Supplemental Appendix

TABLE OF CONTENTS

SOUTHERN COMMUNITY COHORT STUDY 3
MISSING DATA 8
DIFFERENCE-IN-DIFFERENCES MODEL 9
Functional Form *10*
MODELING HEALTH STATUS CHANGES 10
Difference-in-Difference Estimates from the Outcome Transitions Model *13*
Mapping Outcome Transitions to the Three Category Health Status Change Scale *14*
Construction of the Health Status Composite Summary Measure *15*
ASSESSMENT OF THE PARALLEL HEALTH TRENDS ASSUMPTION 18
ASSESSMENT OF COVARIATE BALANCE 22
ACCOUNTING FOR TRUNCATION BY DEATH AND MORTALITY AS AN OUTCOME 25
Sample Selection Issues Arise When Mortality Truncates Observation of Health Outcomes *25*
STATISTICAL METHODS FOR ADDRESSING TRUNCATION BY DEATH 25
ADDRESSING TRUNCATION-BY-DEATH THROUGH PRINCIPAL STRATIFICATION 25
ADDRESSING DEATH THROUGH INVERSE PROBABILITY WEIGHTING 28
DECOMPOSITION OF HEALTH STATUS CHANGES 28
EXACT INFERENCES USING CLUSTER-BASED PERMUTATION 31

STATISTICAL SOFTWARE AND REPLICATION CODE 32

SUPPLEMENTAL EXHIBITS AND FIGURES 33

LIST OF eFIGURES

eFigure 1. Location of SCCS Recruitment Health Clinics and 2013 Uninsured Rate,by Medicaid Expansion Status ... 3
eFigure 2. CONSORT Diagram for Primary Analytic Sample 5
eFigure3. Medical Outcomes Study 12-Item Short-Form General Health Survey (SF-12) as Assessed in the Southern Community Cohort Study Follow-Up Survey 6
eFigure4. Comparison of Physical and Mental Health Summary Scores in the SCCS and an Age- and Gender-Matched Medical Expenditure Panel Survey Sample of Adults in the Southern U.S. with Income <400% FPL 7
eFigure5. SF-12 Nonresponse Patterns 9
eFigure6. Survival Differences Between Expansion and Non-expansion State Samples in the Pre-Expansion Period (January 2010 to December 2013) 21
eFigure7 Kaplan Meier Survival Curves by State Medicaid Expansion Status 34

LIST OF eTABLES

eTable 1. Item Nonresponse for Covariates with Missing Data 8
eTable2. Comparison of Difference-in-Difference Recycled Prediction Point Estimates by Regression Type, Physical Health Limitation Outcome 10
eTable3. Comparison of Post-Expansion Health Status Distribution in the Expansion State Samples 13
eTable4. Comparison of Marginal DiD Estimates from Standard DiD (Eqn 1) and Outcome Transitions DiD Model (Eqn. 2) 14
eTable5. Estimated Health Status Changes for Physical Limitation Outcome (Accomplished Less Due to Physical Limitation) 15
eTable6. Survival Differences Between Expansion and Non-expansion State Samples in the Pre-Expansion Period (January 2010 to December 2013) 20
eTable7. Baseline differences in SF-12 outcomes between Medicaid expansion and non-expansion states 23

eTable8. Comparison of Primary Estimates to Principal Stratification Estimates 27
eTable9. Decomposition of Health Status Changes, Physical Limitation Outcome 29
eTable10. Decomposition of Health Status Changes, Mental Health Limitation Outcome 30
eTable11. Comparison of Rejection Rates Under Alternative Inference Methods 31
eTable12. Individual State Samples in the Analytic SCCS Sample 33
eTable13. Differential changes in SF-12 health outcomes among SCCS participants in Medicaid expansion states, as compared with those in non-expansion states 35
eTable14. Differential Changes in Health Status Transitions Among SCCS Participants in Medicaid expansion states, as compared with those in non-expansion states 38
eTable15. Robustness of Overall Composite Summary Estimates to Alternative Sample and Specification Choices 41

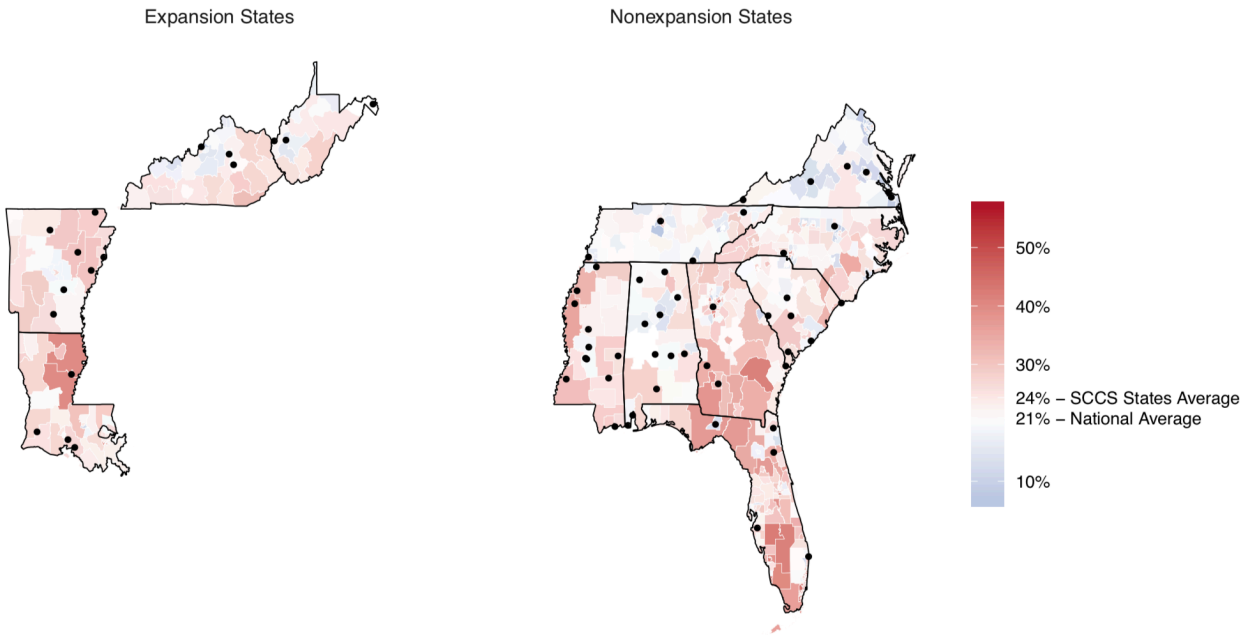
Methodological Appendix

Southern Community Cohort Study

We drew upon survey data and administrative data linkages from the Southern Community Cohort Study (SCCS), the largest-ever recruited cohort of low-income and minority adults in the US. Enrollment into the SCCS occurred between 2002 and 2009 and in total, data were collected on 84,513 adults aged 40 to 79. Details on enrollment and the baseline survey are available in previously published work,[1] as well as on the SCCS program website at https://www.southerncommunitystudy.org/.

Enrollment into the SCCS occurred primarily in health clinics in rural areas. These areas, moreover, had very high uninsured rates in 2013, on the eve of the ACA's insurance expansions in 2014.

eFigure 1. Location of SCCS Recruitment Health Clinics and 2013 Uninsured Rate,by Medicaid Expansion Status



Sources: Authors' analysis of Southern Community Cohort Administrative data an 2013 American Community Survey data .

Notes: Figure shows the geographic location of Southern Community Cohort Study (SCCS) recruitment clinics overlaid on the 2013 uninsured rate among nonelderly adults according to the American Community Survey (ACS).

Because the SF-12 assessment was not included in the baseline enrollment survey, we included only the 53,344 adults who responded to the first follow-up survey (FU1), fielded between 2008 and 2013. We did not use the second follow-up survey (FU2), fielded between 2012 and 2015, because it did not include the SF-12 health assessment.

We developed a third follow-up survey (FU3) targeting the 53,344 FU1 respondents. Of these, 30,841 (57.8%) completed FU3, 17,818 (33.4%) were contacted but did not respond, and 4,546 (8.5%) died before completing the FU3 survey. The remaining 138 participants targeted for FU3 either requested no further contact, dropped out of the study, or were unable to be contacted because a viable address or phone number was not available.

Because death was included in our study outcomes—including SF-12 responses, for which death was coded as a possible response category—our primary study outcomes had an effective response rate of 66.3% (57.8% + 8.5%).

All baseline and follow-up questionnaires are available on the SCCS program website at https://www.southerncommunitystudy.org/questionnaires.html.

Sample Inclusion and Exclusion Criteria

To construct the final analytic sample, we first selected the 43,097 adults who answered the first follow-up survey and also met the following inclusion criteria:

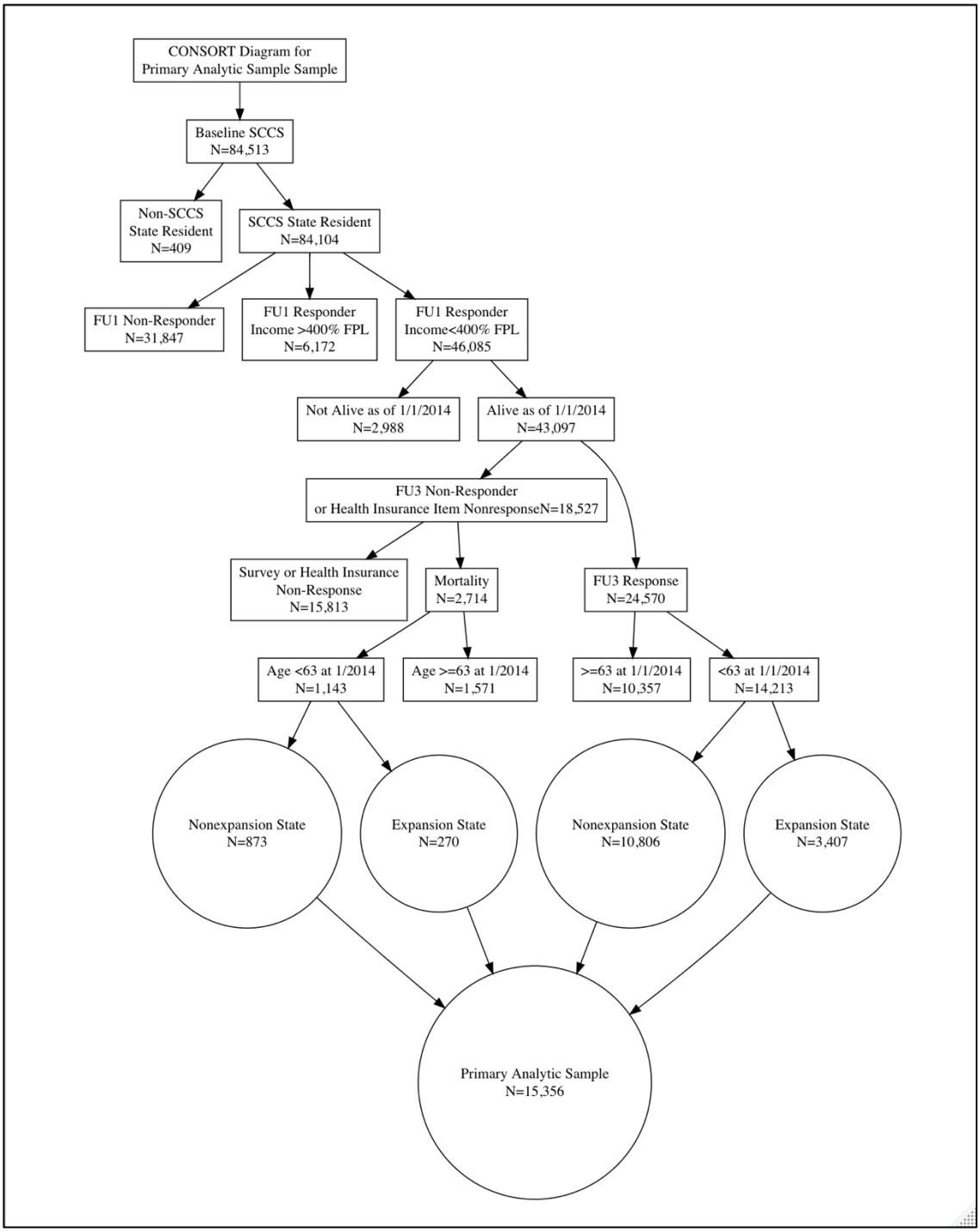
- Primary address in one of the 12 SCCS states.
- Self-reported income at FU1 was approximately 400% of the federal poverty line (FPL) or below.

Among these 43,097 adults, 24,570 were alive as of January 1, 2014, responded to the FU3 survey (fielded from 2015 to 2017) and had non-missing health insurance data. Among FU3 nonresponders, 2,714 did not respond due to death after January 1, 2014 according to vital status data (updated as of Dec 31, 2016).

We further restricted our analytic file to those aged below 63 as of January 1, 2014, to avoid attributing any changes in outcomes to individuals who had aged into Medicare by the end of our study period.

The above study inclusion and exclusion criteria are summarized in the CONSORT diagram in **Error! Reference s ource not found.**, and resulted in a final analytic sample of 15,356 respondents.

