Appendix – For Online Publication

The Rise of Healthcare Jobs

Joshua D. Gottlieb, Neale Mahoney, Kevin Rinz, and Victoria Udalova

March 2025

A Data appendix

A.1 Industry Classification

In Figure 1, we use Current Employment Statistics (CES) from the BLS to plot industry employment over time. Specifically, we plot employment for the five largest major sectors by employment and the manufacturing major sector. We make two adjustments to the baseline CES data as defined by the BLS. First, we define healthcare by removing employees in social assistance from the BLS definition of "healthcare and social assistance."[22] Second, the baseline CES series is restricted to private-sector employees (and excludes public-sector employees). For this figure, we assign employees in public hospitals, public education, and public administrative roles to the healthcare, education, and "administrative, support, and waste management" series, respectively.

The rest of the paper uses the Census and the annual ACS. We code industries based on their 1990 definition and occupations based on their 2010 definitions—the latter since 2010 is the first year with midlevels. In the ACS, we use harmonized 1990 industry codes to define healthcare and manufacturing. We define healthcare as offices of physicians, dentists, chiropractors, optometrists, and other clinicians, hospitals, nursing and care facilities, and all other health facilities. Figure B.1 shows that this definition of healthcare yields employment estimates very close to the CES. We define manufacturing as all employment codes listed under the manufacturing group in the industry 1990 classification aside from printing and publishing industries and logging, which are not considered manufacturing in the CES and NAICS.

[22] The social assistance subsector includes family services, food, housing, and relief services, vocational rehabilitation services, and child care services.

A.2 Occupation Classification

We use Census/ACS occupation codes to create our healthcare occupation categories. Because the harmonized coding scheme offered by IPUMS is based on 1990 codes, which do not break out nurse practitioners and CRNAs from nurses, we instead use codes contemporaneous to the survey year and harmonize accordingly.

Physicians are given a single code through 2018 until they are separated into physicians and surgeons. We include surgeons as physicians in those years. Nurses are defined as registered nurses (RNs) and licensed practical nurses (LPNs) throughout the sample. Until 2010, CRNAs and NPs were assigned the same code as nurses. Midlevels are defined as physician assistants, CRNAs, nurse practitioners, and nurse midwives; the latter two groups are combined by the ACS into a single category. Prior to 2010, the only distinct code in this group was physician assistants, so we omit statistics for midlevels until 2010.

Aides' codes become more granular over our sample. We use the broader grouping provided by IPUMS wherever possible. For example, in 1980, this includes the three occupation codes under "Health Service Occupations", while in 2017, this includes the 11 codes under "Health Support Occupations". The 2018–2022 codes do not include this header, so we select the codes that most closely match the ones in 2017.

"Other" clinical occupations is a catch-all category that includes all other patient-facing clinical occupation codes. For example, in 2018–2022, this includes most codes under "Healthcare Practitioners and Technical Occupations" that were not already covered by other occupation groups. There is considerable heterogeneity in this group, with dentists, physical therapists, and technicians all included. For this reason, we do not focus on this group on its own. "Non-clinical" occupations are defined as all remaining respondents with a healthcare industry code that are not matched into one of the previous groups. We cross-verify our crosswalk for aides, other clinical, and non-clinical using the 1990 codes from IPUMS to ensure we are capturing the same set of occupations each year. A complete list of codes can be found in Ruggles et al. (2024).

A.3 Restricted-Use Data

Our estimates using the Decennial Census or the ACS come from the restricted-use versions of the datasets. There are three advantages critical to our study. First, the internal files contain much larger samples. The public-use Decennial Census files include data from long-form surveys for roughly 5% of the US population. The restricted-use data contains long-form responses for 20% of the population. For ACS years, public-use files contain about two-thirds of all respondents for a given year. The internal files contain the full sample.

Second, income variables contain less income censoring (top-coding) in the internal versions. In the public-use files, income top-coding begins around the 99.5th percentile of the income distribution (McKenna and Haubach, 2019). An important part of our sample—physicians, in particular—has a large fraction of respondents censored in the public data. The restricted-use file has much less censoring. Gottlieb et al. (2023) use New York as an example to show that only 0.1% of the population and less than 0.5% of doctors have their incomes censored.

