TABLE OF CONTENTS

A. List of Medicaid expansion states and expansion dates
B. Conditional logit regression model
C. Social disadvantage index
D. Annual sample sizes of general internists
E. Full regression results for main analyses, including tests of zero trends in the pre-expansion period
F. Social disadvantage categories: hexiles or terciles?
G. Sensitivity analyses
H. Detailed explanation of simulation methods and full simulation results
I. References
A. List of Medicaid expansion states and expansion dates.

Table A.1. List of Medicaid expansion states and expansion dates.¹

| State              | Implementation date of Medicaid expansion |
|--------------------|--------------------------------------------|
| Alaska             | 9/1/2015                                   |
| Arizona            | 1/1/2014                                   |
| Arkansas           | 1/1/2014                                   |
| California         | 1/1/2014                                   |
| Colorado           | 1/1/2014                                   |
| Connecticut        | 1/1/2014                                   |
| Delaware           | 1/1/2014                                   |
| District of Columbia | 1/1/2014                                |
| Hawaii             | 1/1/2014                                   |
| Illinois           | 1/1/2014                                   |
| Indiana            | 2/1/2015                                   |
| Iowa               | 1/1/2014                                   |
| Kentucky           | 1/1/2014                                   |
| Louisiana          | 7/1/2016                                   |
| Maine              | 1/10/2019                                  |
| Maryland           | 1/1/2014                                   |
| Massachusetts      | 1/1/2014                                   |
| Michigan           | 4/1/2014                                   |
| Minnesota          | 1/1/2014                                   |
| Montana            | 1/1/2016                                   |
| Nevada             | 1/1/2014                                   |
| New Hampshire      | 8/15/2014                                  |
| New Jersey         | 1/1/2014                                   |
| New Mexico         | 1/1/2014                                   |
| New York           | 1/1/2014                                   |
| North Dakota       | 1/1/2014                                   |
| Ohio               | 1/1/2014                                   |
| Oregon             | 1/1/2014                                   |
| Pennsylvania       | 1/1/2015                                   |
| Rhode Island       | 1/1/2014                                   |
| Vermont            | 1/1/2014                                   |
| Virginia           | 1/1/2019                                   |
| Washington         | 1/1/2014                                   |
| West Virginia      | 1/1/2014                                   |
B. Conditional logit regression model.

Our analytic goal was to assess whether the probability that new physicians located in each category of choice area changed after implementation of the Medicaid expansion, compared with the pre-expansion period. We based our analyses on the conditional logit regression model, which was developed specifically to analyze situations in which subjects make a single choice (a choice area, in our application) from a set of alternatives.\(^2\) In estimating the regression coefficients, the model explicitly accounts for the characteristics of the alternatives rejected as well as the one chosen, thereby capturing a fundamental aspect of real-life decision making. The model is fit to data that includes information on each choice area for each physician. An indicator variable identifies the choice area chosen by each physician. The estimated coefficients can be used to predict the probability that each physician locates in each choice area. These probabilities can be summed across choice areas as desired.

The basic assumptions underlying the conditional logit model are as follows:

- Individual \(i\) makes a single choice among \(C\) alternatives
- The utility, or value, of alternative \(c\) to individual \(i\) in calendar year \(t\) is \(U_{ict}\), which can be written as:
  \[ U_{ict} = \mu_{ict} + \varepsilon_{ict} \]
  where \(\mu_{ict}\) is a systematic component and \(\varepsilon_{ict}\) is a random error term with a Type I extreme value distribution

Then the probability, \(P_{irt}\), that individual \(i\) chooses alternative \(r\) in year \(t\) is given by equation (1):

\[ P_{irt} = \frac{\exp (\mu_{irt})}{\sum_{c \in C} \exp (\mu_{ict})} \]

Conditional logit regression enables us to model the systematic component of utility, \(\mu_{ict}\), as a function of characteristics of the alternatives and interactions of individual characteristics and characteristics of the alternatives.