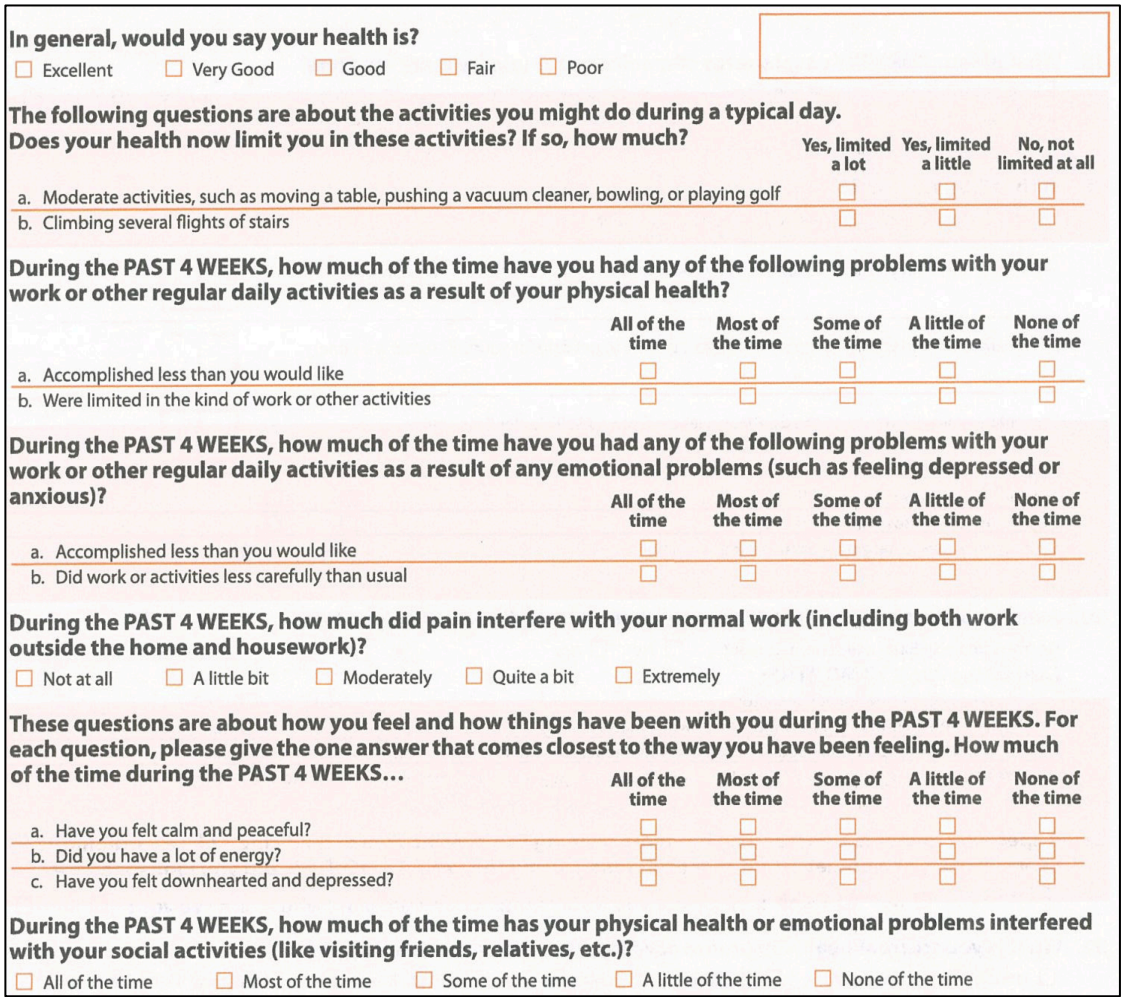
eFigure 2. CONSORT Diagram for Primary Analytic Sample



Primary and Secondary Outcomes

Measures of self-reported health were drawn from the Medical Outcomes Study 12-Item Short-Form General Health Survey (SF-12). The SF-12 questions were asked at FU1 and again at FU3. The questions, as well as possible responses, are provided in eFigure3.

eFigure3. Medical Outcomes Study 12-Item Short-Form General Health Survey (SF-12) as Assessed in the Southern Community Cohort Study Follow-Up Survey



In general, would you say your health is?

☐ Excellent ☐ Very Good ☐ Good ☐ Fair ☐ Poor

The following questions are about the activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

	Yes, limited a lot	Yes, limited a little	No, not limited at all
a. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	☐	☐	☐
b. Climbing several flights of stairs	☐	☐	☐

During the PAST 4 WEEKS, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a. Accomplished less than you would like	☐	☐	☐	☐	☐
b. Were limited in the kind of work or other activities	☐	☐	☐	☐	☐

During the PAST 4 WEEKS, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a. Accomplished less than you would like	☐	☐	☐	☐	☐
b. Did work or activities less carefully than usual	☐	☐	☐	☐	☐

During the PAST 4 WEEKS, how much did pain interfere with your normal work (including both work outside the home and housework)?

☐ Not at all ☐ A little bit ☐ Moderately ☐ Quite a bit ☐ Extremely

These questions are about how you feel and how things have been with you during the PAST 4 WEEKS. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the PAST 4 WEEKS…

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a. Have you felt calm and peaceful?	☐	☐	☐	☐	☐
b. Did you have a lot of energy?	☐	☐	☐	☐	☐
c. Have you felt downhearted and depressed?	☐	☐	☐	☐	☐

During the PAST 4 WEEKS, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?

☐ All of the time ☐ Most of the time ☐ Some of the time ☐ A little of the time ☐ None of the time

The health insurance outcome had the following categories:

1. Private insurance coverage
2. Medicaid coverage
3. Medicare coverage
4. Military coverage
5. Other health insurance coverage ("Other type" in SCCS questionnaire)
6. Uninsured

Because multiple selections were possible, we utilized a hierarchy such that any individual reporting private coverage would receive a value of 1, an individual without private coverage but who reports Medicaid would receive a value of 2, and so on. This resulted in a mutually exclusive categorical health insurance outcome variable. Uninsured was not a category of exclusion, but only recorded for people who marked only that category in their response.

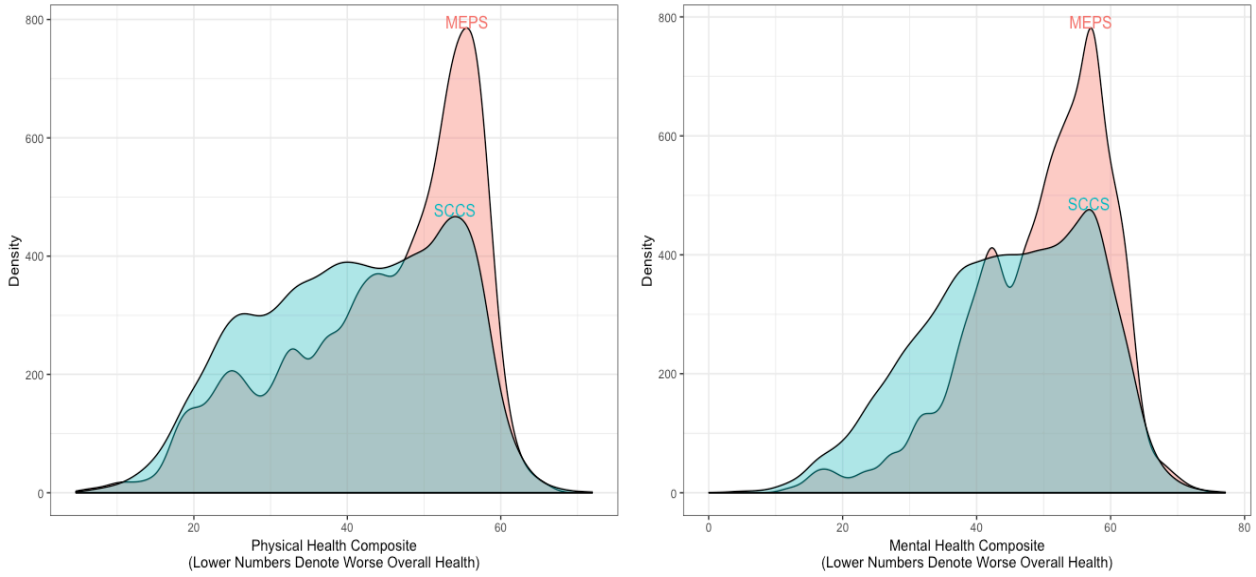
Finally, we also analyzed survival as a secondary outcome. Survival outcomes were assessed via exact administrative matches with the National Death Index (NDI), as well as partial matches performed by trained SCCS personnel based on Social Security and NDI data.

Comparison of SCCS Population to General Low- and Moderate-Income Population

To investigate and compare the overall baseline (FU1) health distribution in the SCCS sample to general low-income populations, we constructed SF-12 summary scores for mental and physical health using standard methods.[2,3] We constructed analogous scores using a sample of adults with income <400% FPL in the South U.S. Census region from the 2013 Medical Expenditure Panel Survey (MEPS). The MEPS sample was age- and gender-matched to match the distribution in our SCCS sample.

Our comparison of the health distribution in these two matched samples is provided in the density plots in eFigure4 below. The distribution of both the physical and mental health summary scores showed that SCCS participants were in worse overall health as compared with a general sample of low-income adults in the south with similar income, age, and gender distributions.

eFigure4. Comparison of Physical and Mental Health Summary Scores in the SCCS and an Age- and Gender-Matched Medical Expenditure Panel Survey Sample of Adults in the Southern U.S. with Income <400% FPL



Missing Data

Based on the inclusion and exclusion criteria outlined above, as well as our primary health outcome definitions (which included death as a category), we observed outcome responses for 66.3% of the cohort targeted for FU3. This was in line with expectations and prior surveys given the demographics of the cohort. However, to test whether there was differential survey nonresponse we fit nonresponse weights using an array of baseline and FU1 characteristics. The model underlying these nonresponse weights was fit using an iterative algorithm to select first- and second-order terms that predicted FU3 response (binary outcome).[4] We then constructed nonresponse weights based on the predicted response probability for each individual, and re-fit the DiD models while weighting for nonresponse. We found no differences between our results with and without adjustment for survey nonresponse—likely because many of the covariates that entered the survey nonresponse model also appeared in our DiD regressions, and our DiD estimates were robust to both the regression specification and the exclusion of controls (see eTable15 in this document).

Item nonresponse for our socio-demographic, socio-economic, and medical history controls was low, as detailed in **Error! Reference source not found.** below. All item nonresponse in our FU3 data was imputed using multiple i mputation via chained equations.[5,6] Our estimates were not sensitive to the use of 5 imputed datasets or a single imputed dataset – again, likely due to the fact that model estimates were robust to the inclusion and exclusion of additional control variables.

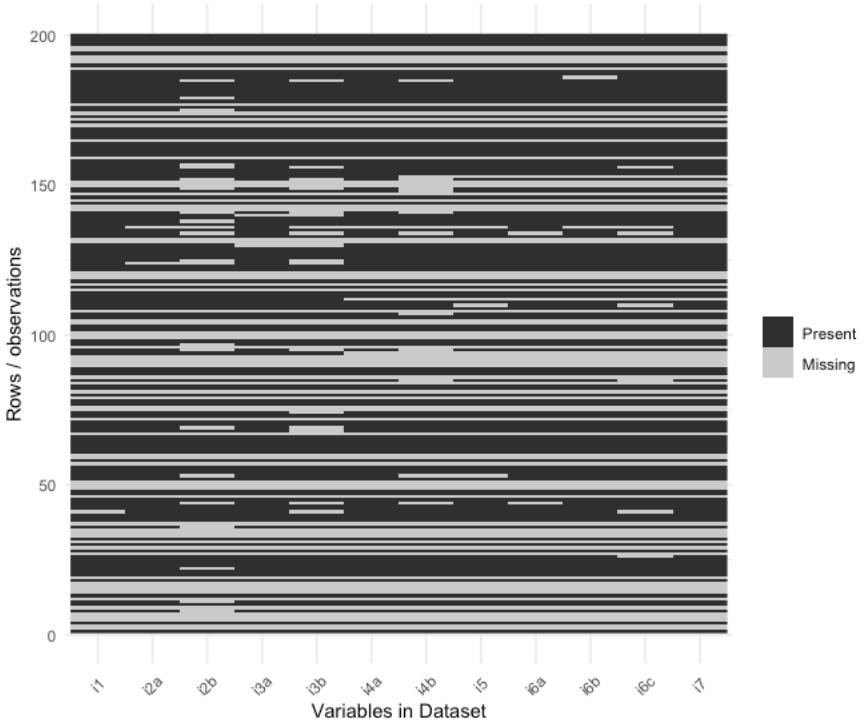
eTable 1. Item Nonresponse for Covariates with Missing Data

Covariate	Percent Missing
Hx Polyps	10.5
Hx Fibroids	8.0
Hx Cancer Diagnosis	4.9
Statin Use	4.6
Hx BPH	4.0
Hx Spine Fracture	3.3
Hx Hip Fracture	2.7
Aspirin Use	2.5
Hx Stroke	2.2
Current Marital Status	2.0
Current Employment Status	2.0
Highest Level of Education	1.9
Hx Multiple Sclerosis	1.8
Hx Parkinsons	1.8
Hx Acute Myocardial Infarction	1.8
Current Smoker	0.90
Race / Ethnicity	0.90

Our sample selection criteria ensured that we had no missing data for the DiD model matrix variables (i.e., a vector of state indicators and a post-expansion indicator), as state residency was determined at baseline for all SCCS participants.