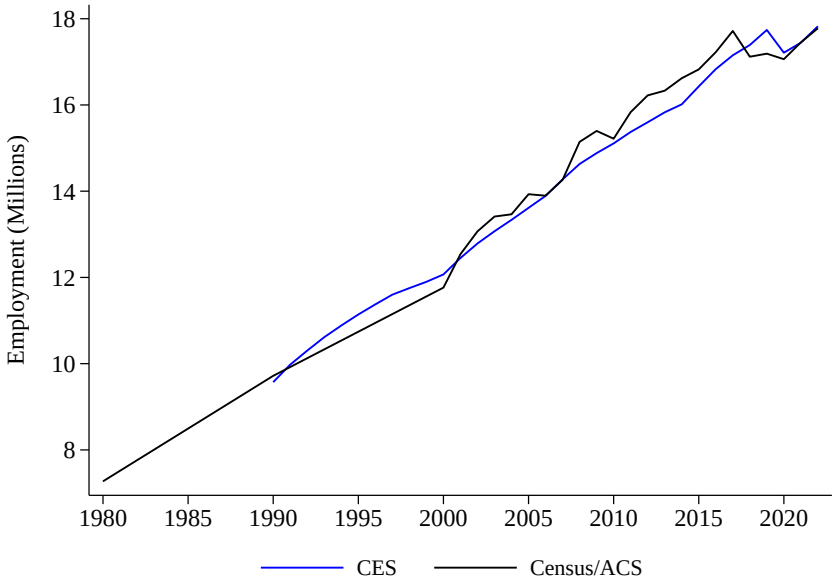
Third, we observe far more detailed geographic units in the restricted files. Most importantly, we observe counties, which allows us to conduct analysis using stable MSAs over time. Public-use files identify public-use microdata areas, which are no smaller than 100,000 people and are not consistent over time.

A.4 Calculating Healthcare's Contribution to Female Employment

In Section 4.1, we note that 3.3 million of the 13.1 million women who found employment in 2022 due to rising female employment shares did so because of the growth in healthcare. We calculate 13.1 million by taking the difference between the share of women employed in 2022 and 1980 and multiplying by the working-age female population. We calculate 3.3 million by taking the difference in healthcare's share of employment from 1980 to 2022 and multiplying by the working-age female population and the share of women employed in 2022.

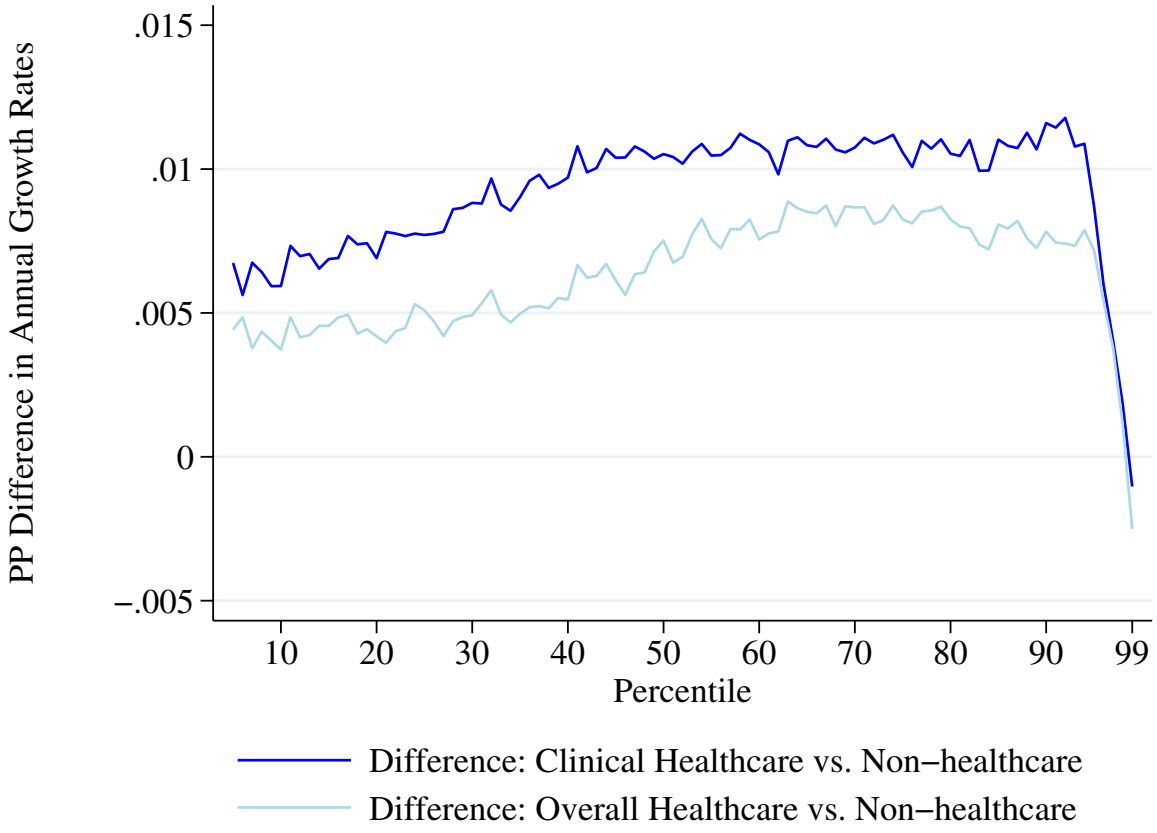
B Additional exhibits

Figure B.1: Healthcare Employment Trends by Data Source



Notes: This figure shows the time series of employment for healthcare in the United States. Data come from the 1980, 1990, and 2000 Decennial Census, the 2001–2022 annual ACS survey, and the Current Employment Statistics (CES) from 1990-2022. We group ACS occupation codes to match BLS healthcare definitions. This iteration of the CES healthcare line based on NAICS codes is only available beginning in 1990. Appendix A contains further detail. DRB number: CBDRB-FY24-0442.