In our study, we fit the conditional logit model using a dataset where each physician had an observation for each of the 764 choice areas. The dependent variable was an indicator for choice area chosen by the physician. The independent variables—i.e., the determinants of the systematic component of utility—were four indicators of the physician’s prior contact with each choice area (whether the choice area was in the physician’s state of birth, medical school, or GME and whether the physician completed GME in the commuting zone [CZ]), state fixed effects, CZ-level sociodemographic variables (see manuscript), main effects (indicators) for the disadvantage level of the CZ, indicators for the urban-rural category of the CZ, indicators for whether the CZ had an allopathic medical school and an osteopathic medical school, an indicator for continuation of the primary care fee bump, the GPCI, and 5 vectors of interaction terms constructed by interacting indicator variables for calendar years 2009-2019 with indicators for 5 categories of choice areas: (1) high disadvantage choice areas in expansion states, (2) medium disadvantage choice areas in expansion states, (3) low disadvantage choice areas in expansion states, (4) high disadvantage choice areas in non-expansion states, and (5) medium disadvantage choice areas in non-expansion states. The sixth category, low disadvantage choice areas in non-expansion states, was the “reference category,” meaning that the interaction terms were always zero for these choice areas.

In mathematical terms, we modelled the systematic component of utility, \(\mu_{ist}\), as follows:
\[ \mu_{ict} = \alpha_1 \text{SameBrthSt}_{ic} + \alpha_2 \text{SameMedSchSt}_{ic} + \alpha_3 \text{SameGMESt}_{ic} + \alpha_4 \text{SameGMECZ}_{ic} + \beta_1 \text{SocDem}_{ct} + \beta_2 \text{UrbRur}_c + \beta_3 \text{MedSch}_{ct} + \beta_4 \text{FeeBump}_{ct} + \beta_5 \text{GPCI}_c + \gamma_s + \mu_1 \text{HighDis}_{ct} + \mu_2 \text{MedDis}_{ct} \]
\[ + \sum_{u=2009}^{2019} \varphi_{1u} \text{Exp}_c \times \text{HighDis}_{cu} \times Y(u = t) \]
\[ + \sum_{u=2009}^{2019} \varphi_{2u} \text{Exp}_c \times \text{MedDis}_{cu} \times Y(u = t) + \sum_{u=2009}^{2019} \varphi_{3u} \text{Exp}_c \times \text{LowDis}_{cu} \times Y(u = t) \]
\[ + \sum_{u=2009}^{2019} \varphi_{4u} \text{NExp}_c \times \text{HighDis}_{cu} \times Y(u = t) + \sum_{u=2009}^{2019} \varphi_{5u} \text{NExp}_c \times \text{MedDis}_{cu} \times Y(u = t) \]

- \text{SameBrthSt}_{ic}, \text{SameMedSchSt}_{ic}, \text{and SameGMESt}_{ic} are, respectively, indicators for whether physician \( i \) was born, attended medical school, or completed GME in the state where choice area \( c \) is located.
- \text{SameGMECZ}_{ic} is an indicator for whether physician \( i \) was completed GME in the commuting zone where choice area \( c \) is located.
- \text{SocDem}_{ct} are sociodemographic characteristics of the commuting zone where choice area \( c \) is located in calendar year \( t \).
- \text{UrbRur}_c is a set of indicators for the urban-rural category of the commuting zone where choice area \( c \) is located.
- \text{MedSch}_{ct} are two indicators for whether the commuting zone where choice area \( c \) is located has an allopathic medical school and an osteopathic medical school.
- \text{FeeBump}_{ct} is an indicator for whether the state where choice area \( c \) is located continued the Medicaid fee bump in calendar year \( t \) (starting in 2015).
- \text{GPCI}_c is the geographic practice cost index in choice area \( c \).
- \( \gamma_s \) are state fixed effects.
- \text{HighDis}_{ct}, \text{MedDis}_{ct}, \text{and LowDis}_{ct} are indicators for whether the disadvantage level of the commuting zone where choice area \( c \) is located is high, medium, or low, respectively, in calendar year \( t \).
- \text{Exp}_c is an indicator variable for whether the state where choice area \( c \) is located was an expansion state.
- \text{NExp}_c is an indicator variable for whether the state where choice area \( c \) is located was a non-expansion state.
- \( Y(u = t) \) is an indicator variable that takes on the value “1” when the argument in parentheses is satisfied (that is, when \( u = t \)) and is “0” otherwise (these are the calendar year indicators).
The \( \varphi \)'s are the coefficients of main interest. The five summation terms containing the \( \varphi \)'s in the last three lines of the equation are the 5 vectors of interaction terms described in the Materials and Methods section of the manuscript. They are constructed by interacting indicator variables for calendar years 2009-2019 with indicators for 5 categories of choice areas: (1) high disadvantage choice areas in expansion states, (2) medium disadvantage choice areas in expansion states, (3) low disadvantage choice areas in expansion states, (4) high disadvantage choice areas in non-expansion states, and (5) medium disadvantage choice areas in non-expansion states. The sixth category, low disadvantage choice areas in non-expansion states, is the “reference category” in the model, so a vector of interaction terms does not appear for this category. We consider 2009-2013 the pre-expansion period and 2014-2019 the post-expansion period.