Item nonresponse for the SF-12 health outcomes was also low (range 1% to 14%), though we did note systematically higher nonresponse for the "b" items in the SF-12 question couplets (these couplets can be seen in the alternating shaded regions in eFigure3 above). This systematic nonresponse pattern can be seen in the columns with differentially higher grey patterns in the visualization of observed item response shown in eFigure5 below, which was constructed based on a random sample of 200 observations from our primary analytic sample. We hypothesized that this systematic couplet nonresponse was due to participants' confusion over whether both (a) and (b) questions should have been answered in the mail survey. We did not detect any differential couplet nonresponse across Medicaid expansion and non-expansion states samples, and answers to other SF-12 questions were highly predictive of the missing SF-12 information, so we elected to include all individuals with imputed SF-12 item nonresponse in our sample.

eFigure5. SF-12 Nonresponse Patterns



Difference-in-Differences Model

Our estimates were based on a difference-in-differences (DiD) regression model:

$$h(E(Y_{its})) = \beta_0 + \beta_1 \text{State}_s + \beta_2 \text{Post}_t + \beta_3 \text{Expand}_s \times \text{Post}_t + \beta_4 \text{Socio}_{i,t=pre} + \beta_5 \text{Clinical}_{i,t=pre} \quad (1)$$

where Y_{its} is the health outcome for individual i at time $t \in \{\text{pre}, \text{post}\}$ and in state s, and h is a link function appropriate for Y. In addition, State_s is a vector of state indicators, Expand_s is an indicator for whether the individual resides in a state that expanded its Medicaid program, Post_t is an indicator for whether the observation is in the post-expansion period, $\text{Socio}_{i,t=pre}$ is a vector of baseline sociodemographic and socioeconomic characteristics (e.g., education, gender, marital status, employment status, race/ethnicity, age) and $\text{Health}_{i,t=pre}$ is a vector of baseline medical history controls (e.g., previous diagnosis of cancer, AMI, multiple sclerosis, etc.).

This model specification was used to fit the health insurance outcome and served as the basis for the SF-12 outcome transitions regression. In this model, β_3 estimates the differential change in Y for non-elderly adults in expansion states with incomes under 400% of FPL compared to concurrent changes for similar adults in non-expansion states.

Functional Form

For our primary analyses we used an identity link function and fit the model using multiple multivariate regression. Since all outcomes were categorical, this was equivalent to jointly fitting a series of linear probability models to each of the outcome response categories.

We used multiple multivariate regression over a non-linear functional form for categorical outcomes (e.g., multinomial logit) because we felt it was important to include individual state fixed effects in the primary specification. Logistic models, including the multinomial logit, have been shown to produce inconsistent estimates of regression parameters when including panel-level fixed effects (e.g., group or time dummies) due to the "incidental parameters" problem.[7,8]

Nevertheless, as can be seen in
eTable2 below, our multiple multivariate regression produced identical DiD estimates as a multinomial logit specification when we fit a basic DiD model without state fixed effects (i.e., when swapping in a binary indicator for expansion status for the vector of state dummies in the regression specification above).
eTable2, and in subsequent examples in this section, shows the estimates of health status changes for the SF-12 measure on whether the individual was able to accomplish less than they would have liked due to a physical health limitation. To facilitate comparisons across specification types, the estimates shown below are derived using recycled predictions for the DiD estimate of the change in health status.

eTable2. Comparison of Difference-in-Difference Recycled Prediction Point Estimates by Regression Type, Physical Health Limitation Outcome

Accomplished Less Due to Physical Limitation	Multiple Multivariate Regression	Multinomial Logit Regression
None of the Time	0.019	0.019
A Little of the Time	-0.003	-0.003
Some of the Time	0.009	0.009
Most of the Time	-0.018	-0.018
All of the Time	-0.005	-0.005
Death	-0.002	-0.002

Modeling Health Status Changes

To model health status transitions, we augmented the regression specification in equation (1) above to include interactions between baseline health ($Y_{i,t=pre}$) and each of the DiD model matrix variables. This allowed the DiD estimates to vary by baseline health category.

More formally, the health status transition DiD model took the following form:

$$h(E(Y_{its})) = \beta_0 + \beta_1 \text{State}_s + \beta_2 \text{Post}_t + \beta_3(\text{Expand}_s \times \text{Post}_t) + \beta_4 \text{Socio}_{i,t=pre} + \beta_5 \text{Clinical}_{i,t=pre} + \beta_6 Y_{i,t=pre} + \beta_7(\text{Expand}_s \times Y_{i,t=pre}) + \beta_8(\text{Post}_t \times Y_{i,t=pre}) + \beta_9(\text{State}_s \times Y_{i,t=pre}) + \beta_{10}(\text{Expand}_s \times \text{Post}_t \times Y_{i,t=pre}) \quad (2)$$

where $Y_{i,t=pre}$ is the baseline health category for indivdiual i and the other variables are defined as in equation (1) above. This pooled DiD model specification was equivalent to fitting a series of DiD models separately on the sub-sample of individuals in each baseline health category.

The primary quantities of interest from the transitions DiD regression model were estimates on health status changes associated with Medicaid expansion. As noted in the main text, we identified three mutually exclusive health status change categories:

- No change in health status
- Health status improvement
- Health status decline (including death)

In addition, we were also interested in decomposing the nature of observed health status changes. For example, if fewer people experienced a health status decline, was that the result of fewer people in excellent health status moving to good health status? Or was it because fewer people in fair health status transitioned to poor health status? Or was it because fewer people died?

Below, we detail how we used recycled predictions from the above DiD transitions model to produce estimates on these quantities of interest.

Health Status Changes as a Discrete Time Markov Process

A useful way to represent health status changes is using a Markov trace which multiplies a vector summarizing pre-expansion occupancy in each health status category by a transition probability matrix. For this example, we will consider changes over a two-period cycle, as we have in our study—though the general framework described here also could be used in a multi-period cycle by sampling transition probabilities from cumulative hazards estimated by a multi-state model fit to longitudinal data.

More formally, for the physical health limitation outcome (i.e., accomplished less due to physical limitation), define the pre-expansion health status occupancy vector as follows:

$$\mathbf{p} = \begin{pmatrix} p_{all} \\ p_{most} \\ p_{some} \\ p_{little} \\ p_{none} \\ p_{death} \end{pmatrix}$$

where p_c is the fraction of the sample in each category c in the pre-expansion period.

We can similarly define a post-expansion occupancy vector as follows:

$$\mathbf{p}' = \begin{pmatrix} p'_{all} \\ p'_{most} \\ p'_{some} \\ p'_{little} \\ p'_{none} \\ p'_{death} \end{pmatrix}$$

Finally, define the transition probability matrix **R**:

$$\mathbf{R} = \begin{pmatrix} r_{all,all} & r_{all,most} & r_{all,some} & r_{all,little} & r_{all,none} & r_{all,death} \\ r_{most,all} & r_{most,most} & r_{most,some} & r_{most,little} & r_{most,none} & r_{most,death} \\ r_{some,all} & r_{some,most} & r_{some,some} & r_{some,little} & r_{some,none} & r_{some,death} \\ r_{little,all} & r_{little,most} & r_{little,some} & r_{little,little} & r_{little,none} & r_{little,death} \\ r_{none,all} & r_{none,most} & r_{none,some} & r_{none,little} & r_{none,none} & r_{none,death} \\ r_{death,all} & r_{death,most} & r_{death,some} & r_{death,little} & r_{death,none} & r_{death,death} \end{pmatrix}$$

where $r_{pre,post}$ is the probability of transitioning from baseline category pre to follow-up category $post$.

Basic matrix algebra links the two occupancy vectors as follows:

$$\begin{pmatrix} p_{all} \\ p_{most} \\ p_{some} \\ p_{little} \\ p_{none} \\ p_{death} \end{pmatrix}^T \cdot \begin{pmatrix} r_{all,all} & r_{all,most} & r_{all,some} & r_{all,little} & r_{all,none} & r_{all,death} \\ r_{most,all} & r_{most,most} & r_{most,some} & r_{most,little} & r_{most,none} & r_{most,death} \\ r_{some,all} & r_{some,most} & r_{some,some} & r_{some,little} & r_{some,none} & r_{some,death} \\ r_{little,all} & r_{little,most} & r_{little,some} & r_{little,little} & r_{little,none} & r_{little,death} \\ r_{none,all} & r_{none,most} & r_{none,some} & r_{none,little} & r_{none,none} & r_{none,death} \\ r_{death,all} & r_{death,most} & r_{death,some} & r_{death,little} & r_{death,none} & r_{death,death} \end{pmatrix} = \begin{pmatrix} p'_{all} \\ p'_{most} \\ p'_{some} \\ p'_{little} \\ p'_{none} \\ p'_{death} \end{pmatrix}^T$$

From our sample we can obtain estimates of $\mathbf{p}_{expand}$ and $\mathbf{R}_{expand}$ using simple tabulations and cross-tabulations of the physical limitation outcome:

$$\widehat{\mathbf{p}}_{expand} = \begin{pmatrix} p_{all} = 0.14 \\ p_{most} = 0.18 \\ p_{some} = 0.21 \\ p_{little} = 0.17 \\ p_{none} = 0.29 \\ p_{death} = 0.0 \end{pmatrix}$$

$$\widehat{\mathbf{R}}_{expand} = \begin{pmatrix} 0.36 & 0.26 & 0.16 & 0.043 & 0.049 & 0.14 \\ 0.20 & 0.28 & 0.25 & 0.080 & 0.083 & 0.11 \\ 0.13 & 0.18 & 0.33 & 0.15 & 0.12 & 0.086 \\ 0.087 & 0.095 & 0.24 & 0.23 & 0.28 & 0.060 \\ 0.040 & 0.060 & 0.15 & 0.16 & 0.54 & 0.043 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

and likewise we can also obtain the baseline occupancy vector and transition probability matrix in the non-expansion state sample:

$$\widehat{\mathbf{p}}_{nonexp} = \begin{pmatrix} p_{all} = 0.13 \\ p_{most} = 0.17 \\ p_{some} = 0.23 \\ p_{little} = 0.16 \\ p_{none} = 0.31 \\ p_{death} = 0.0 \end{pmatrix}$$

$$\widehat{\mathbf{R}}_{nonexp} = \begin{pmatrix} 0.34 & 0.24 & 0.16 & 0.070 & 0.061 & 0.12 \\ 0.21 & 0.28 & 0.24 & 0.089 & 0.074 & 0.10 \\ 0.10 & 0.21 & 0.33 & 0.13 & 0.13 & 0.095 \\ 0.081 & 0.13 & 0.26 & 0.21 & 0.25 & 0.069 \\ 0.061 & 0.064 & 0.15 & 0.15 & 0.52 & 0.051 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

We verified that we had accumulated the necessary information by comparing the estimated post-expansion distribution of the outcome versus a simple tabulation of the outcome in the post period. The following table shows this comparison for the expansion state sample:

eTable3. Comparison of Post-Expansion Health Status Distribution in the Expansion State Samples

	All	Most	Some	A Little	None	Death
Markov (1 cycle)	0.143	0.160	0.224	0.140	0.252	0.081
Tabulated	0.143	0.160	0.224	0.140	0.252	0.081

Analogous occupancy vectors and transition probability matrices were defined for the other SF-12 outcomes, as well.

Difference-in-Difference Estimates from the Outcome Transitions Model

The outcome transitions model represented by Equation (2) yields parameter estimates that can be combined via recycled predictions to produce DiD estimates on the overall change in each outcome response category (i.e., estimates equivalent to the coefficient on β_3 in Equation (1)). More importantly, these parameters can also be combined to yield DiD estimates for the transition probability matrix $\mathbf{R}$. It is this estimate of $\mathbf{R}_{DiD}$ that forms the basis of our estimates on health status changes, as well as decompositions of those changes.

To validate our recycled predictions approach we used estimates of $\widehat{R}_{expand}$ and $\widehat{R}_{nonexp}$ to construct unadjusted DiD estimates for the transition probability matrix. This follows since the underlying assumption of DiD is that the experience of the comparison group represents the counterfactual experience for the intervention group in the absence of the intervention. That is, the pre-post transitions in outcomes reflect the first difference, and the difference between expansion and non-expansion states reflects the second difference in a difference-in-differences model. More generally, this approach fits into the *changes-in-changes* framework described by Athey and Imbens (2006).[9]

Thus, under standard DiD assumptions the DiD estimate for the transition probability matrix can be represented as follows:

$$\widehat{\mathbf{R}}_{DiD} = \widehat{\mathbf{R}}_{expand} - \widehat{\mathbf{R}}_{expand} = \begin{pmatrix} 0.018 & 0.011 & -0.0024 & -0.027 & -0.012 & 0.012 \\ -0.0090 & -0.00022 & 0.0021 & -0.0080 & 0.0098 & 0.0054 \\ 0.032 & -0.029 & -0.0067 & 0.025 & -0.012 & -0.0085 \\ 0.0057 & -0.035 & -0.020 & 0.025 & 0.033 & -0.0086 \\ -0.021 & -0.0043 & 0.0051 & 0.0091 & 0.020 & -0.0087 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix}$$

To obtain counterfactual predictions for the health status transitions based on this DiD transition matrix, it is useful to adopt a potential outcomes framework. For treatment status $Z \in \{0,1\}$, define the post-expansion health status

distribution as $\mathbf{p'(Z)}$. That is, $\mathbf{p'(1)}$ is the health status distribution under Medicaid expansion, and $\mathbf{p'(0)}$ is the counterfactual distribution without expansion.

