

Supplemental Data from
“The Earnings and Labor Supply of U.S. Physicians”

Joshua D. Gottlieb Maria Polyakova
Kevin Rinz Hugh Shiplett
Victoria Udalova

The Quarterly Journal of Economics, May 2025

Overview

Machine-readable data files for exhibits in Gottlieb, Polyakova, Rinz, Shiplett, and Udalova “The Earnings and Labor Supply of U.S. Physicians”, *The Quarterly Journal of Economics*, Volume 140, Issue 2, May 2025, Pages 1243-1298, <https://doi.org/10.1093/qje/qjaf001>.

This data product is intended to inform interested parties of ongoing research and to encourage discussion. Any views expressed are those of the authors and not those of the U.S. Census Bureau. Access to the data has been approved for this project by U.S. Census Bureau, CMS, IRS, and SSA because it benefits Census Bureau programs in accordance with Title 13 of the U.S. Code. The Census Bureau’s Disclosure Review Board and Disclosure Avoidance Offices have reviewed this data product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release, authorization nos. CBDRB-FY23-0319, CBDRB-FY2023-CES005-024, and CBDRB-FY24-0456.

Details regarding data sources and construction can be found in the replication package on Harvard Dataverse at 10.7910/DVN/CJ1UM1, and Appendix B of the paper.

File names of data products below match exhibit numbers in the published manuscript and the online appendix. For example, the estimates graphed in Appendix Figure E.1 are contained in `docinc_figure_e1.csv`. The README PDF defines each variable in each CSV file distributed here.

Files

`docinc_figure_2.csv`

- `specialty_markers`: Medicare specialty or NRMP specialty group.
- `sample`: Physicians in our 2005-2017 panel, of ages 20-70 (“ageall”) or ages 40-55 (“age4055”).
- `ptotinc`: Mean individual total income within sample and specialty group (in 1,000s of 2017 dollars).
- `wkh`: Average weekly working hours for physicians in specialty (measured using the ACS).
- `resdur_allyrs`: Average imputed years of training.
- `n_pub_sn_match`: Self-reported number of abstracts, presentations, and publications by applicants to specialty (data from the NRMP).
- `pred_share_us_sn`: Share of physicians with U.S. degree, conditional on work hours and years of training.
- `pred_ptotinc`: Mean of individual total income, conditional on work hours and years of training (in 1,000s of 2017 dollars).
- `freq`: Number of physicians in sample and specialty.

`docinc_figure_3.csv`

- `sample`: 2017 observations, by occupation (physicians or lawyers) and age group (ages 20-70 or ages 40-55).

- `stfips`: State FIPS code.
- `n`: Number of observations in `sample`.
- `mean_age`: Mean age of sample in years.
- `mean_female`: Share of sample that is female.
- `mean_ranked_school`: Share of physicians that graduated from a ranked medical school.
- `mean_ptotinc`: Mean individual total income in 2017 dollars.
- `mean_logptotinc`: Mean log individual total income, where income is measured in 2017 dollars.
- `mean_totw2comp`: Mean total Form W-2 earnings in 2017 dollars.
- `mean_adjgross`: Mean adjusted gross income in 2017 dollars.
- `mean_top1`: Share in top 1% of AGI national distribution.
- `mean_with25kbizinc`: Share with more than 25,000 2017 dollars of business income.
- `mean_retired_1099ssa`: Share receiving Form 1099-SSA.
- `mean_wkh`: Mean weekly working hours for subsample observed in the ACS.
- `mean_income_acs`: Mean income in 2017 dollars for subsample observed in the ACS.
- `mean_primary_care`: Share of physicians in primary care.
- `mean_fe_cz`: Mean CZ fixed effects from equation 1.
- `mean_fe_id`: Mean individual fixed effect from equation 1.
- `mean_logptotinctwfe`: Mean log individual total income, where income is measured in 2017 dollars, for physicians in the estimation sample for equation 1.