An additional point is that, to make the model estimable, some constraint must be placed on the year coefficients for each category of choice areas. Otherwise the model would not be estimable due to collinearity. In many cases investigators opt to omit the last pre-intervention year from the model, which is equivalent to constraining its coefficient to equal zero. The potential shortcoming of this approach, however, is that every other year in the analysis is thereby compared to a single year—the omitted year—which may be idiosyncratically high or low if there is random variation in the data. In our study, rather than omitting a particular year, we imposed the constraints that the average of the coefficients of the indicators for the 5 pre-expansion years (2009-2013) equal zero for each category of choice areas. This approach helps smooth out random year-to-year variation in the data.

In this setup, the \( \varphi \)'s can be used to assess whether there were any pre-expansion trends toward increasing or declining probabilities that new general internists located in any of the 6 categories of choice areas. The results of this analysis are summarized in the Results section of the manuscript and more fully reported in Appendix E.

Further, the anti-logarithm of the difference of the estimated coefficients for 2 different categories of choice areas in the same year approximates the ratio of the probabilities that, other things equal, a physician located in those 2 categories in the particular year relative to the average ratio of the same probabilities during the pre-expansion period. We refer to this quantity as the relative rate ratio (RRR). As an example, consider the anti-logarithm of the difference in the coefficients for high disadvantage choice areas in expansion states and non-expansion states in 2014, i.e., \( \varphi_{1,2014} - \varphi_{4,2014} \). An RRR=1.0 (>1.0) would mean that in 2014 physicians were equally likely (more likely) to locate in high disadvantage choice areas in expansion states relative to high disadvantage choice areas in non-expansion states than they were, on average, during the pre-expansion period. Figure 2 in the manuscript presents the findings of this type of analysis for the three disadvantage levels.

C. Social disadvantage index.

As discussed in the manuscript, we developed a social disadvantage index at the commuting zone (CZ) level using four sociodemographic variables: percentage of the population with less than a high school diploma, percentage with a bachelor’s degree or more, percentage poor, and per capita income. We first standardized each variable, using the mean and standard deviation across CZs, to calculate a z-score for each CZ. We then added the z-scores for the four variables.

We assessed the reliability, or internal consistency, of the social disadvantage index using Crohnbach’s alpha. Reliability in this context refers to the extent to which the index is a consistent measure of a concept, and Cronbach’s alpha is one way of measuring the strength of that consistency. We obtained the following results for the three years that we checked:
Table C.1. Cronbach’s alpha.

| Year | Crohnbach’s alpha |
|------|-------------------|
| 2009 | 0.885             |
| 2014 | 0.877             |
| 2019 | 0.880             |

These are strong values of Crohnbach’s alpha without being so high as to suggest that that some of the variables in the index may be redundant.3

We also assessed the consistency of the index values across time by calculating the Pearson correlation coefficient across CZs for every pair of years. The correlation matrix is shown below:

Table C.2. Correlation coefficients.

