

# Readme: Mean Physician Earnings by Commuting Zone and Specialty\*

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

First release: July 21, 2020  
Data last updated: July 21, 2020  
Readme last updated: July 21, 2020

This document contains a brief description of the mean physician earnings data accompanying “Who Values Human Capitalists’ Human Capital? Healthcare Spending and Physician Earnings.” The full paper is available [here](#). Data are available for all physicians in the 125 largest commuting zones, and by specialty within these commuting zones as sample sizes permit. Files are provided in Stata and CSV formats. The two versions are identical in content and layout.

## 1 Sample

The sample used to create this physician earnings database includes all physicians with a National Provider Identifier (NPI) in the 2018 National Plan & Provider Enumeration System (NPES) who filed Form 1040 or had Form W-2 filed on their behalf in 2017, for whom a medicare specialty code could be determined, and who were between the ages of 30 and 65. Physicians’ ages are obtained from the Social Security Administrations Numerical Identification System (Numident) database. For the purposes of this database, a “physician” is defined as anyone with a primary taxonomy code starting with “20” in the 2018 NPES. Additional details regarding data sources and construction can be found in Section 2.2 of the paper.

## 2 Geography

Data were aggregated to 1990 Commuting Zones (CZs). A description of the 1990 CZs and the methodology used to delineate them can be found [here](#).

## 3 Physician Specialty

Physicians are classified into specialties using Medicare specialty codes. Medicare specialty codes were obtained from physicians’ primary taxonomy code as recorded in the 2018 NPES using the crosswalk available [here](#).

---

\*Any views expressed herein are those of the authors and not those of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The statistical summaries reported in this paper have been cleared by the Census Bureau’s Disclosure Review Board, release authorization number CBDRB-FY20-349.

## 4 Variable Descriptions

**cz90:** Five-digit CZ code. The first three digits are the Labor Market Area (LMA) code where the CZ is located, last two digits are the CZ number within the LMA (“00” if only one CZ in an LMA).

**czname:** Description of CZ. This is the name of the LMA where the CZ is located, followed by the CZ number in parentheses if the LMA contains multiple CZs. The CZ number is taken from the final digit of the CZ fips code (*cz90*), and may not start from “1” if data for some CZs in the LMA have been suppressed due to disclosure restrictions.

**speccode:** Two-character Medicare specialty code. A small number of specialty codes are non-numeric. Code “00” is not a Medicare Specialty Code; it indicates estimates that apply to the sample of all physicians.

**specdesc:** Description of the Medicare Provider/Supplier Type associated with speccode.

**incper:** CZ-level mean of physician professional earnings (W-2 earnings + tax unit self employment earnings) in 2017. Mean is calculated using only positive values. This corresponds to the primary income measure used in the paper. See Section 2.3 and Appendix Section A.3 for a detailed description of how this variable was constructed.

**log\_incper:** CZ-level mean of  $\log(\textit{incper})$  in 2017.

**n\_incper:** Number of physicians in the CZ in 2017 used in calculating *incper* and *log\_incper*. Since only positive values were included, both *incper* and *log\_incper* use the same samples. This includes all physicians in a CZ with an NPI in the 2018 NPPES who had positive *incper* in 2017.

**agi:** 2017 CZ-level mean of physician adjusted gross income from Form 1040. Mean is calculated using only positive values.

**log\_agi:** CZ-level mean of  $\log(\textit{agi})$  in 2017.

**n\_agi:** Number of physicians in the CZ in 2017 used in calculating *agi* and *log\_agi*. Since only positive values were included, both *agi* and *log\_agi* use the same samples. This includes all physicians in a CZ with an NPI in the 2018 NPPES who had positive *agi* from Form 1040 in 2017. *n\_agi* is weakly smaller than *n\_incper*, since it does not include individuals who had W-2 income but did not file Form 1040.

## 5 Additional Data Notes

1. All estimates have been rounded in accordance with rules established by the Census Bureau’s Disclosure Review Board.
2. *agi* has been suppressed in some rows in order to avoid disclosing information based on small implicit cells generated by differences between the construction of *incper* and *agi*.