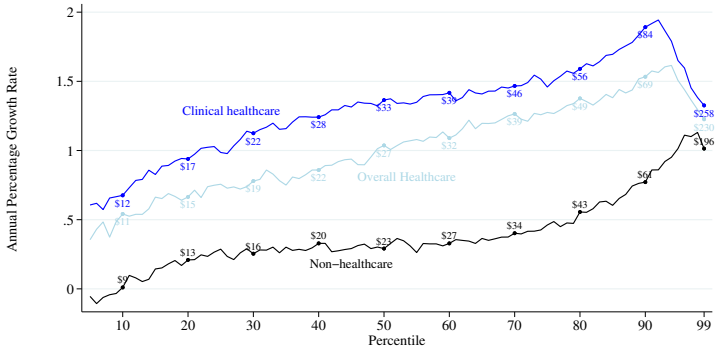
Figure B.2: Percentage Point Difference in Annual Wage Growth



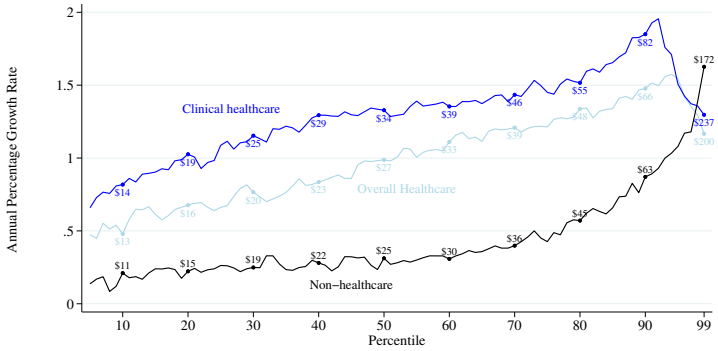
Notes: This figure shows the difference in wage growth between specified industries from 1980 to 2022 at each percentile of the distribution in Figure 2c. Data come from the 1980 Decennial Census and the 2022 American Community Survey. To compute hourly wage, we take the self-reported annual wage and divide by the usual hours worked per week multiplied by the number of weeks worked. We restrict to respondents working at least 14 weeks a year for at least 20 hours on average to minimize measurement error when computing hourly wage. We omit the bottom five percentiles where average wages are close to zero and average annual growth is noisy. We calculate real annual growth rates assuming continuously compounding. The growth rate of a variable Y from its value Y_1 in year T_1 to Y_2 in year T_2 is $\frac{(\ln(Y_2)-\ln(Y_1))}{(T_2-T_1)}$. Wages are inflation-adjusted to 2022 dollars using the CPI-U. DRB number: CBDRB-FY24-0442.