| Year | 2009   | 2020 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|------|--------|------|------|------|------|------|------|------|------|------|------|
| 2009 | 1.000  | 0.994| 0.989| 0.982| 0.976| 0.970| 0.967| 0.966| 0.964| 0.964| 0.961|
| 2010 | 1.000  | 0.995| 0.989| 0.982| 0.977| 0.971| 0.970| 0.968| 0.969| 0.964|      |
| 2011 | 1.000  | 0.996| 0.991| 0.987| 0.982| 0.979| 0.977| 0.977| 0.977| 0.972|      |
| 2012 | 1.000  | 0.996| 0.993| 0.989| 0.985| 0.980| 0.980|      |      |      |      |
| 2013 | 1.000  | 0.997| 0.994| 0.990| 0.986| 0.982|      |      |      |      |      |
| 2014 | 1.000  | 0.996| 0.993| 0.987| 0.985| 0.977|      |      |      |      |      |
| 2015 | 1.000  | 0.996| 0.993| 0.992| 0.988| 0.981|      |      |      |      |      |
| 2016 | 1.000  | 0.996| 0.993| 0.987|      |      |      |      |      |      |      |
| 2017 | 1.000  | 0.996| 0.992|      |      |      |      |      |      |      |      |
| 2018 | 1.000  | 0.996|      |      |      |      |      |      |      |      |      |
| 2019 | 1.000  |      |      |      |      |      |      |      |      |      |      |

As noted in the manuscript, for analysis we grouped CZs into terciles based on their social disadvantage index values and labeled the terciles high disadvantage, medium disadvantage, and low disadvantage. Consequently, we also assessed the degree of separation in each of the four sociodemographic component variables across the three terciles of the social disadvantage index. The figure shows the mean and standard deviation for each of the four component variables in each of the disadvantage terciles. As anticipated (and desired), the separation in the means of each variable across the terciles is considerable.
In Appendix F, we report the analyses that led us to choose terciles of the index after considering hexiles as a more granular approach.

D. Annual sample sizes of general internists.

Table D.1. Annual sample sizes.

| Year | General internal medicine |
|------|---------------------------|
| 2009 | 2820                      |
| 2010 | 2656                      |
| 2011 | 2564                      |
| 2012 | 2555                      |
| 2013 | 2714                      |
| 2014 | 2668                      |
| 2015 | 2912                      |
| 2016 | 3090                      |
| 2017 | 3268                      |
| 2018 | 3378                      |
| 2019 | 3477                      |
| Total| 32102                     |
E. Full regression results for the main analyses, including tests of zero trends in the pre-expansion period.

Table E.1. below presents the results of the main analyses reported in the manuscript. These analyses employed the specification described in the manuscript and explained in greater detail in Section B above. The table reports estimated coefficients, with their standard errors in parentheses just below the coefficients. (The coefficients of the state fixed effects are omitted in the interest of space.) The relative rate ratios (RRRs) shown in Figure 2 of the manuscript are obtained by taking the anti-logarithm of the difference in the coefficients of the expansion indicator for a particular disadvantage level and year and the non-expansion indicator for the same disadvantage level and year.

Table E.1.