For the expansion state sample, $\mathbf{p_{expand}}'(\mathbf{1})$ is simply the observed distribution of the outcome in the post-expansion period:

$$\mathbf{p_{expand}}'(\mathbf{1})^T = \overbrace{\mathbf{p}_{expand}^T \cdot \mathbf{R}_{expand}}^{\text{Observed Post-Expansion Distribution}} \quad (3)$$

The counterfactual for the expansion state sample is this observed distribution net of the estimated impact of expansion (i.e., the DiD estimate):

$$\mathbf{p_{expand}}'(\mathbf{0})^T = \overbrace{\mathbf{p}_{expand}^T \cdot \mathbf{R}_{expand}}^{\text{Observed Post-Expansion Distribution}} - \overbrace{[\underbrace{(\mathbf{p'}_{expand}^T - \mathbf{p}_{expand}^T)}_{\text{Difference in Treated Group}} - \underbrace{(\mathbf{p'}_{nonexp}^T - \mathbf{p}_{nonexp}^T)}_{\text{Difference in Control Group}}]}^{Difference-in-Differences}$$

which is equivalent to:

$$\mathbf{p_{expand}}'(\mathbf{0})^T = \mathbf{p}_{expand}^T \cdot \mathbf{R}_{expand} - [(\mathbf{p}_{expand}^T \cdot \mathbf{R}_{expand} - \mathbf{p}_{expand}^T) - (\mathbf{p}_{nonexpansion}^T \cdot \mathbf{R}_{nonexpansion} - \mathbf{p}_{nonexpansion}^T)]$$

Since $\mathbf{R}_{DiD} = \mathbf{R}_{expand} - \mathbf{R}_{nonexpand}$ we can express the counterfactual as

$$\mathbf{p_{expand}}'(\mathbf{0})^T = \mathbf{p}_{expand}^T \cdot \mathbf{R}_{expand} - \left[(\mathbf{p}_{expand}^T \cdot \mathbf{R}_{expand} - \mathbf{p}_{expand}^T) - (\mathbf{p}_{nonexpansion}^T \cdot (\mathbf{R}_{expand} - \mathbf{R}_{DiD}) - \mathbf{p}_{nonexpansion}^T)\right] \quad (4)$$

We verified that DiD estimates on the marginal change in the outcome using the regression specification in equation (2), and the difference in potential outcomes estimated using equations (3) and (4), exactly matched DiD estimates on the marginal changes estimated by β_3 in equation (1). This is demonstrated in the table below.

eTable4. Comparison of Marginal DiD Estimates from Standard DiD (Eqn 1) and Outcome Transitions DiD Model (Eqn. 2)

	All	Most	Some	A Little	None	Death
Standard DiD (Equation 1)	-0.00489	-0.0184	0.00945	-0.00274	0.0186	-0.00203
DiD Outcome Transitions Model (Equation 2)	-0.00489	-0.0184	0.00945	-0.00274	0.0186	-0.00203

Mapping Outcome Transitions to the Three Category Health Status Change Scale

The above modeling framework allowed us to map each outcome transition represented in the transition probability matrix to the three-category health status change scale. To do so we defined binary health status change matrices. For the physical limitation outcomes these matrices were defined as follows:

$$\mathbf{C}_{\text{no\_change}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{C}_{\text{improvement}} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{C}_{\text{decline}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

We similarly defined health status change matrices for all SF-12 outcomes. We then integrated these change matrices into the potential outcomes equations so that they yielded DiD estimates on each health status change outcome:

$$\mathbf{p}_{expand}{}'(1)^T = \mathbf{p}_{expand}^T \cdot \left(\mathbf{C}_k \circ \mathbf{R}_{expand}\right). \qquad (5)$$

$$\mathbf{p}_{expand}{}'(0)^T = \mathbf{p}_{expand}^T \cdot (\mathbf{C}_k \circ \mathbf{R}_{expand}) - [(\mathbf{p}_{expand}^T \cdot (\mathbf{C}_k \circ \mathbf{R}_{expand}) - \mathbf{p}_{expand}^T) - . \\ (\mathbf{p}_{nonexpansion}^T \cdot (\mathbf{C}_k \circ (\mathbf{R}_{expand} - \mathbf{R}_{DiD})) - \mathbf{p}_{nonexpansion}^T)]. \qquad (6)$$

where $\circ$ is the element-wise matrix multiplication operator. eTable5 below provides estimates for these health status change outcomes for the physical limitation outcome.

eTable5. Estimated Health Status Changes for Physical Limitation Outcome (Accomplished Less Due to Physical Limitation)

	Observed $\mathbf{p}_{expand}{}'(1)$	Counterfactual $\mathbf{p}_{expand}{}'(0)$	DiD Estimate $\mathbf{p}'_{expand}(1) - \mathbf{p}_{expand}{}'(0)$
No Change	0.3692	0.3628	0.006325
Health Improvement	0.2554	0.2419	0.01356
Health Decline	0.3754	0.3953	-0.01988

Thus, our DiD show that the probability of experiencing a health status decline declined by 1.99 percentage points. We describe our procedure for producing inference estimates on this estimated quantity of interest later in this document.

Construction of the Health Status Composite Summary Measure

Equations (5) and (6) formed the backbone of our primary results on health status changes. To aggregate these results into composite summary measures, we took the average of the DiD health status change estimates across the

twelve SF-12 questions for each of the three categories. That is, we separately took the averages of the 12 no change estimates, the 12 health status improvement estimates, and the 12 health status decline estimates.

We also created physical and mental health sub-scores by averaging across the following subsets of SF-12 outcomes:

Physical Health Sub-Score

- Does your health now limit you in moderate activities?
- Does your health now limit you in climbing several flights of stairs?
- During the past 4 weeks, how much of the time have you accomplished less than you would like?
- During the past 4 weeks, how much of the time were you limited in the kind of work or other activities?
- During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

Mental Health Sub-Score

- During the past 4 weeks, how much of the time have you accomplished less than you would like as a result of any emotional problems?
- During the past 4 weeks, how much of the time did you work or conduct activities less carefully than usual as a result of any emotional problems?
- During the past 4 weeks, have you felt calm and peaceful?
- During the past 4 weeks, did you have a lot of energy?
- During the past 4 weeks, have you felt downhearted and depressed?
- During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities?

Medical Outcomes Study 12-Item Short-Form General Health Survey (SF-12) domains, questions, and possible responses

Overall Health Items

In general, would you say your health is?
[*Excellent; very good; good; fair; poor]*

Physical Health Items

Does your health now limit you in these activities? If so, how much during a typical day?

Moderate activities such as moving a table, pushing a vacuum cleaner, bowling, or playing golf?
[*No, not limited at all; yes, limited a little ; yes, limited a lot*]

Climbing several flights of stairs
[*No, not limited at all; yes, limited a little ; yes, limited a lot*]

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health:

Accomplished less than you would like
[*None of the time; a little of the time; some of the time; most of the time; all of the time]*

Were limited in the kind of work or other activities
[*None of the time; a little of the time; some of the time; most of the time; all of the time*]

During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?
[*Not at all; a little bit; moderately; quite a bit; extremely]*

Mental Health Items

During the past 4 weeks, how much of the time have you had any of the following problems with your work? or other regular daily actvities as a result of any emotional problems (such as feeling depressed or anxious)?

Accomplished less than you would like
[*None of the time; a little of the time; some of the time; most of the time; all of the time*]

Did work or activities less carefully than usual
[None of the time; a little of the time; some of the time; most of the time; all of the time]

For the past 4 weeks, how much of the time ...

Have you felt calm and peaceful?
[None of the time; a little of the time; some of the time; most of the time; all of the time]

Did you have a lot of energy?
[None of the time; a little of the time; some of the time; most of the time; all of the time]

Have you felt downhearted and depressed?
[None of the time; a little of the time; some of the time; most of the time; all of the time]

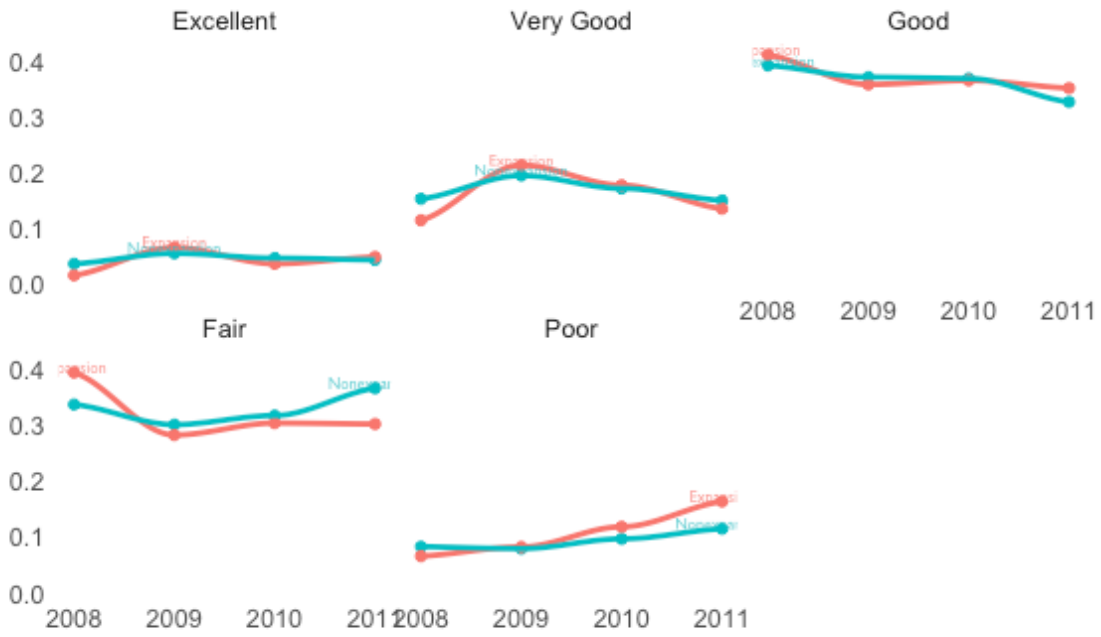
During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)
[None of the time; a little of the time; some of the time; most of the time; all of the time]

Assessment of the Parallel Health Trends Assumption

Identification of the effects of Medicaid expansion based on the DiD model required the assumption that the experience of the non-expansion state sample in the post-expansion period represented, on average, what the experience of the expansion state sample would have been in the absence of expansion. It is therefore common when using a DiD design to assess this "parallel trends assumption" based on examination of trends in the outcome prior to the intervention.

Fortunately, the (pre-expansion) FU1 survey was not administered all at once, but was rather spread over a roughly 4 year period. Thus, we can leverage the staggered nature of FU1 surveying to look cross-sectionally at whether outcome trends were diverging in the expansion and non-expansion state samples from 2008-2011.

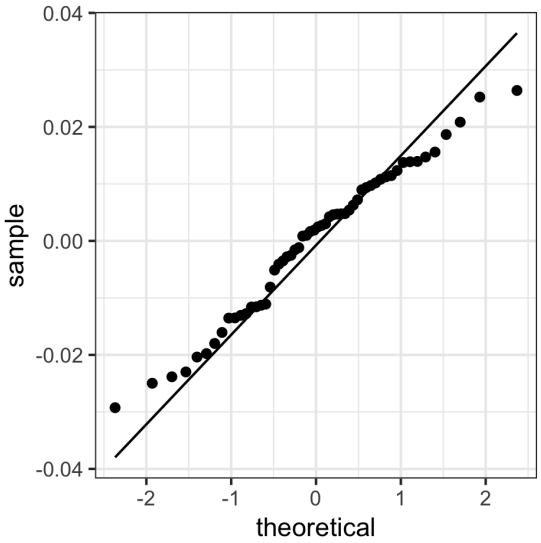
The plot below fits LOESS curves cross-sectional changes (by year) for the expansion and nonexpansion state samples for the general health status question. Note that surveying timing was non-random (the SCCS targets certain populations for surveying first), so the levels bounce around a bit owing to the compositional differences at each point in time.



As seen in the figure, most outcome categories trended similarly between the two groups. To the extent there are possible trend differences for certain categories, they point in the direction of expansion state cohort members being slightly more likely to be in *poorer* health over time. This pattern goes against our main results, since to explain our results & inferences, the four expansion states would need to be differentially *improving* in health status over time—i.e., the opposite of what is suggested in the plot.