Figure B.3: Average Annual Wage Growth by Percentile

(a) All employed respondents

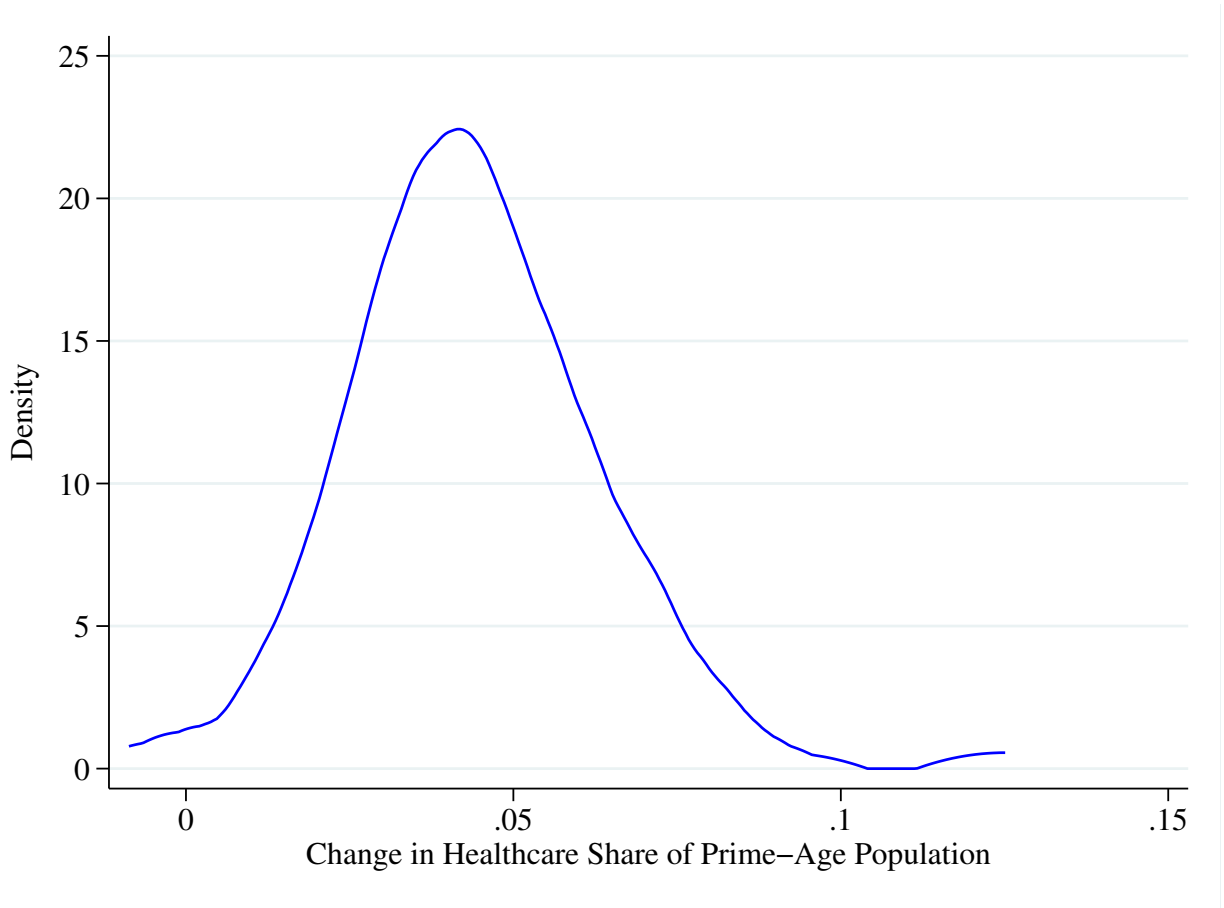


(b) Respondents working at least 48 hours and 35 weeks



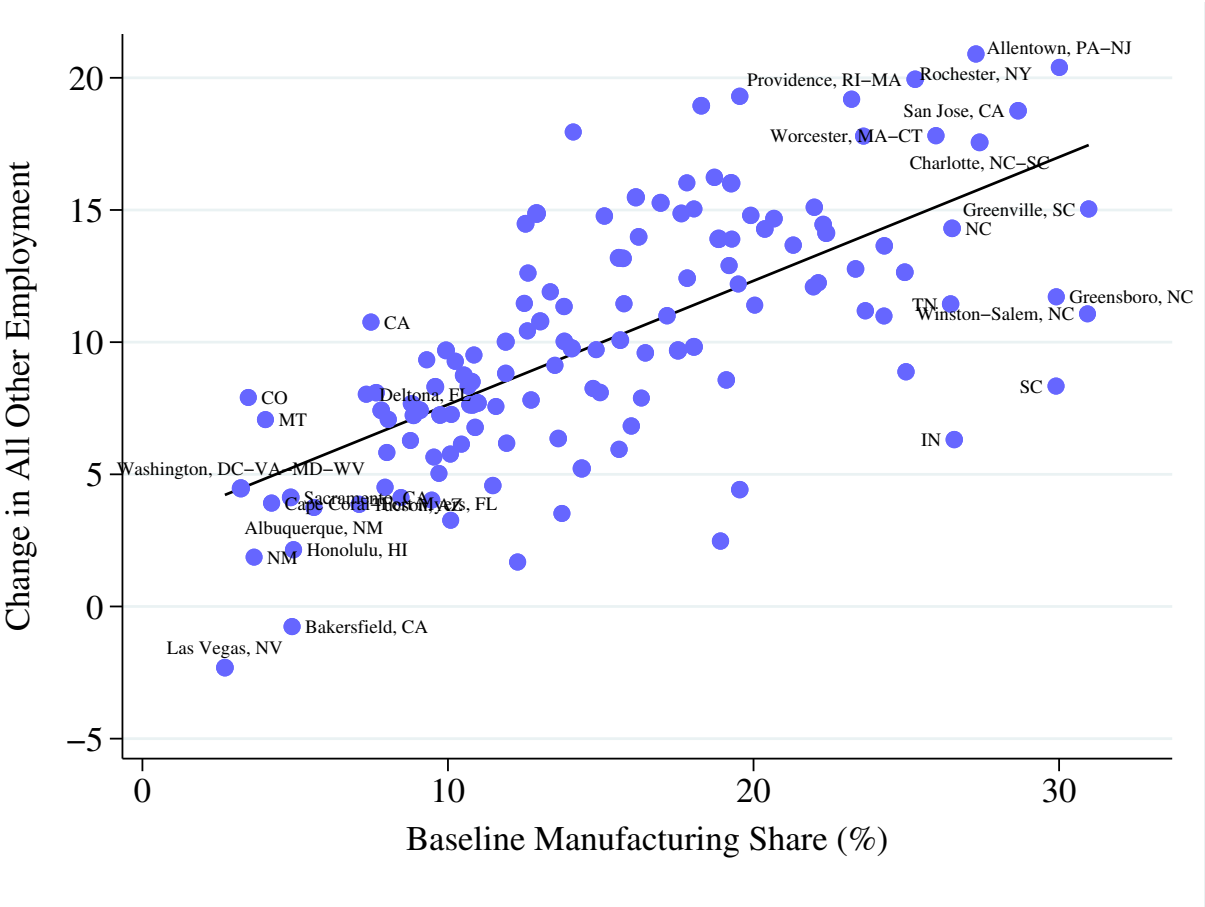
Notes: This figure shows the growth in wages by industry from 1980 to 2022 at each percentile of the wage distribution. Data come from the 1980 Decennial Census and the 2022 American Community Survey. To compute hourly wage, we take the self-reported annual wages and divide by the product of usual hours worked per week and the number of weeks worked. In Panel (a), we include all respondents. In Panel (b), we restrict to respondents working at least 48 weeks and at least 35 hours per week on average. We calculate real annual growth rates assuming continuously compounding. The growth rate of a variable Y from its value Y_1 in year T_1 to Y_2 in year T_2 is $\frac{(\ln(Y_2)-\ln(Y_1))}{(T_2-T_1)}$. Wages are inflation-adjusted to 2022 dollars using the CPI-U. DRB number: CBDRB-FY24-0442.