| Independent variable | Coefficient (Standard error) | Independent variable | Coefficient (Standard error) |
|----------------------|------------------------------|----------------------|------------------------------|
| Birth state          | 1.578** (0.0428)             | Medium dis*2011*non-expansion | -0.118 (0.0838)             |
| Med sch state        | 1.543** (0.0398)             | High dis*2012*expansion    | 0.206 (0.122)               |
| GME state            | 2.022** (0.0243)             | Medium dis*2012*expansion | -0.00747 (0.0739)          |
| GME commuting zone   | 2.725** (0.0243)             | Low dis*2012*expansion    | 0.0225 (0.0596)             |
| Continue Medicaid fee bump | 0.129* (0.0556)           | High dis*2012*non-expansion | 0.116 (0.0958)             |
| Medium metro area    | 0.144** (0.0260)             | Medium dis*2012*non-expansion | 0.109 (0.0816)             |
| Small metro area     | 0.257** (0.0383)             | High dis*2013*expansion    | -0.0827 (0.128)             |
| Nonmetro area        | -0.220** (0.0544)            | Medium dis*2013*expansion | 0.0356 (0.0698)             |
| Total population (log) | 0.885** (0.0118)            | Low dis*2013*expansion    | 0.00574 (0.0571)            |
| Percentage poor      | 0.0225 (0.704)               | High dis*2013*non-expansion | -0.0243 (0.0960)           |
| Per capita income    | 1.40e-06 (6.50e-06)          | Medium dis*2013*non-expansion | -0.0272 (0.0800)           |
| Percentage less than high school | -1.636** (0.579)        | High dis*2014*expansion    | 0.267** (0.135)             |
| Percentage college or more | 0.0412 (0.444)            | Medium dis*2014*expansion | 0.0936 (0.0871)             |
| Percentage foreign born | 0.152 (0.80)               | Low dis*2014*expansion    | 0.0869 (0.80)               |
| Percentage black | -0.388* (0.164) | High dis*2014*non-expansion | -0.331* (0.137) |
|------------------|------------------|-----------------------------|------------------|
| Percentage Hispanic | -0.0155 (0.191) | Medium dis*2014*non-expansion | 0.245** (0.0926) |
| Percentage other minority | 0.692** (0.261) | High dis*2015*expansion | 0.0538 (0.141) |
| Percentage female | 5.222** (1.551) | Medium dis*2015*expansion | -0.0385 (0.0893) |
| Percentage < 5 years old | 3.515 (2.721) | Low dis*2015*expansion | 0.111 (0.0689) |
| Percentage 5-17 years old | 0.633 (1.294) | High dis*2015*non-expansion | -0.235 (0.126) |
| Percentage > 64 years old | -0.966 (0.699) | Medium dis*2015*non-expansion | 0.15 (0.0891) |
| Percentage insurance not Medicaid | 0.734 (0.381) | High dis*2016*expansion | 0.0926 (0.133) |
| GPCI | -4.519** (0.668) | Medium dis*2016*expansion | 0.0761 (0.0829) |
| Has allopathic medical school | 0.133** (0.0238) | Low dis*2016*expansion | 0.0412 (0.0649) |
| Has osteopathic medical school | -0.00954 (0.0227) | High dis*2016*non-expansion | -0.102 (0.114) |
| High disadvantage (main effect) | 0.0324 (0.0578) | Medium dis*2016*non-expansion | 0.0448 (0.0881) |
| Medium disadvantage (main effect) | -0.0987** (0.0315) | High dis*2017*expansion | 0.208 (0.128) |
| High dis*2009*expansion | -0.0285 (0.124) | Medium dis*2017*expansion | 0.200* (0.0823) |
| Medium dis*2009*expansion | -0.0657 (0.0707) | Low dis*2017*expansion | 0.165* (0.0662) |
| Low dis*2009*expansion | 0.0136 (0.0562) | High dis*2017*non-expansion | -0.275* (0.123) |
| High dis*2009*non-expansion | -0.118 (0.0967) | Medium dis*2017*non-expansion | 0.147 (0.0825) |
| Medium dis*2009*non-expansion | -0.0685 (0.0809) | High dis*2018*expansion | 0.0412 (0.131) |
| High dis*2010*expansion | -0.177 (0.136) | Medium dis*2018*expansion | 0.178* (0.0817) |
| High dis, expansion states | Medium dis, expansion states | Low dis, non-exp states | High dis, non-exp states | Medium dis, non-exp states |
|---------------------------|-----------------------------|------------------------|-------------------------|-------------------------|
| Chi sq                    | 4.35                        | 1.05                   | 1.51                    | 4.04                    | 4.96                    |
| P value                   | 0.36                        | 0.90                   | 0.83                    | 0.40                    | 0.29                    |

Note: Low disadvantage choice areas in non-expansion states comprise the reference category.