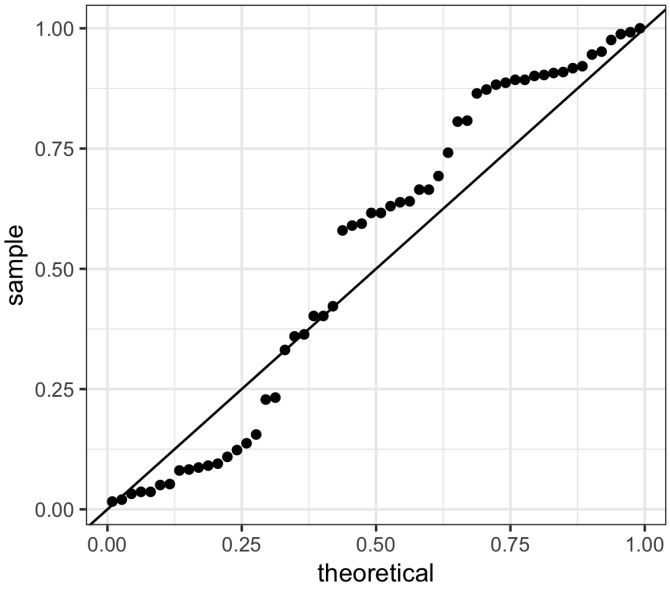
Moreover, the figure above provides trend estimates for the 5 outcome categories in the general health status SF-12 measure; any observed differences in trends could be the result of statistical noise rather than evidence of true trend differences. To more rigorously assess whether there were any meaningful differences in trends across the SF-12, we next fit a series of linear regressions on pre-2014 data for each of the 56 possible outcome categories. We then plot these 56 trend estimates against a hypothetical normal distribution with mean 0 to assess whether the trend estimates

cluster above or below 0, or center around 0 (i.e., no difference in trend). These estimates are shown in the quantile-quantile (QQ) plot in the figure below:



We see here that the 56 trend estimates generally follow a normal distribution with mean 0, with no systematic clustering of trend estimates either well below or well above 0. This evidence indicates that there were few meaningful differences in outcome trends across the SF-12 measures.

We next producing measures summarizing our statistical inferences (the quantile ranking) based on the randomization inference procedure described in the main text and in this document. That is, for each of the 56 trend estimates we obtain a quantile ranking based on the distribution of 495 estimates produced by permuting state expansion status. Below, we plot a similar quantile-quantile plot of a uniform distribution, under the assumption that, if there were no systematic differences in trends, the quantile rankings for the 56 trend estimates should follow a uniform distribution between 0 and 1.



We see, again, no evidence of systematic pooling of quantile rankings at the extreme values (i.e., at around 0, and 1). Again, this indicates that the quantile rankings for the 56 trend estimates for the SF-12 follow a fairly uniform distribution, with values falling more evenly over the range of 0 to 1.

In addition to the assessment of outcome trends above, the SCCS data also included administrative linkages to Death Registry and Social Security data. This allowed us to test for parallel trends using mortality as a proxy measure of health status. A distinct advantage of this approach was that it provided a continuously-measured health outcome over a four-year period.

To analyze differences in survival trends between the expansion and non-expansion state groups we constructed an analogous sample to our main analytic sample but with one difference: rather than selecting individuals alive as of 12/31/2013 (on the eve of expansion) we selected individuals alive as of 1/1/2010 (the year the ACA was passed). We then censored survival in this sample at 12/31/2013 to limit our analysis to the pre-expansion period.

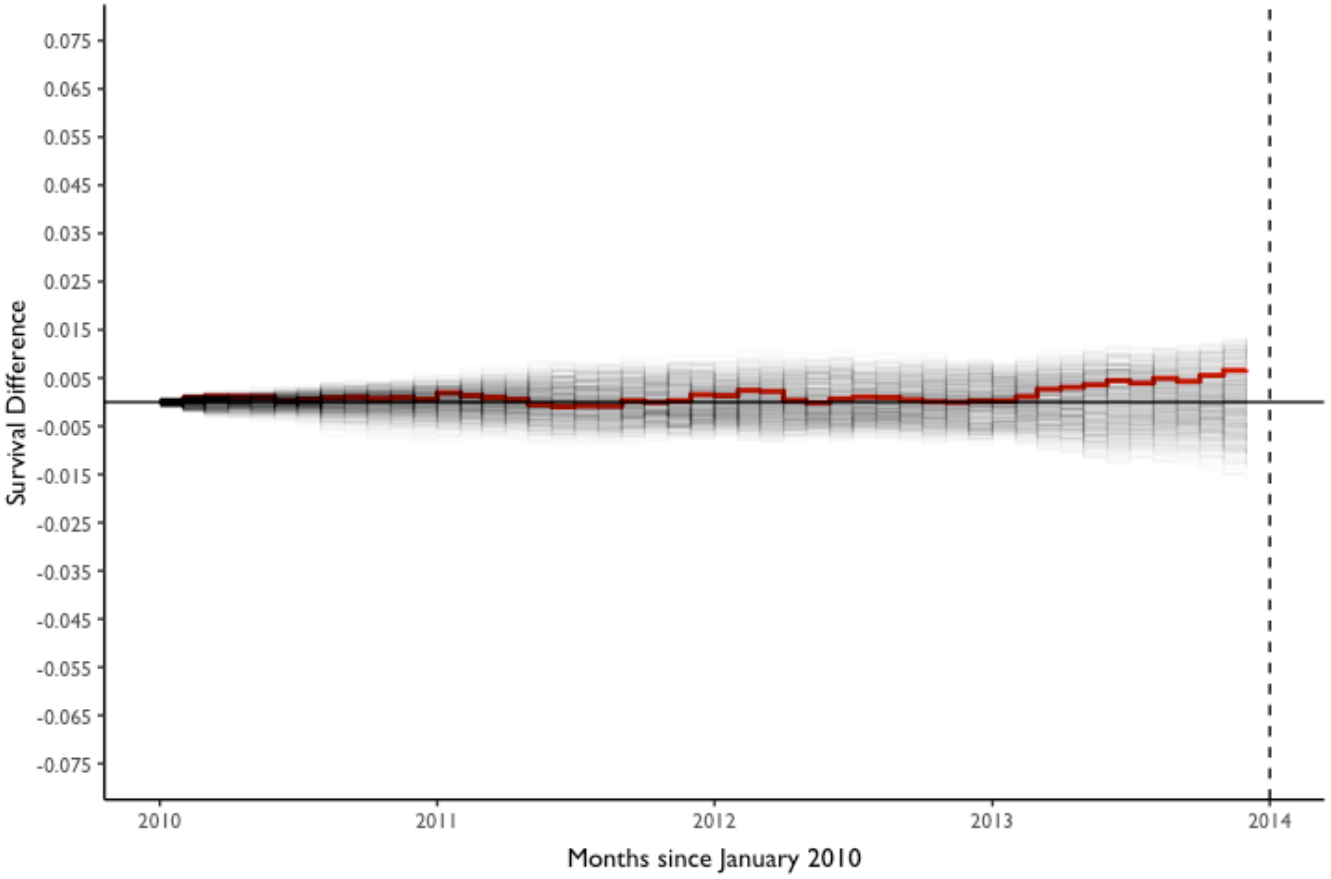
We fit non-parametric Kaplan-Meier survival estimates to the sample separately for the expansion and non-expansion state samples. To test for differences in survival trends we re-fit the survival estimates using the permutation inference procedure as described below and in the main text. We used the permutation estimates to test for differences in survival at various dates in the pre-expansion period.

The survival difference between the expansion and non-expansion state samples, along with the associated percentile rankings, are provided in the table below. In addition, the figure plots the survival difference in each month during the pre-expansion period (red line) along with the survival difference for each of the 495 permuted samples (grey lines). At no point during the pre-expansion period do we detect statistically or substantively significant differences in survival between the two groups.

eTable6. Survival Differences Between Expansion and Non-expansion State Samples in the Pre-Expansion Period (January 2010 to December 2013)

Date	Survival Difference	Percentile Ranking
As of December 2010	0.05	57.6
As of December 2011	0.15	65.3
As of December 2012	0.03	51.7
As of December 2013	0.67	80.2

eFigure6. Survival Differences Between Expansion and Non-expansion State Samples in the Pre-Expansion Period (January 2010 to December 2013)



Assessment of Covariate Balance

While the difference-in-difference study design relies on the parallel trends assumption—i.e., the design does not rely on treated and control units being balanced on observable (or unobservable) characteristics, as in a randomized design[7,9]—it is nevertheless useful to assess covariate balance. Balance assessments serve as a useful diagnostic to investigate the potential for statistical model dependence,[10] the potential for differences between groups to contribute to differences in outcome trends over time (i.e., potential violations of the parallel trends assumption),[11] and the possibility of bias due to regression-to-the-mean effects.[12]

We assessed covariate imbalance by estimating the standardized mean difference (SMD) between sample observations in Medicaid expansion versus non-expansion states. For our SMD estimates we used the unified approach for continuous and categorical variables outlined in Yang and Dalton (2012).[13] The use of SMD in sample balance assessments has a long history in the propensity score literature,[11,14,15] and has the advantage of putting mean differences all on the same (standard deviation) scale—thereby facilitating apples-to-apples comparisons across outcomes and covariates measured on different scales. In this literature, a SMD of 0.1 or less (in absolute magnitude) is generally regarded as a negligible difference—though regression adjustment has been shown in simulation studies to address imbalance with SMDs of 0.25 or less (in absolute magnitude).[14,16]

As shown in Exhibit 2 in the main text, we found that our sample was well balanced on all outcomes and nearly all covariates. Importantly, SMDs for the pre-expansion SF-12 measures were low, as detailed in the table below; the largest SMD is 0.121, though most SMDs range between 0.03 and 0.07. These negligible SMDs indicate that our difference-in-difference estimates do not suffer from bias from regression-to-the-mean effects in the time-varying outcomes and covariates in our sample.

eTable7. Baseline differences in SF-12 outcomes between Medicaid expansion and non-expansion states

	Nonexpansion State Sample (N=11679)	Expansion State Sample (N=3677)	Standardized Mean Difference
General Health Status, No. (%)			0.072
Excellent	576 (4.93)	187 (5.09)	
Very good	2073 (17.75)	694 (18.87)	
Good	4300 (36.82)	1326 (36.06)	
Fair	3698 (31.66)	1085 (29.51)	
Poor	1032 (8.84)	385 (10.47)	
Health limitation: moderate daily activities, No. (%)			0.021
Limited a lot	2948 (25.24)	921 (25.05)	
Limited a little	3383 (28.97)	1100 (29.92)	
Not limited at all	5348 (45.79)	1656 (45.04)	
Health limitation: climbing stairs, No. (%)			0.041
Limited a lot	3755 (32.15)	1254 (34.10)	
Limited a little	3524 (30.17)	1077 (29.29)	
Not limited at all	4400 (37.67)	1346 (36.61)	
Accomplished less than would have liked due to physical limitation, No. (%)			0.062
All the time	1547 (13.25)	532 (14.47)	
Most of the time	2022 (17.31)	672 (18.28)	
Some of the time	2659 (22.77)	788 (21.43)	
A little of the time	1914 (16.39)	634 (17.24)	
None of the time	3537 (30.29)	1051 (28.58)	
Physical limitation limits work or other activities, No. (%)			0.066
All the time	1779 (15.23)	629 (17.11)	
Most of the time	1875 (16.05)	578 (15.72)	
Some of the time	2443 (20.92)	716 (19.47)	
A little of the time	1697 (14.53)	572 (15.56)	
None of the time	3885 (33.26)	1182 (32.15)	
Accomplished less than would have liked due to emotional problem, No. (%)			0.032
All the time	1002 (8.58)	327 (8.89)	
Most of the time	1578 (13.51)	525 (14.28)	
Some of the time	2575 (22.05)	788 (21.43)	
A little of the time	1963 (16.81)	634 (17.24)	
None of the time	4561 (39.05)	1403 (38.16)	

[a] Medicaid expansion states include Arkansas, Kentucky, Louisiana and West Virginia while the nonexpansion state sample includes Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (which had not yet expanded during the post-expansion surveying period).

NOTE: Table continues on next page

NOTE: This table continues from the previous page

	Nonexpansion State Sample (N=11679)	Expansion State Sample (N=3677)	Standardized Mean Difference
Did work or actvities less carefully due to emotional problem, No. (%)			0.029
All the time	1039 (8.90)	316 (8.59)	
Most of the time	1285 (11.00)	427 (11.61)	
Some of the time	2188 (18.73)	662 (18.00)	
A little of the time	2094 (17.93)	653 (17.76)	
None of the time	5073 (43.44)	1619 (44.03)	
Pain interferes with work and other activities, No. (%)			0.084
Not at all	3163 (27.08)	894 (24.31)	
A little bit	2737 (23.44)	865 (23.52)	
Moderately	1694 (14.50)	567 (15.42)	
Quite a bit	2662 (22.79)	825 (22.44)	
Extremely	1423 (12.18)	526 (14.31)	
Felt calm and peaceful, No. (%)			0.069
All the time	1100 (9.42)	281 (7.64)	
Most of the time	4393 (37.61)	1394 (37.91)	
Some of the time	3482 (29.81)	1097 (29.83)	
A little of the time	1994 (17.07)	677 (18.41)	
None of the time	710 (6.08)	228 (6.20)	
Have a lot of energy, No. (%)			0.087
All the time	675 (5.78)	162 (4.41)	
Most of the time	2885 (24.70)	841 (22.87)	
Some of the time	3725 (31.89)	1201 (32.66)	
A little of the time	2657 (22.75)	921 (25.05)	
None of the time	1737 (14.87)	552 (15.01)	
Felt depressed, No. (%)			0.063
All the time	773 (6.62)	260 (7.07)	
Most of the time	1572 (13.46)	520 (14.14)	
Some of the time	3034 (25.98)	918 (24.97)	
A little of the time	3011 (25.78)	1018 (27.69)	
None of the time	3289 (28.16)	961 (26.14)	
Physical or emotional problem interfered with social activities, No. (%)			0.057
All the time	1066 (9.13)	337 (9.17)	
Most of the time	1836 (15.72)	648 (17.62)	
Some of the time	2797 (23.95)	827 (22.49)	
A little of the time	1928 (16.51)	616 (16.75)	
None of the time	4052 (34.69)	1249 (33.97)	

[a] Medicaid expansion states include Arkansas, Kentucky, Louisiana and West Virginia while the nonexpansion state sample includes Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (which had not yet expanded during the post-expansion surveying period).

