hospitals for the financial year 2017-18. Researchers should contact the UK Department of Health and Social Care to discuss access arrangements for these data.

The required proprietary datasets are:

- Admitted Patient Care (APC) Hospital Episode Statistics 2003-04, 2004-05, 2005-06, 2006-07, 2007-08, 2008-09, 2009-10, 2010-11, 2011-12, 2012-13, 2013-14, 2014-15, 2015-16, 2016-17, 2017-2018
- A&E Hospital Episode Statistics 2007-08, 2008-09, 2009-10, 2010-11, 2011-12, 2012-13, 2013-14, 2014-15, 2015-16, 2016-17, 2017-2018
- Outpatient Hospital Episode Statistics 2005-06, 2006-07, 2007-08, 2008-09, 2009-10, 2010-11, 2011-12, 2012-13, 2013-14, 2014-15, 2015-16, 2016-17, 2017-2018
- ONS mortality data for all Acute Myocardial Infarction (AMI) patients attending NHS hospitals between April 2005 and March 2018
- NHS Electronic Staff Record data for cardiologists employed by NHS hospitals in 2017-18

For a full list of variables in the Hospital Episode Statistics and mortality records datasets see: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary

Publicly available and non-disclosive data required for the code are also saved in the folder. These files are:

- types – lists NHS trust types in 2017-18 (used to identify teaching hospitals)
- msoa_vars – socioeconomic characteristics at the 2001 lower layer super output area
- Msoa_vars – socioeconomic characteristics at the 2001 middle layer super output area
- lsoa_areatypes – geographic information at the 2001 lower layer super output area
- oper_imp_cost – imputed costs for each recorded operation (used to cost previous inpatient treatment) based on NHS tariff payments averaged at the primary operation level
- diag_no_oper_imp_cost – imputed costs at the diagnosis level (used to cost previous inpatient treatment where operation data is missing) based on NHS tariff payments averaged at the primary diagnosis level

Code

After obtaining the data specified above, all .do files can be executed to produce the figures and tables in the paper and appendix, after updating that paths in "0. master_dofile.do" to match your machine.

These globals are:

- rawdata – folder where raw inpatient data are saved
- rawdata_op – folder where raw outpatient data are saved
- eddata – folder where raw A&E data are saved
- mortalitydata – folder where raw mortality data are saved
- consids – folder where doctor IDs are saved for pre-2010 data (note: this would likely be unnecessary when applying for the data now, where doctor IDs (pconsult) could be requested with the initial dataset)
- savedata – data created by the do-files
- dofile – folder where dofiles are saved
- inputs – calls input data (the 6 datasets outlined above) once saved in here
- results – stores all results
- staffdata – calls Electronic Staff Record dataset

Running the file 0. master_dofile.do will run all files in correct order, and contains details of what each file does. A summary is listed below:

1) Files that construct the baseline data
    - 1. extract_records.do - extracts patient records for AMI patients
    - 2. make_cardiologist_list.do – makes list of IDs for all cardiologists who treat AMI patients between 2003 and 2017
    - 3. build_patient_data.do - cleans patient dataset and identifies episode of care that is studied
    - 4. make_analysis_dataset_part1.do - Merges in various datasets and starts to create analysis dataset. Also calls do-file (4a) that create records of previous treatment

- 5. make_analysis_dataset_part2.do" - finalises creating the dataset used in the baseline analysis. Also calls do-files (5a and 5b) that create connected sets and create measures of cardiologist experience

2) Descriptive results
   - 6. make_sumstats.do - Produces results in Tables 1 and 2, Appendix Table A1 and Appendix Figures A1 and A2

3) Results (Section 5.1)
   - 7a. make_table3.do - produces results in Table 3
   - 7b. make_figure1_figureA3.do - produces Figure 1 and Figure A3
   - 7c. make_TableA2.do - produces Table A2

4) Robustness checks (Sections 5.2 and 5.3)
   - 8a. make_figure2.do - produces Figure 2
   - 8b. make_pastop.do - creates record of past outpatient care for AMI patients, used in results in Table 4
   - 8c. make_pastae.do - creates record of past A&E (ED) care for AMI patients, used in results in Table 4
   - 8d. make_gpadmits.do - creates record of elective and emergency hospital admissions at the GP practice level, used in results in Table 4
   - 8e. make_table4.do - produces results displayed in Table 4
   - 8f. make_tableA3.do - produces results displayed in Table A3; also calls two other dofiles (8fi and 8fii) to create the analysis dataset using the sample from 2009 onwards
   - 8g. make_figureA4.do - produces Figure A4
   - 8h. make_tableA4_cols1and2.do - produces columns 1 and 2 in Table A4; also calls dofile 8hi to create the sample of hospitals that always conducted PCI throughout the period
   - 8i. make_tableA4_cols3and4.do - produces columns 3 and 4 in Table A4; calls dofiles 8ii and 8iii to create the sample of patients without a previous interaction with the treating doctor
   - 8j. make_figure_A5_and_A6.do - produces Figures A5 and A6
   - 8k. test_trust_specific_seasonality.do - Seasonality test in footnote 31
   - 8l. rsquared_analysis.do - R-squared comparisons between models in Section 5.2.2
   - 8m. make_Table5.do - produces results in Table 5
   - 8o. make_Table6.do - produces results in Table 6 (columns 2 and 3, column 1 is the same as column 4 in Table 3)
   - 8p. make_TableA5.do - produces results for Table A5
   - 8q. KSS_footnote37.do - sets up data and produces comparison results for KSS tests in footnote 37

5) Hospital correlates (Section 5.4)
   - 9a. make_Figure3_and_TableA6.do - makes Figure 3 and results for Table A6
   - 9b. make_FigureA7.do - makes Figure A7
   - 9c. make_Table7.do - makes results Table 7

- 9d. make_figureA8.do - makes Figure A8

6) Reallocation exercises (Section 6)
   - 10a. make_Figure4_and_FigureA9.do -makes Figure 4 and Figure A9
   - 10b. reallocation_calculations_section6.do - calculates mortality gains from various reallocation exercises discussed in Section 6.2