*P<.05   **P<.01

A key issue in our analysis was to assess whether there were any trends during the pre-expansion period (2009-2013) toward increasing or declining probabilities that new general internists located in any of the 6 categories of choice areas. Therefore, we conducted tests of the null hypothesis that the 5 pre-expansion coefficients were simultaneously equal to zero for each category of choice areas. (These are chi-square tests with 4 degrees of freedom, rather than 5, because of the constraint that the sum of the pre-expansion coefficients equal zero.) The findings are below:

None of the tests approached significance, indicating that the pre-expansion trends were parallel and there was no trend toward increasing or declining probabilities that new general internists located in any of the 6 categories of choice areas.
F. Social disadvantage categories: hexiles or terciles?

As noted in the manuscript, in our analyses we grouped CZs into terciles based on their social disadvantage index values and labeled the terciles high disadvantage, medium disadvantage, and low disadvantage. We chose terciles after considering hexiles as a more granular approach and determining that pairs of adjacent hexiles could be combined. The analyses that led to our conclusion are shown in Table F.1. Specifically, the table reports the difference between expansion and non-expansion states in the estimated coefficient of the indicator variable for the 6 post-expansion years (see manuscript), by disadvantage level hexile. Hexile 1 is the most disadvantaged hexile, whereas hexile 6 is the least disadvantaged.

Table F.1. Results for hexiles.

| Hexile | Difference in coefficient, expansion and non-expansion states |
|--------|-------------------------------------------------------------|
| Hexile 1 | 0.544**                                                     |
| Hexile 2 | 0.352**                                                     |
| Hexile 3 | 0.076                                                       |
| Hexile 4 | 0.029                                                       |
| Hexile 5 | 0.127**                                                     |
| Hexile 6 | 0.103*                                                      |

*P<.05. ** P<.01.

As shown, the coefficients of hexiles 1 and 2 are roughly similar in magnitude, as are the coefficients of hexiles 3 and 4 and the coefficients of hexiles 5 and 6. Additional analyses found that the coefficients for hexiles 1 and 2 were not significantly different from each other (P=.134), whereas each of these coefficients was highly significantly different from each of the other four coefficients. There were no significant differences in pairwise comparisons of the coefficients of hexiles 3, 4, 5, and 6. Of note, our analysis suggests that we could have combined hexiles 3 through 6 to create a single category, but we opted instead to preserve the tercile structure.

G. Sensitivity analyses.

As discussed in the Materials and Methods section of the manuscript, we used our conditional logit model to conduct 4 sensitivity analyses: (1) using data for 2007–2019 to extend the pre-expansion period, (2) adding indicators for caps on punitive and noneconomic damages and for full nurse practitioner scope of practice to the models, (3) using the disadvantage levels in 2014 for all years rather than allowing the disadvantage level to vary across years, and (4) limiting the analyses to states that either expanded in 2014 or never expanded during the study period. These models used a single indicator variable for the 6 years following the Medicaid expansion (see manuscript).
Table G.1 reports the difference between expansion and non-expansion states in the estimated coefficient of the indicator variable for the 6 post-expansion years, by disadvantage level, for the main analysis and for each of the 4 sensitivity analyses. (See manuscript for additional details.) As shown, the differences in the estimated coefficients for each disadvantage level are similar in magnitude across analyses.

### Table G.1. Results of sensitivity analyses.

| Sensitivity analysis          | High disadv, difference expansion and non-expansion | Medium disadv, difference expansion and non-expansion | Low disadv, difference expansion and non-expansion |
|-----------------------------|---------------------------------------------------|------------------------------------------------------|--------------------------------------------------|
| Main analysis               | 0.412**                                           | 0.048                                                | 0.113**                                          |
| Sensitivity analysis 1      | 0.384**                                           | 0.017                                                | 0.089**                                          |
| Sensitivity analysis 2      | 0.400**                                           | 0.038                                                | 0.108**                                          |
| Sensitivity analysis 3      | 0.406**                                           | 0.019                                                | 0.119**                                          |
| Sensitivity analysis 4      | 0.508**                                           | 0.046                                                | 0.134**                                          |

*P<.05.  ** P<.01.