Accounting for Truncation by Death and Mortality as an Outcome

Sample Selection Issues Arise When Mortality Truncates Observation of Health Outcomes

When an intervention affects both health and mortality, observed health outcomes in the group exposed to the intervention will reflect outcomes among two types of individuals: (1) those who would only survive with the intervention and (2) those who would survive both with and without the intervention. By comparison, post-intervention health outcomes among individuals who would die without the intervention are unobserved in a comparison group not exposed to the intervention, since these observations can only be made among the living.[17]

So-called "truncation by death" results in differential sample selection that may bias estimates of outcome changes. Suppose, for example, that the group of individuals who would die without the intervention (i.e., Medicaid expansion) are in poorer health, on average. If Medicaid confers a survival benefit—as previous research has found[18,19]—then failure to address death as a competing risk may result in analyses that show self-reported poor health differentially *increased* in expansion as compared with non-expansion states. However, this differential increase in poor health is the result of a health status improvement (i.e., expansion leaves more individuals alive to report their health status) and not a health status decline.

Statistical Methods for Addressing Truncation by Death

If the health outcomes under consideration are categorical, the simplest and most straightforward method for accounting for death is to include mortality as an additional outcome response category. This is the approach we adopted for our main results. It is also the approach advocated in earlier work by Polsky et al. in their analyses of Medicare health impacts.[20,21]

A general challenge with this approach, however, is that not all health outcomes are categorical. If the primary outcome of interest is continuous (e.g., a clinical test or biomarker value, or a composite measure constructed from a set of health measures), then death cannot be easily incorporated. Doing so would either require strong assumptions on the weight placed on death relative to the other health outcome values, or nonsensical values being imposed on the outcome (e.g., imputing a clinical test value of zero for individuals who die). Previously developed (continuous) SF-12 composite summary scores fall under this category, because they were not developed to incorporate death as a health outcome.[2,3,22]

Earlier work by McWilliams et al. (2009) identifies an additional challenge when including death as an outcome category. If the underlying processes generating changes in the health outcome and mortality differ—which may be the case if the mortality effects of expanded access to coverage take time to be realized—then jointly estimating health status and mortality changes using the same statistical model could lead to biased results.[23] It is for this reason that we explore the robustness of our findings using an exploratory principal stratification approach.

Addressing Truncation-by-Death Through Principal Stratification

The basic idea underlying principal stratification is to stratify the sample on a post-exposure variable. In the context of our study, the post-exposure variable was survival.

That is, each sample member falls into one of four mutually exclusive strata:

(1) Individuals who would survive both with and without expanded access to Medicaid.
(2) Individuals who would survive with Medicaid, but would die without it.
(3) Individuals who would die with Medicaid, but would survive without it.
(4) Individuals would die both with and without Medicaid.

Using observed data it is not possible to uniquely assign each individual to their stratum because post-expansion outcomes observed in the treatment group will reflect mixture of individuals from groups (1) and (2); similarly, post-expansion outcomes will be observed in the control group among a mixture of individuals from groups (1) and (3). Thus, using observed data and information on treatment assignment we can narrow down, but not conclusively assign, each individual to their unique stratum.

Principal stratification methods rely on additional assumptions to isolate treatment effect estimates within the sub-strata of always-survivors (i.e., stratum #1). That is, principal stratification analyses sidestep issues of sample selection bias that arise when death competes with the primary health outcome because in the sub-stratum of always-survivors, there are no decedents.

We detail the intuition behind these assumptions below, however because we do not rely on the principal stratification results for our primary results, we leave many of the statistical details to the existing literature.[17,24,25]

Assumption 1: Quasi-Randomization and Parallel Trends

Principal stratification methods were developed in the context of clinical trials in which assignment to the intervention is random.[17,24,26] Thus, randomization is an important assumption underlying nearly all prior work on principal stratification methods.

In the context of our study we did not have explicit randomization of the intervention, and relied on additional assumptions and an alternative study design (differences-in-differences, or DiD) to identify treatment effects. We detail the assumptions of DiD and explore evidence of the appropriateness of the DiD assumptions elsewhere in this document. Below, we demonstrate the models and assumptions needed to produce principal stratification estimates using a DiD model.

Under random assignment the average treatment effect of an intervention (Z) on an outcome (Y) can be estimated using the following model:

$$Y_i = \gamma_0 + \gamma_1 Z_i + \epsilon_i \qquad (7)$$

for individual i, and where Z_i is a treatment indicator equal to 1 if the individual is in the intervention group. Under this model γ_1 identifies the average treatment effect of the intervention.

Alternatively, in the case of a two-period (i.e., pre and post), individual-level panel data setting it is useful to consider the following difference-in-differences model:

$$y_{it} = \alpha + \beta_0 Post_t + \beta_1 G_{it} + \mu_i + \epsilon_{it} \qquad (8)$$

where $Post_t$ is a binary indicator for whether the observation is in the post-intervention period at time t (i.e., $t = 2$), G_{it} is a binary indicator for whether the individual receives the intervention at time t, μ_i summarizes individual-level heterogeneity, and ϵ_{it} is an idiosyncratic error term. In this model the coefficient β_1 estimates the treatment effect under standard DiD parallel trends assumptions.[7]

A simple estimation procedure is to net out μ_i from the model above by taking a first difference:

$$(y_{i2} - y_{i1}) = \beta_0 + \beta_1 (G_{i2} - G_{i1}) + (\epsilon_{i2} - \epsilon_{i1}) \qquad (9)$$

The first-differencing eliminates the (fixed) individual-level heterogeneity summarized by μ_i . Moreover, since Z_{i1} is 0 for all individuals in the first time period (i.e., in the pre-expansion period) we can re-write this model as

$$(y_{i2} - y_{i1}) = \beta_0 + \beta_1 Z_i + \eta_{it} \qquad (10)$$

where $\eta_{it} = (\epsilon_{i2} - \epsilon_{i1})$ and, as above, Z_i is an indicator for whether individual i is in the intervention group. That is, the DiD panel model can be estimated by regressing the individual change score on a treatment indicator.

Equation (10) is the DiD estimator we use to fit our principal stratification results. This model is useful because it provides a direct linkage to the regression model underlying standard principal stratification approaches in equation (7) above (i.e., a regression of the outcome on the treatment indicator). However, underlying this model is an assumption of effective (quasi) randomization of the treatment exposure. Since decisions on Medicaid expansion were effectively political decisions, this may be a reasonable assumption, particularly within the 12 states we consider within the South. Indeed, our comparison of baseline characteristics (see main text Exhibit 2), as well as survival trend differences, demonstrate reasonable balance across expansion state groupings in our sample (most standardized mean differences are <0.10)—lending some evidence in support of this assumption.

Assumption 2: Monotonicity

An additional assumption required for principal stratification is monotonicity. Under this assumption, the intervention can only improve survival, not decrease it. In effect, monotonicity rules out the existence of sub-strata #3, i.e., individuals who would die with Medicaid and who would survive without it.

By ruling out stratum 3 by the monotonicity assumption, we can identify the fraction of the sample who are always-survivors (i.e., stratum 1) using observed survival in the non-expansion state sample. This follows because under monotonicity, those surviving in the non-expansion state sample can only come from among the sub-strata of always-survivors.

Assumption 3: Principal Ignorability

A final assumption—principal ignorability—is needed to differentiate among the remaining strata. For this assumption, we rely on the rich set of socio-economic, demographic, health status and medical history variables collected by the SCCS at baseline and in the first follow-up survey to predict stratum membership using a principal score model.[24] The principal score model was analogous to fitting a propensity score for stratum membership.

Our principal stratification estimates are based on using the estimated probability of stratum membership for each individual to recover an average treatment effect for the Medicaid expansion state sample. Fitting the principal stratification model yielded the following estimates for the physical health limitation outcome:

eTable8. Comparison of Primary Estimates to Principal Stratification Estimates

	DiD Estimate Primary Model (Eq. 2)	DiD Estimate Change Score Model (Eq. 10)	Principal Stratification DiD Estimate
No Change	0.006325	.0064	-0.0045
Health Improvement	0.014	0.014	0.005
Health Decline	-0.020	-0.018	-0.030

As can be seen in the table, principal stratification results in somewhat smaller estimates for health improvement category, but a larger estimate (in absolute magnitude) for the no change and health status decline categories.

Addressing Death Through Inverse Probability Weighting

An alternative method for addressing death in a longitudinal sample is through inverse probability weighting (IPW). This approach accounts for sample attrition due to death by effectively "weighting up" surviving units that are similar to decedents. That is, a statistical model (typically a logit) is fit predicting death status—and predictions from this model are used to construct inverse probability of death weights. In effect, non-decedents who are observably similar to decedents receive higher analytic weights and thus are used to "stand-in" for the decedents in any analysis. Conceptually, the theory underlying IPW treats health outcomes among the dead as missing data that are imputed using the observed health outcomes among the living. The target population in an IPW analysis is therefore a "pseudo-population" in which health status is always defined because individuals never die.[27]

While principal stratification methods also focus analyses on the living, the relationship between health outcomes and death is conceptualized in a fundamentally different way. Under principal stratification, health outcomes among decedents are not missing, they are undefined[26]—an individual who has died cannot have a functional limitation that limits their ability work. Thus, estimates obtained via principal stratification have a very different population in mind: the stratum of "always survivors." It is important to emphasize that this stratum is not conceptualized to live forever; rather, it is the stratum of individuals who, over the study period, would have survived both with and without the intervention.

In short, IPW methods produce estimates by up-weighting sample units observably similar to decedents, whereas principal stratification produce sub-group estimates isolated to the substratum of individuals who survive both with and without the intervention.

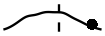
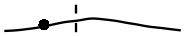
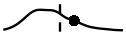
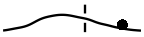
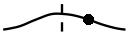
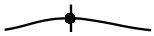
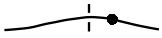
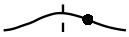
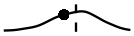
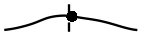
Decomposition of Health Status Changes

Our main results demonstrated that the Medicaid expansions were associated with health improvements via two mechanisms: a higher likelihood of maintaining pre-expansion levels of health (i.e., an increase in the probability of no change in health status) and a lower likelihood of experiencing a health status decline.

Using the DiD estimate for the transition probability matrix, we decomposed these overall findings to explore whether they were concentrated in particular areas of the health distribution. For example, the decrease in the probability of a health decline could reflect fewer transitions to severe health limitations, fewer transitions to minor health limitations, or both.

eTable9 below performs such as decomposition for the physical limitation outcome (i.e., "Does your health now limit you in moderate activities such as moving a table, pushing a vacuum cleaner, bowling, or playing golf?"). Within each health status change category panel, the rows sum to the total DiD estimate for the physical health status limitation in the first column of eTable8 (though note that those estimates have not been multiplied by 100, as they are in eTable9).

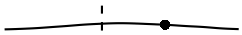
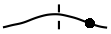
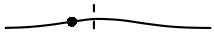
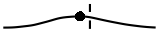
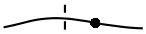
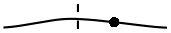
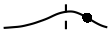
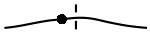
eTable9. Decomposition of Health Status Changes, Physical Limitation Outcome

	Difference–in–Differences Estimate	Inference Sparkline	Percentile Rank
No Change			
None of the Time	–0.33		22.4
A Little of the Time	0.61		98.0
Some of the Time	–0.59		5.5
Most of the Time	0.26		86.3
All of the Time	0.69		97.6
Death	–		–
Health Improvement			
None of the Time	0.49		83.8
A Little of the Time	–0.02		47.5
Some of the Time	0.43		76.6
Most of the Time	0.46		84.4
All of the Time	–		–
Death	–		–
Health Decline			
None of the Time	–		–
A Little of the Time	0.00		54.7
Some of the Time	–0.22		29.9
Most of the Time	–1.6		0.2
All of the Time	0.05		54.1
Death	–0.22		33.5

As seen in the table, our overall estimate for a health status decline (-1.7 ppts) is mostly comprised of a differential reduction in individuals experiencing a decline to the "Most of the Time" (-1.5 ppts) category.

eTable10 below provides analogous decomposition estimates for a key mental health outcome ("Accomplished less than you would like due to an emotional problem"). Again, we see that the overall DiD estimate on the health status decline category is concentrated primarily among a reduction in declines to severe ("Most of the Time") mental health limitations.

eTable10. Decomposition of Health Status Changes, Mental Health Limitation Outcome

	Difference–in–Differences Estimate	Inference Sparkline	Percentile Rank
No Change			
None of the Time	0.95		77.2
A Little of the Time	0.46		95.4
Some of the Time	–0.33		20.4
Most of the Time	–0.15		51.1
All of the Time	0.46		86.3
Death	–		–
Health Improvement			
None of the Time	0.55		82.4
A Little of the Time	0.44		75.2
Some of the Time	–0.50		20.2
Most of the Time	0.37		95.2
All of the Time	–		–
Death	–		–
Health Decline			
None of the Time	–		–
A Little of the Time	0.32		88.7
Some of the Time	–0.35		22.6
Most of the Time	–0.78		2.4
All of the Time	–1.2		0.2
Death	–0.21		33.7

Exact Inferences Using Cluster-Based Permutation

As noted in the main text, a growing body of statistical and econometric research has demonstrated that standard inference approaches yield unacceptably high false discovery rates in the context of a group-level intervention in which the total number of groups (clusters) is small.[28–33] We therefore adopted a cluster permutation-based inference procedure that allowed our observational study to be analyzed like a cluster randomized trial.