Figure B.4: Change in Healthcare Share of Prime Age Population, MSAs (1980–2022)



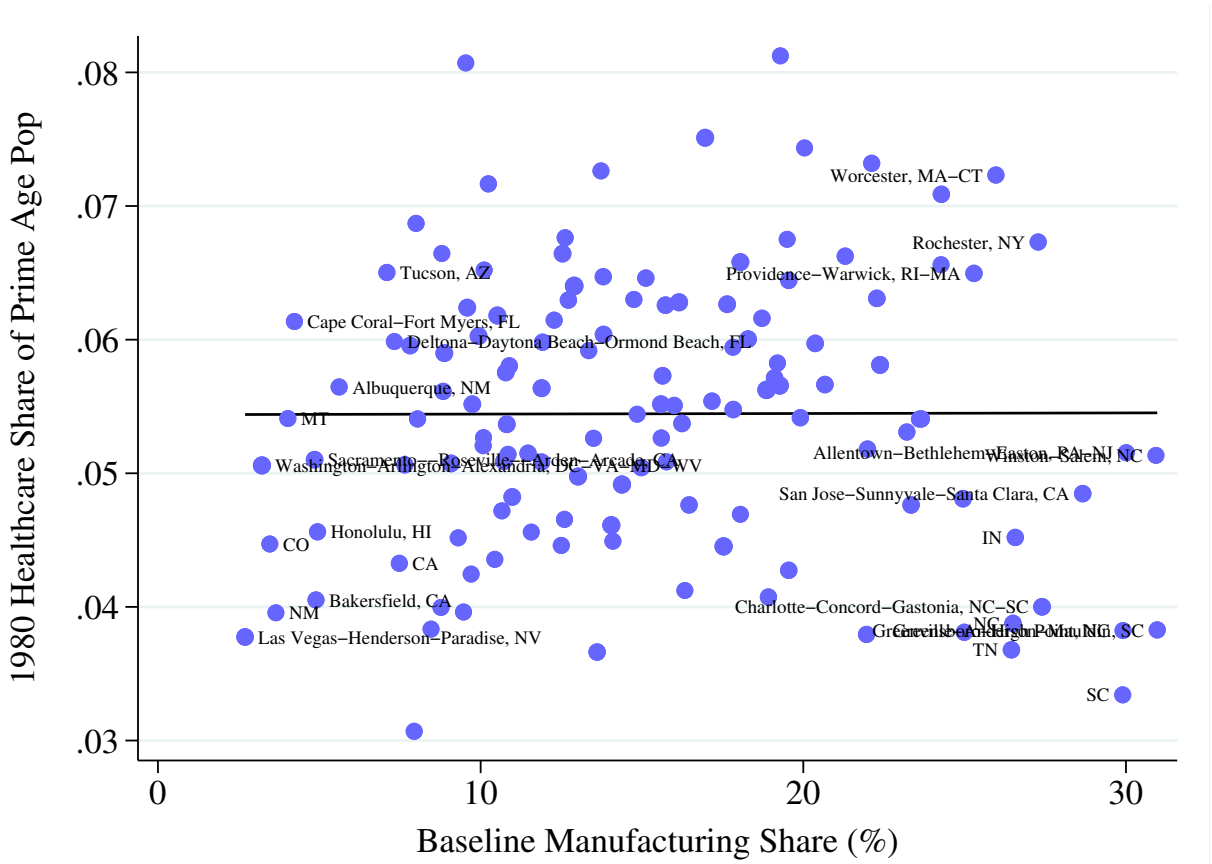
Notes: This figure shows the density of the change in the healthcare share of the prime-age population from 1980–2022 for all MSAs and non-MSA state areas. Data come from the 1980 Decennial Census and the 2022 American Community Survey. The density curve is weighted by 2022 population. Prime-age population is defined as the total population of the geography between the ages of 25 and 54. DRB number: CBDRB-FY24-0442.