H. Detailed explanation of simulation methods and full simulation results.

Our regression coefficient estimates enabled us to predict the probability that each physician located in each choice area in each year from 2014 through 2019 as a function of the state fixed effects, the physician’s previous ties to each choice area (i.e., indicator variables for whether the physician was born, attended medical school, and completed GME in the state and an indicator for whether the physician completed GME in the choice area), an indicator for continuing the Medicaid fee bump, the CZ-level sociodemographic variables, main effects (indicators) for the disadvantage levels, indicators for the urban-rural categories, indicators for whether the CZ had an allopathic medical school and an osteopathic medical school, the GPCI, and 5 vectors of interaction terms constructed by interacting indicator variables for calendar years 2009-2019 with indicators for 5 categories of choice areas.

To carry out the simulations, we first predicted the probability that each physician located in each choice area in each post-expansion year (2014-2019) using the actual values of the independent variables for each choice area. We summed the probabilities across physicians to obtain the predicted number of physicians who located in each choice area in each year. We then summed the predicted numbers across choice areas grouped by disadvantage level and whether the state was an expansion state. This is the scenario we refer to as the “observed patterns of Medicaid expansion across the states” in the text of the manuscript. Naturally, these predicted numbers reproduce the observed numbers, but it is important to do it this way so we can obtain standard errors for the numbers of physicians gained or lost.

Next, we repeated the same calculations, but this time we treated every state as an expansion state in the calculations (i.e., we “switched on” the indicator for Medicaid expansion states to “1” for every state). This is the scenario we refer to as the “a hypothetical alternative in which all states expanded Medicaid.”
Finally, we calculated the differences in the predicted number of physicians locating in choice areas in actual expansion states and non-expansion states, grouped by disadvantage level, under the first scenario and under the second scenario. These differences represent the predicted number of physicians gained or lost.

Figure 3 in the manuscript reports the findings of the analysis of physicians gained or lost. Table H.1 reports full simulation results, including the number of physicians predicted to locate in each category of choice areas under the two scenarios. The numbers in the “Difference” column are those presented in Figure 3 in the manuscript. We obtained standard errors and 95% confidence intervals for the values in the “Difference column” (also shown in Table H.1 and Figure 3) using the delta method.

Table H.1.

| State expansion status | Social disadvantage category | Observed patterns of Medicaid expansion | Hypothetical alternative | Difference | 95% CI for Difference |
|------------------------|-----------------------------|----------------------------------------|--------------------------|------------|----------------------|
| Expansion              | High                        | 526                                    | 507                      | 19         | (10, 27)             |
|                        | Medium                      | 2753                                   | 2678                     | 75         | (40, 109)            |
|                        | Low                         | 10469                                  | 10191                    | 278        | (153, 404)           |
| Non-expansion          | High                        | 535                                    | 767                      | -232       | (-315, -149)         |
|                        | Medium                      | 1887                                   | 1883                     | 4          | (-102, 110)          |
|                        | Low                         | 2623                                   | 2767                     | -144       | (-262, -25)          |

I. References.

1. Status of State Action on the Medicaid Expansion Decisions: Interactive Map. The Henry J. Kaiser Family Foundation, 2020. [https://www.kff.org/health-reform/state-indicator/state-activity-around-expanding-medicaid-under-the-affordable-care-act/?activeTab=map&currentTimeframe=0&selectedDistributions=status-of-medicaid-expansion-decision&sortModel=%7B%22collId%22%22Location%22%22sort%22%22asc%22%7D](https://www.kff.org/health-reform/state-indicator/state-activity-around-expanding-medicaid-under-the-affordable-care-act/?activeTab=map&currentTimeframe=0&selectedDistributions=status-of-medicaid-expansion-decision&sortModel=%7B%22collId%22%22Location%22%22sort%22%22asc%22%7D)

2. McFadden DL. Conditional Logit Analysis of Qualitative Choice Behavior. In: Frontiers in Econometrics, Zarembka P, 105-142. New York: Academic Press, 1973. [https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf](https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf)

3. Tavakol M, Dennick R. Making sense of Crohnbach’s alpha. Int J Med Educ. 2011; 2: 53–55. Published online 2011 Jun 27. doi: 10.5116/ijme.4dfb.8dfd