Under this inference procedure, we re-produced each estimate under all the possible 495 combinations of 4 expansion states among 12 total states. We then compared the "extremeness" of the observed estimate obtained using the 4 true expansion states to the overall distribution of 495 estimates. Rather than produce p-values we obtained the quantile ranking of each estimate within this distribution. We did not estimate 95% confidence intervals using the exact inference approach because doing so required specifying a constant additive treatment effect,[30,34] and in our view an assumption of no heterogeneity in the impact of Medicaid on outcomes across population sub-groups was unlikely to hold in practice.

The permutation-based inference method had a number of unique advantages. First, the procedure did not require parametric assumptions on the sampling distribution of our estimates. Second, its flexibility lended itself well to the the complex estimation procedure for obtaining DiD estimates based on the transition probability matrix, as described above. That is, our DiD estimates were based on linear combinations of transition matrices and occupancy vectors—so producing analogous estimates 495 times was a fairly trivial exercise. Third, we demonstrated using a simulation study that this method yielded Type I error rate in line with expectations. By comparison, other approaches yielded high false discovery rates under standard hypothesis testing with α=0.05.

The results of our simulation study are summarized in eTable11 below. The table provides rejection rates for alternative inference methods based our simulation study. For this simulation study, we specified a DiD data generation process under Equation 1 above and with a null treatment effect. We then then simulated M = 1,000 datasets, each with N = 10,000 observations, in which we drew 4 treated clusters from a total of 12 clusters (states). That is, this simulation study replicated key aspects of our observational data and analytic approach. We then fit a DiD model based on equation (1) to these simulated data and performed hypothesis testing using several widely used methods:

- Standard inference methods (i.e., no adjustment for clustering of treatment within groups).
- Cluster-robust inference methods (i.e., clustering standard errors at the state level)
- Exact inference via cluster permutation

Rejection rates (α=0.05) are summarized in the table below:

eTable11. Comparison of Rejection Rates Under Alternative Inference Methods

Inference Method	Rejection Rate (α=0.05)
Standard Inference	0.93
Cluster Robust Inference	0.113
Cluster Permutation	0.042

As shown in eTable11, our cluster permutation procedure produced rejection rates of 4.2%--slightly conservative, but in line with expectations when α=0.05. By comparison, the other inference methods yield rejection rates far above 5%.

Statistical Software and Replication Code

Software and versions of all programs and packages used in the analysis are provided below. In addition, replication code for all data construction, analysis, and figures and tables is available upon request. SCCS data are unable to be provided directly to interested researchers. Researchers interesting in utilizing the SCCS data can submit an ancillary study request using the submission tool available at https://www.southerncommunitystudy.org/.

```
R version 3.5.1 (2018-07-02)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS High Sierra 10.13.6

Matrix products: default
BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBLAS.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
 [1] parallel  tcltk     grid      stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] naniar_0.3.1            UpSetR_1.3.3            styler_1.0.2            lubridate_1.7.4         furrr_0.1.0             future_1.8.1            tictoc_1.0
 [8] doSNOW_1.0.16           snow_0.4-2              ri_0.9                  nnet_7.3-12             doParallel_1.0.11       iterators_1.0.9         foreach_1.4.4
[15] gridExtra_2.3           kableExtra_0.9.0        patchwork_0.0.1         broom_0.4.5             plotly_4.7.1            here_0.1                cem_1.1.19
[22] rgenoud_5.8-2.0         Matching_4.9-3          MASS_7.3-50             quantreg_5.36           knitr_1.20              directlabels_2018.05.22 survminer_0.4.2
[29] ggpubr_0.1.7            DiagrammeR_1.0.0        VIM_4.7.0               data.table_1.11.4       colorspace_1.3-2        mice_3.1.3              ggthemes_3.5.0
[36] stargazer_5.2.2         janitor_1.0.0           haven_1.1.2             lintr_1.0.2             jsonlite_1.5            scales_0.5.0            readxl_1.1.0
[43] magrittr_1.5            rms_5.1-2               SparseM_1.77            Hmisc_4.1-1             Formula_1.2-3           survival_2.42-6         lattice_0.20-35
[50] MatchIt_3.0.2           usethis_1.3.0           rstudioapi_0.7          repurrrsive_0.1.0       testthat_2.0.0          shiny_1.1.0             roxygen2_6.0.1
[57] rlang_0.2.1             devtools_1.13.6         covr_3.1.0              forcats_0.3.0           readr_1.1.1             stringr_1.3.1           tidyr_0.8.1
[64] dplyr_0.7.6             tibble_1.4.2            purrr_0.2.5             ggplot2_3.0.0           sccs_0.0.0.9000

loaded via a namespace (and not attached):
  [1] tidyselect_0.2.4    lme4_1.1-17         htmlwidgets_1.2     combinat_0.0-8      munsell_0.5.0       codetools_0.2-15    withr_2.1.2         rgexf_0.15.3
  [9] robustbase_0.93-1   vcd_1.4-4           Rttf2pt1_1.3.7      listenv_0.7.0       mnormt_1.5-5        KMsurv_0.1-5        rprojroot_1.3-2     downloader_0.4
 [17] TH.data_1.0-9       randomForest_4.6-14 R6_2.2.2            rex_1.1.2           assertthat_0.2.0    promises_1.0.1      multcomp_1.4-8      gtable_0.2.0
 [25] globals_0.12.1      sandwich_2.4-0      MatrixModels_0.4-1  cmprsk_2.2-7        splines_3.5.1       extrafontdb_1.0     lazyeval_0.2.1      acepack_1.4.1
 [33] brew_1.0-6          checkmate_1.8.5     reshape2_1.4.3      abind_1.4-5         backports_1.1.2     httpuv_1.4.4.2      extrafont_0.17      tools_3.5.1
 [41] influenceR_0.1.0    psych_1.8.4         RColorBrewer_1.1-2  Rcpp_0.12.17        plyr_1.8.4          base64enc_0.1-3     visNetwork_2.0.4    rpart_4.1-13
 [49] viridis_0.5.1       zoo_1.8-2           cluster_2.0.7-1     openxlsx_4.1.0      lmtest_0.9-36       mvtnorm_1.0-8       packrat_0.4.9-3     mitml_0.3-5
 [57] hms_0.4.2           mime_0.5            evaluate_0.10.1     xtable_1.8-2        XML_3.98-1.11       rio_0.5.10          compiler_3.5.1      minqa_1.2.4
 [65] htmltools_0.3.6     later_0.7.3         visdat_0.5.1        expm_0.999-2        boot_1.3-20         Matrix_1.2-14       car_3.0-0           quadprog_1.5-5
 [73] bindr_0.1.1         pan_1.6             igraph_1.2.1        pkgconfig_2.0.1     km.ci_0.5-2         foreign_0.8-70      laeken_0.4.6        sp_1.3-1
 [81] xml2_1.2.0          rvest_0.3.2         digest_0.6.16       rmarkdown_1.10      cellranger_1.1.0    survMisc_0.5.5      htmlTable_1.12      Rook_1.1-1
 [89] curl_3.2            commonmark_1.5      jomo_2.6-2          nloptr_1.0.4        nlme_3.1-137        bindrcpp_0.2.2      carData_3.0-1       viridisLite_0.3.0
 [97] pillar_1.2.3        httr_1.3.1          DEoptimR_1.0-8      glue_1.2.0          zip_1.0.0           class_7.3-14        stringi_1.2.4       polspline_1.1.13
[105] latticeExtra_0.6-28 memoise_1.1.0       e1071_1.6-8
```

Supplemental Exhibits and Figures

eTable12. Individual State Samples in the Analytic SCCS Sample

	Sample Size
Alabama	3317
Arkansas	962
Florida	1048
Georgia	1559
Kentucky	1394
Louisiana	601
Mississippi	2541
North Carolina	433
South Carolina	814
Tennessee	1194
Virginia	773
West Virginia	720

eFigure7 Kaplan Meier Survival Curves by State Medicaid Expansion Status

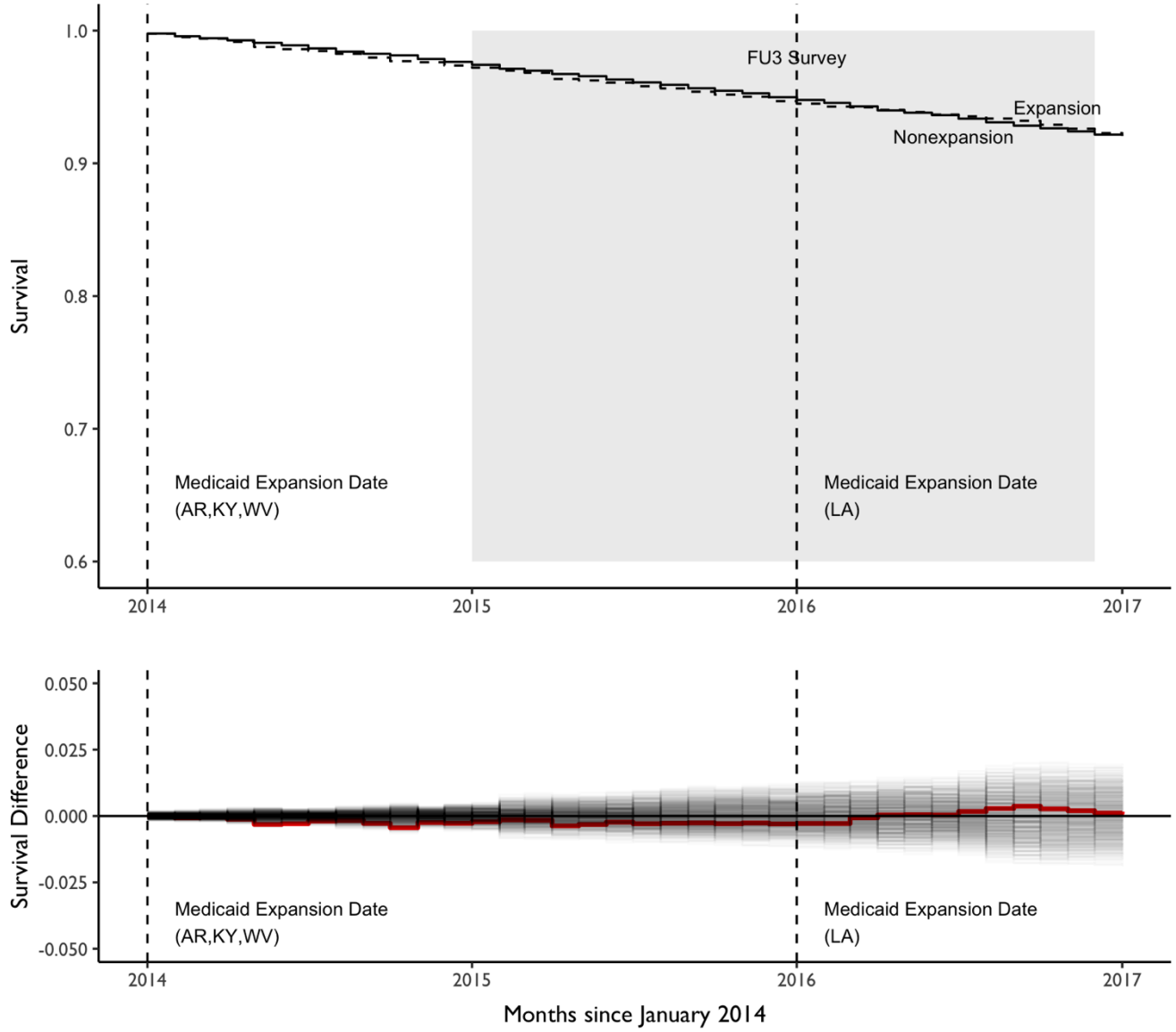
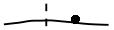


Figure shows non-parametric (Kaplan-Meier) survival curves separately by state Medicaid expansion status. Bottom panel shows difference in survival (red line) between expansion and non-expansion groups, along with estimated survival differences for all 495 possible permutations of 4 expansion and 8 non-expansion states (grey lines).