Figure B.5: Change in Non-Manufacturing, Non-healthcare Employment



Notes: This figure plots the change in the non-manufacturing, non-healthcare employed share of prime-age population against the manufacturing share of the prime-age population in 1980. Data come from the 1980 Decennial Census and the 2022 American Community Survey. "Non-manufacturing" and "Non-healthcare" refers to all employed respondents in an industry other than manufacturing or healthcare. Each dot represents a 2013 MSA or the parts of a state not in an MSA; only sufficiently large MSAs to satisfy Census Bureau release requirements are plotted. Prime-age population is defined as the total population of the geography between the ages of 25 and 54. DRB number: CBDRB-FY24-0442.

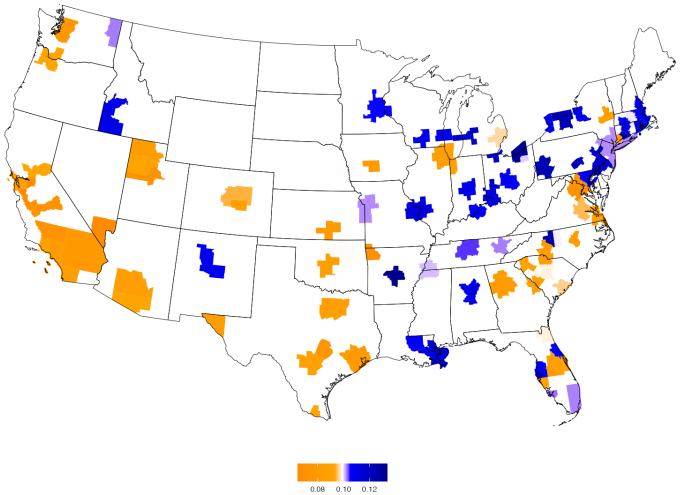
Figure B.6: 1980 Healthcare and 1980 Manufacturing Shares



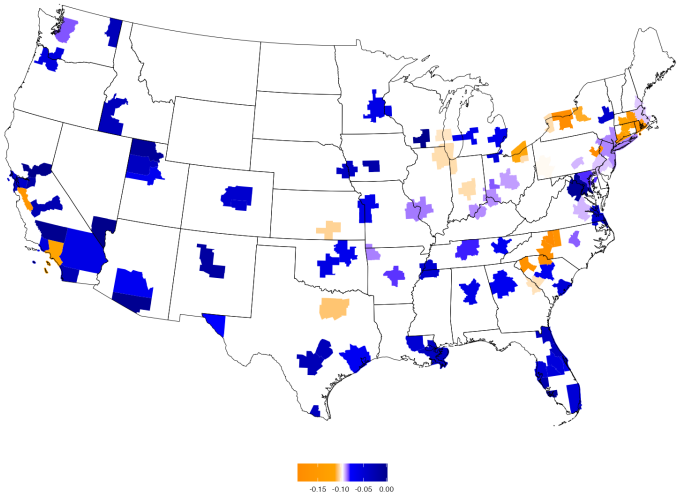
Notes: This figure plots the 1980 manufacturing share of the prime-age population against the 1980 healthcare share of the prime-age population. Data come from the 1980 Decennial Census and the 2022 American Community Survey. Each dot represents a 2013 MSA or the parts of a state not in an MSA; only sufficiently large MSAs to satisfy Census Bureau release requirements are plotted. Prime-age population is defined as the total population of the geography between the ages of 25 and 54. DRB number: CBDRB-FY24-0442.

Figure B.7: Employment Across Space (MSAs)

(a) 2022 Healthcare Employment



(b) Change in Manufacturing Share, 1980–2022



Notes: Figure (a) maps the 2022 healthcare share of employment for MSAs. Figure (b) maps the change in the manufacturing share of prime-age population from 1980–2022 for MSAs. Data come from the 1980 Decennial Census and the 2022 American Community Survey. The 95 MSAs shown are sufficiently large to be disclosable under Census Bureau rules. We exclude Honolulu for visual purposes. DRB number: CBDRB-FY24-0442.

Table B.1: Earnings Growth, Education Decomposition

	Earnings Growth (1980-2022)		
	(1)	(2)	(3)
	Actual (%)	Reweighted (%)	Reweighted / Actual
Physicians	0.74	0.74	1.000
Nurses	1.29	1.09	0.843
Aides	0.83	0.75	0.907

Notes: Data from the 1980, 1990, and 2000 Census and the 2001–2022 ACS. We compute reweighted average annual growth following DiNardo, Fortin and Lemieux (1996) by reweighing the 2022 observations so that their education distribution matches the base period education distribution, and then estimating the growth rate. We calculate real annual growth rates assuming continuous compounding. The growth rate of a variable Y from its value Y_1 in year T_1 to Y_2 in year T_2 is $\frac{(\ln(Y_2)-\ln(Y_1))}{(T_2-T_1)}$. We omit midlevels because these occupations have only emerged recently, and their job roles and educational standards are still evolving. DRB number: CBDRB-FY24-0442.