`docinc_figure_e1.csv`

- `var`: Log individual total income for 20- to 70-year-old U.S. physicians in 2017. Income is measured in 2017 dollars.
- `varlab`: Log annual income in 2017 dollars.
- `xaxis`: x-axis position of each income bin (for graphing).
- `frac`: y-axis height for each income bin (for graphing).
- `mean`: Mean log individual total income.
- `sd`: Standard deviation of log individual total income.
- `p1,p5,p25,p50,p75,p95,p99`: The 1st, 5th, 25th, 50th, 75th, 95th, and 99th percentiles of log individual total income.

docinc_figure_e2.csv

- **var**: Log adjusted gross income (AGI) for 20- to 70-year-old U.S. physicians in 2017. Income is measured in 2017 dollars.
- **varlab**: Log AGI.
- **xaxis**: x-axis position of each income bin (for graphing).
- **frac**: y-axis height for each income bin (for graphing).
- **mean**: Mean log AGI.
- **sd**: Standard deviation of log AGI.
- **p1,p5,p25,p50,p75,p95,p99**: The 1st, 5th, 25th, 50th, 75th, 95th, and 99th percentiles of log AGI.

docinc_figure_e3.csv

- **year**: Data year.
- **notadj_einsize1**: Share of physicians in firms (EIN) with only one physician.
- **notadj_einsizemissing**: Share of physicians where firm size (EIN) is missing.
- **notadj_ptotinc**: Individual total income (in 1,000s of 2017 dollars).
- **notadj_top1**: Share of physicians in the top 1% of the national income (AGI) distribution.

All values are computed using the sample of all physicians aged 20 to 70 observed during a given year. For each variable with the **notadj** prefix, there is a corresponding variable with the **adj** prefix. The values of these variables have been regression adjusted to equalize the composition of age, sex, Medicare specialties, and states across time to 2017 levels.

docinc_figure_e4.csv

- **agebin**: Age category.
- **restriction**: Sample definition.
- **ptotinc**: Individual total income in 2017 dollars.
- **income_acs**: Income as reported in the ACS in 2017 dollars.
- **totw2comp**: Total wage earnings as reported on Form W-2, in 2017 dollars.
- **pftloss**: Total business income in 2017 dollars (adjusted for married filing jointly physician couples).
- **wkh_imputed**: Weekly working hours as reported in the ACS. For individuals who report they are “not in the labor force”, we impute working hours to be 0.
- **schd_c**: Share filing Schedule C.
- **received_1099ssa**: Share receiving Form 1099-SSA.

docinc_figure_e5.csv

- **ranked_top5**: Physicians in sample went to a top-5 MD program, as ranked by the U.S. News & World Report.
- **spec_category_code**: 9-category specialty code.
- **spec_category_desc**: Specialty name.
- **top5_school_share**: Share of physicians from top-5 MD programs.
- **mean_hourly_income**: Mean hourly income within specialty, in 2017 dollars.

All values were calculated using our sample of physicians aged 40-55 for years 2005-2017.

docinc_table_1.csv

- **sample**: Sample of physicians, by year and age-bracket.
- **depvar**: Dependent variable.
- **count**: Number of physicians in subsample.
- **no_unique_npi**: Number of unique physicians in sample.
- **no_phys_year**: Number of physician-year observations in sample.
- **mean**: Mean of dependent variable.
- **median**: Median of dependent variable.
- **stdev**: Standard deviation of dependent variable.
- **share_sample**: Share of the sample that meets X criteria (defined by **depvar**).
- **subpanel**: Label for subpanel within Table 1.

All income values are in 2017 dollars.

docinc_table_2.csv

- **panel**: Panel titles as they appear in Table 2.
- **rownames**: Names of rows as they appear in Table 2.
- **top{x}percent**: Top X% of physicians by income in the group defined by **rownames**.

All values are calculated using our sample of 40-55 physicians in 2017. All income values are in 2017 dollars.

docinc_table_e1.csv

- **specialty_category**: 9-category specialty code.

- `medicare_specialty_code`: Medicare specialty code.
- `medicare_specialty_description`: Medicare specialty description.

`docinc_table_e5.csv`

- `sample`: Physician age-group.
- `depvar`: Outcome variable. `ptotinc` is individual total income in 2017 dollars; `adjgross` is adjusted gross income (AGI) in 2017 dollars; `top1` is an indicator for belonging to the top 1% of households in the national AGI distribution; `totw2comp` is total wages as reported on Form W-2, in 2017 dollars.
- `mean`: Mean of `depvar` for sample.
- `median`: Median of `depvar` for sample.
- `stdev`: Standard deviation of `depvar` for sample.
- `spec_category_desc`: 9-category specialty description.
- `year`: Data year.