Table B.2: Regression estimates, changing employment composition

	Δ in Share of Prime-Age Population					
	Disclosable MSAs			All MSAs		
	(1)	(2)	(3)	(4)	(5)	(6)
	Manufacturing	Healthcare	All Other	Manufacturing	Healthcare	All Other
Baseline Manufacturing Share	-0.60	0.07	0.60	-0.54	0.03	0.63
	(0.04)	(0.02)	(0.08)	(0.02)	(0.01)	(0.07)
Constant	0.02	0.03	0.05	0.02	0.04	0.05

Notes: This table reports coefficients from regressing the change in the share of the prime-age population employed in selected industries against the baseline manufacturing share. Data come from the 1980 Decennial Census and the 2022 American Community Survey. Prime-age population is defined as the total population of the geography between the ages of 25 and 54. The unit of observation is a 2013 MSA or the parts of a state not in an MSA. Disclosable MSAs satisfy Census Bureau release requirements. DRB number: CBDRB-FY24-0442.

Table B.3: Change in Employment and Baseline Manufacturing Employment

	1980 Share	2011 Share	Regression: emp. chg. vs. 1980 mfg. share	$\frac{\text{Regression: health emp. chg. vs. 1980 mfg. share}}{\text{Regression: mfg. emp. chg. vs. 1980 mfg. share}}$
Panel A: All				
Manufacturing	0.160	0.080	-0.575	
Healthcare	0.056	0.084	0.058	0.100
Panel B: White Male (37%)				
Manufacturing	0.228	0.119	-0.748	
Healthcare	0.026	0.034	0.022	0.029
Panel C: White Female (37%)				
Manufacturing	0.095	0.046	-0.387	
Healthcare	0.078	0.130	0.094	0.243
Panel D: Nonwhite Male (13%)				
Manufacturing	0.206	0.095	-0.902	
Healthcare	0.039	0.044	-0.021	-0.023
Panel E: Nonwhite Female (14%)				
Manufacturing	0.114	0.048	-0.591	
Healthcare	0.115	0.135	0.052	0.088

Notes: This table relates the change in healthcare and manufacturing employment to a region's baseline manufacturing share. Data come from the 1980 Decennial Census and the 2011 American Community Survey. Healthcare and manufacturing are defined based on 1990 Census industry codes. We compute the 1980 and 2011 shares by taking the number of individuals ages 25–54 employed in manufacturing or healthcare and dividing by the total number of individuals ages 25–54. The third column reports the regression slope for the change in manufacturing or healthcare employment against a PUMA's baseline manufacturing share. The fourth column reports the ratio of the change in healthcare slope and the change in manufacturing slope. PUMAs refer to 1980–2011 "consistent" PUMAs defined by IPUMS.

Table B.4: Change in Employment and Baseline Manufacturing Employment

	1980 Share	2011 Share	Regression: emp. chg. vs. 1980 mfg. share	$\frac{\text{Regression: health. emp. chg. vs. 1980 mfg. share}}{\text{Regression: mfg. emp. chg. vs. 1980 mfg. share}}$
Panel A: All				
Manufacturing	0.160	0.080	-0.575	
Healthcare	0.056	0.084	0.058	0.100
Panel B: $\leq$ Bachelor's Degree (70%)				
Manufacturing	0.172	0.084	-0.612	
Healthcare	0.050	0.073	0.054	0.089
Panel C: $\geq$ Bachelor's Degree (30%)				
Manufacturing	0.110	0.070	-0.274	
Healthcare	0.080	0.110	0.059	0.215

Notes: This table relates the change in healthcare and manufacturing employment to a region's baseline manufacturing share. Data come from the 1980 Decennial Census and the 2011 American Community Survey. Healthcare and manufacturing are defined based on 1990 Census industry codes. We compute the 1980 and 2011 shares by taking the number of individuals ages 25–54 employed in manufacturing or healthcare and dividing by the total number of individuals ages 25–54. The third column reports the regression slope for the change in manufacturing or healthcare employment against a PUMA's baseline manufacturing share. The fourth column reports the ratio of the change in healthcare slope and the change in manufacturing slope. PUMAs refer to 1980–2011 "consistent" PUMAs defined by IPUMS.

C Theory appendix

We present a more detailed derivation of the theoretical result we discuss in Section 4.

C.1 Overview

Consider an economy with three industries: healthcare, manufacturing, and other. Let $s = \{H, M, O\}$ index those industries. We assume that each industry faces isoelastic exogenous demand

$$Q_s = Z_s P_s^{\alpha} \tag{C.1}$$

where Q_S is the industry's output, P_s the unit price of the output, Z_s is a demand shifter, and $\alpha < 0$ the demand elasticity. Each industry has a Cobb-Douglas production function of the form

$$Q_s = A_s L_s^{\beta} K_s^{1-\beta} \tag{C.2}$$

where L_s is labor, K_s is capital, and A_s is a productivity shifter. Total labor supply is inelastic at $\bar{L} = \sum_s L_s$ and workers are perfectly elastic across industries so equilibrium wages are constant and equal to the marginal revenue product of labor. Capital is also supplied perfectly elastically at a constant rental rate.

C.2 Cost minimization

With exogenous demand and perfect competition, each industry minimizes costs $wL + rK$ subject to equation (C.2). This yields a constant capital-to-labor expenditure ratio

$$\frac{K_s}{L_s} = \frac{w}{r}\frac{1-\beta}{\beta}. \tag{C.3}$$

Substituting this into the production function (C.2) and solving for L_s in terms of Q yields

$$L_s = \frac{Q_s}{A_s}\beta^{1-\beta}\left(1-\beta\right)^{-(1-\beta)} w^{-(1-\beta)} r^{1-\beta}. \tag{C.4}$$

We now solve for cost. Note that $wL + rK = wL\left[1 + \frac{rK}{wL}\right]$. Using (C.3), we can simplify and solve for cost

$$C(Q) = \frac{Q_s}{A_s}\beta^{-\beta}(1-\beta)^{-(1-\beta)}w^{\beta}r^{1-\beta}. \tag{C.5}$$

Perfect competition implies that the industry's output price is equal to the marginal cost $C'(Q)$:

$$P = \frac{1}{A_s}\beta^{-\beta}(1-\beta)^{-(1-\beta)}w^{\beta}r^{1-\beta}. \tag{C.6}$$

Substituting this into the demand equation (C.1) yields

$$Q_s = Z_s\left(\frac{1}{A_s}\beta^{-\beta}(1-\beta)^{-(1-\beta)}w^{\beta}r^{1-\beta}\right)^{\alpha}. \tag{C.7}$$

For simplicity, we collect constants by defining

$$\lambda = \beta^{-\beta}(1-\beta)^{-(1-\beta)}r^{1-\beta} \tag{C.8}$$

which allows us to express demand in terms of wages and productivity:

$$Q_s = Z_s\left(\frac{w^{\beta}\lambda}{A_s}\right)^{\alpha}. \tag{C.9}$$

C.3 Labor demand and labor shares

To compute labor demand for industry s, we use the constant expenditure share $wL_s = \beta C(Q_s)$. We substitute our expression for the cost function, solve for L_s, and substitute for Q_s to obtain:

$$L_s = \frac{\beta\lambda w^{\beta-1}Q_s}{A_s} \tag{C.10}$$

$$= \frac{\beta w^{\beta-1}\lambda}{A_s}\left[Z_s\left(\frac{w^{\beta}\lambda}{A_s}\right)^{\alpha}\right] \tag{C.11}$$

$$= \beta w^{\beta(\alpha+1)-1}\lambda^{\alpha+1}\left[Z_s A_s^{-(\alpha+1)}\right]. \tag{C.12}$$

Recall the adding-up constraint $\sum_s L_s = \overline{L}$. Healthcare's share of labor is thus

$$S_H = \frac{L_H}{\overline{L}} = \frac{Z_H A_H^{-(\alpha+1)}}{\sum_s Z_s A_s^{-(\alpha+1)}} \tag{C.13}$$

C.4 Reallocation following demand shock

We ask how healthcare and "other" (non-healthcare, non-manufacturing) labor demand changes following a manufacturing shock.

Suppose there is a demand shock to manufacturing (via the demand shifter).[23] Then we have

$$\frac{\partial S_H}{\partial Z_M} = -\frac{A_M^{-(\alpha+1)} Z_H A_H^{-(\alpha+1)}}{\left(\sum_s Z_s A_s^{-(\alpha+1)}\right)^2} \tag{C.14}$$

and

$$\frac{\partial S_O}{\partial Z_M} = -\frac{A_M^{-(\alpha+1)} Z_O A_O^{-(\alpha+1)}}{\left(\sum_s Z_s A_s^{-(\alpha+1)}\right)^2} \tag{C.15}$$

A positive manufacturing demand shock decreases healthcare and other's shares of the labor force. To understand how these changes relate to baseline shares, we scale the changes in equations (C.14) and (C.15) by the industries' respective baseline shares:

$$\frac{\frac{\partial S_H}{\partial Z_M}}{S_H} = \frac{A_M^{-(\alpha+1)} Z_H A_H^{-(\alpha+1)}}{\left(\sum_s Z_s A_s^{-(\alpha+1)}\right)^2} \cdot \frac{\sum_s Z_s A_s^{-(\alpha+1)}}{Z_H A_H^{-(\alpha+1)}} = \frac{A_M^{-(\alpha+1)}}{\sum_s Z_s A_s^{-(\alpha+1)}} \tag{C.16}$$

$$\frac{\frac{\partial S_O}{\partial Z_M}}{S_O} = \frac{A_M^{-(\alpha+1)} Z_O A_O^{-(\alpha+1)}}{\left(\sum_s Z_s A_s^{-(\alpha+1)}\right)^2} \cdot \frac{\sum_s Z_s A_s^{-(\alpha+1)}}{Z_O A_O^{-(\alpha+1)}} = \frac{A_M^{-(\alpha+1)}}{\sum_s Z_s A_s^{-(\alpha+1)}} \tag{C.17}$$

These two quantities are identical. This means that each non-manufacturing industry has the same proportional response to a shock to manufacturing demand. This matches the results of our statistical analysis empirically connecting manufacturing job losses to healthcare job gains.

[23] The conclusion is identical when there is a shock to manufacturing productivity.