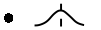
eTable13. Differential changes in SF-12 health outcomes among SCCS participants in Medicaid expansion states, as compared with those in non-expansion states

	Expected Percent Experiencing Change	Diff–in–Diff Estimate	Inference Sparkline	Percentile Rank
General health status				
Excellent	4.5	0.32		86.7
Very Good	17.8	–0.29		25.9
Good	33.8	0.32		65.3
Fair	27.0	0.55		74.5
Poor	9.4	–0.67		7.7
Health limits in moderate activities				
Not limited at all	33.9	2.6		99.6
Limited a little	31.7	–2.3		3.2
Limited a lot	26.9	–0.16		45.7
Health limits in climbing several flights of stairs				
Not limited at all	35.2	–1.4		9.9
Limited a little	26.1	2.1		89.9
Limited a lot	31.2	–0.42		39.4
Accomplished less than you would like				
None of the Time	23.4	1.9		97.6
A Little of the Time	14.3	–0.27		39.2
Some of the Time	21.6	0.95		84.4
Most of the Time	18.0	–1.8		8.1
All of the Time	15.0	–0.48		26.1

NOTE: Table continues on next page

The Percent at Baseline column lists the percentage of the Medicaid expansion state sample in each category in the pre-expansion period, which is based on the follow-up 1 survey (FU1) fielded from 2008-2013. Medicaid expansion states include Arkansas, Kentucky, Louisiana and West Virginia while the non-expansion state sample includes Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (which had not yet expanded during the FU3 surveying period). The difference-in-differences (DiD) column provides estimates of the differential change in the category attributed to Medicaid expansion.
Inference sparklines plot the density distribution of DiD estimates for all 495 possible state expansion status permutations; the observed estimate (i.e., the estimate in the DiD column) is also denoted by a solid point in the inference sparkline. The percentile ranking of the DiD estimate in the permutation distribution is provided in the last column; the most extreme of the 495 values receive percentile rankings of 0.01 and 100.0.

NOTE: This table continues the table from the previous page

	Expected Percent Experiencing Change	Diff–in–Diff Estimate	Inference Sparkline	Percentile Rank
Limited in work or other activities				
None of the Time	22.8	2.2		100.0
A Little of the Time	11.5	0.21		57.8
Some of the Time	16.2	1.3		92.1
Most of the Time	13.8	0.36		64.0
All of the Time	28.1	–3.8		0.2
Accomplished less as a result of any emotional problems				
None of the Time	34.6	2.4		97.6
A Little of the Time	15.0	0.82		79.4
Some of the Time	19.5	–0.58		28.7
Most of the Time	12.6	–1.3		3.0
All of the Time	10.9	–1.1		1.2
Conduct activities less carefully as result of emotional problems				
None of the Time	35.1	3.0		99.8
A Little of the Time	12.7	0.59		74.5
Some of the Time	16.1	0.60		72.3
Most of the Time	12.6	–0.89		3.2
All of the Time	16.0	–3.1		0.2
Pain interfere with your normal work				
Not at all	19.4	1.0		95.6
A little bit	19.8	0.26		60.2
Moderately	15.7	0.61		73.9
Quite a bit	22.8	–1.1		14.3
Extremely	14.8	–0.56		14.1

NOTE: Table continues on next page

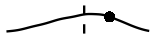
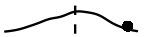
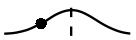
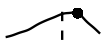
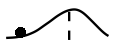
The Percent at Baseline column lists the percentage of the Medicaid expansion state sample in each category in the pre-expansion period, which is based on the follow-up 1 survey (FU1) fielded from 2008-2013. Medicaid expansion states include Arkansas, Kentucky, Louisiana and West Virginia while the non-expansion state sample includes Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (which had not yet expanded during the FU3 surveying period). The difference-in-differences (DiD) column provides estimates of the differential change in the category attributed to Medicaid expansion.
Inference sparklines plot the density distribution of DiD estimates for all 495 possible state expansion status permutations; the observed estimate (i.e., the estimate in the DiD column) is also denoted by a solid point in the inference sparkline. The percentile ranking of the DiD estimate in the permutation distribution is provided in the last column; the most extreme of the 495 values receive percentile rankings of 0.01 and 100.0.

NOTE: This table continues the table from the previous page

	Expected Percent Experiencing Change	Diff–in–Diff Estimate	Inference Sparkline	Percentile Rank
Feel calm and peaceful				
None of the Time	4.7	0.21		79.6
A Little of the Time	14.8	–0.35		30.9
Some of the Time	26.3	0.42		64.0
Most of the Time	35.8	1.5		90.5
All of the Time	10.9	–1.6		0.2
Have a lot of energy				
None of the Time	11.2	2.6		100.0
A Little of the Time	24.6	–0.81		12.7
Some of the Time	32.1	–2.4		0.2
Most of the Time	20.3	0.72		75.2
All of the Time	4.3	0.07		55.8
Feel downhearted and depressed				
None of the Time	27.2	1.3		95.4
A Little of the Time	25.7	–0.17		41.2
Some of the Time	21.5	0.34		64.6
Most of the Time	11.1	–0.78		8.1
All of the Time	7.0	–0.49		27.3
Physical health or emotional problems interfered with social activities				
None of the Time	30.7	1.1		94.7
A Little of the Time	15.7	–0.12		41.0
Some of the Time	22.4	0.05		55.2
Most of the Time	16.4	–1.4		0.2
All of the Time	7.2	0.56		82.6

The Percent at Baseline column lists the percentage of the Medicaid expansion state sample in each category in the pre-expansion period, which is based on the follow-up 1 survey (FU1) fielded from 2008-2013. Medicaid expansion states include Arkansas, Kentucky, Louisiana and West Virginia while the non-expansion state sample includes Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (which had not yet expanded during the FU3 surveying period). The difference-in-differences (DiD) column provides the estimated of the differential change in the category attributed to Medicaid expansion.
Inference sparklines plot the density distribution of DiD estimates for all 495 possible state expansion status permutations; the observed estimate (i.e., the estimate in the DiD column) is also denoted by a solid point in the inference sparkline. The percentile ranking of the DiD estimate in the permutation distribution is provided in the last column; the most extreme of the 495 values receive percentile rankings of 0.01 and 100.0.

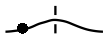
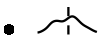
eTable14. Differential Changes in Health Status Transitions Among SCCS Participants in Medicaid expansion states, as compared with those in non-expansion states

	Expected Percent Experiencing Change	Diff–in–Diff Estimate	Inference Sparkline	Percentile Rank
General health status				
No Change	46.9	1.0		79.6
Health Improvement	22.0	0.05		52.3
Health Decline	31.1	−1.1		9.9
Health limits in moderate activities				
No Change	51.1	2.2		98.8
Health Improvement	15.9	−0.41		28.1
Health Decline	33.0	−1.7		2.4
Health limits in climbing several flights of stairs				
No Change	48.6	3.0		100.0
Health Improvement	22.2	−1.8		0.2
Health Decline	29.2	−1.2		10.1
Accomplished less than you would like				
No Change	36.5	0.64		75.2
Health Improvement	24.4	1.4		94.3
Health Decline	39.1	−2.0		1.2

NOTE: Table continues on next page

The Percent Experiencing Change column lists the percentage of the Medicaid expansion state sample that experienced the health status change from the pre-expansion to post-expansion period. Medicaid expansion states include Arkansas, Kentucky, Louisiana and West Virginia while the non-expansion state sample includes Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (which had not yet expanded during the FU3 surveying period). The difference-in-differences (DiD) column provides estimates of the differential change in the category attributed to Medicaid expansion.
Inference sparklines plot the density distribution of DiD estimates for all 495 possible state expansion status permutations; the observed estimate (i.e., the estimate in the DiD column) is also denoted by a solid point in the inference sparkline. The percentile ranking of the DiD estimate in the permutation distribution is provided in the last column; the most extreme of the 495 values receive percentile rankings of 0.01 and 100.0.

NOTE: This table continues the table from the previous page

	Expected Percent Experiencing Change	Diff–in–Diff Estimate	Inference Sparkline	Percentile Rank
Limited in work or other activities				
No Change	33.8	2.6		100.0
Health Improvement	20.5	1.3		99.2
Health Decline	45.7	–3.9		0.2
Accomplished less as a result of any emotional problems				
No Change	39.2	1.4		88.7
Health Improvement	25.5	0.84		93.7
Health Decline	35.3	–2.2		2.8
Conduct activities less carefully as a result of emotional problems				
No Change	36.3	2.9		100.0
Health Improvement	21.7	1.2		99.8
Health Decline	42.0	–4.1		0.2
Pain interfere with your normal work				
No Change	37.3	0.89		81.6
Health Improvement	23.6	0.35		68.5
Health Decline	39.1	–1.2		0.8

NOTE: Table continues on next page

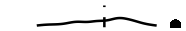
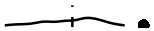
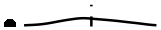
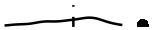
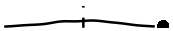
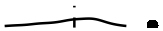
The Percent Experiencing Change column lists the percentage of the Medicaid expansion state sample that experienced the health status change from the pre-expansion to post-expansion period. Medicaid expansion states include Arkansas, Kentucky, Louisiana and West Virginia while the non-expansion state sample includes Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (which had not yet expanded during the FU3 surveying period). The difference-in-differences (DiD) column provides estimates of the differential change in the category attributed to Medicaid expansion. Inference sparklines plot the density distribution of DiD estimates for all 495 possible state expansion status permutations; the observed estimate (i.e., the estimate in the DiD column) is also denoted by a solid point in the inference sparkline. The percentile ranking of the DiD estimate in the permutation distribution is provided in the last column; the most extreme of the 495 values receive percentile rankings of 0.01 and 100.0.

NOTE: This table continues the table from the previous page

	Expected Percent Experiencing Change	Diff–in–Diff Estimate	Inference Sparkline	Percentile Rank
Feel calm and peaceful				
No Change	38.2	2.5		100.0
Health Improvement	22.8	–0.37		31.1
Health Decline	39.0	–2.1		0.2
Have a lot of energy				
No Change	38.5	0.14		53.7
Health Improvement	26.5	1.4		98.4
Health Decline	35.0	–1.5		0.2
Feel downhearted and depressed				
No Change	36.6	0.48		75.4
Health Improvement	30.8	0.05		50.9
Health Decline	32.6	–0.54		30.5
Physical health or emotional problems interfered with social activities				
No Change	39.3	–0.95		17.8
Health Improvement	26.6	1.4		97.2
Health Decline	34.1	–0.43		31.7

The Percent Experiencing Change column lists the percentage of the Medicaid expansion state sample that experienced the health status change from the pre-expansion to post-expansion period. Medicaid expansion states include Arkansas, Kentucky, Louisiana and West Virginia while the non-expansion state sample includes Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (which had not yet expanded during the FU3 surveying period). The difference-in-differences (DiD) column provides estimates of the differential change in the category attributed to Medicaid expansion. Inference sparklines plot the density distribution of DiD estimates for all 495 possible state expansion status permutations; the observed estimate (i.e., the estimate in the DiD column) is also denoted by a solid point in the inference sparkline. The percentile ranking of the DiD estimate in the permutation distribution is provided in the last column; the most extreme of the 495 values receive percentile rankings of 0.01 and 100.0.

eTable15. Robustness of Overall Composite Summary Estimates to Alternative Sample and Specification Choices

	Expected Percent Experiencing Change	Diff–in–Diff Estimate	Inference Sparkline	Percentile Rank
Primary Sample (matches main text)				
No Change	40.2	1.4		100.0
Health Improvement	23.5	0.45		90.3
Health Decline	36.3	−1.8		0.2
Add Race x Time Interaction				
No Change	40.2	1.4		100.0
Health Improvement	23.8	0.19		71.3
Health Decline	36.0	−1.6		0.2
No Demographic or Health Controls				
No Change	40.2	1.4		100.0
Health Improvement	23.6	0.38		84.4
Health Decline	36.2	−1.7		0.2
Parallel Trends Sample				
No Change	37.1	1.6		100.0
Health Improvement	21.8	0.57		98.6
Health Decline	41.1	−2.1		0.2
Excluding Louisiana				
No Change	40.2	1.5		100.0
Health Improvement	23.5	0.72		100.0
Health Decline	36.2	−2.3		0.2

The Percent Experiencing Change column lists the percentage of the Medicaid expansion state sample that experienced the health status change from the pre-expansion to post-expansion period. Medicaid expansion states include Arkansas, Kentucky, Louisiana and West Virginia while the non-expansion state sample includes Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (which had not yet expanded during the FU3 surveying period). The difference-in-differences (DiD) column provides estimates of the differential change in the category attributed to Medicaid expansion. Inference sparklines plot the density distribution of DiD estimates for all 495 (full sample) or 165 (excluding LA) possible state expansion status permutations; the observed estimate (i.e., the estimate in the DiD column) is also denoted by a solid point in the inference sparkline. The percentile ranking of the DiD estimate in the permutation distribution is provided in the last column; the most extreme of the 495 values receive percentile rankings of 0.01 and 100.0.

1. Signorello LB, Hargreaves MK, Blot WJ. The southern community cohort study: investigating health disparities. Journal of health care for the poor and underserved 2010;21(1):26.

2. Johnson JA, Maddigan SL. Performance of the RAND-12 and SF-12 summary scores in type 2 diabetes. Quality of Life Research 2004;13(2):449–456.

3. Johnson JA, Coons SJ. Comparison of the EQ-5D and SF-12 in an adult US sample. Quality of Life Research 1998;7(2):155–166.

4. Imbens G, Rubin D. Causal Inference in Statistics, and In the Social and Biomedical Sciences. Cambridge University Press; 2010.

5. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? International journal of methods in psychiatric research 2011;20(1):40–49.

6. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine 2011;30(4):377–399.

7. Wooldridge J. Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge, MA: MIT Press; 2010.

8. Heckman JJ. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence. University of Chicago Center for Mathematical studies in Business and Economics; 1987.

9. Athey S, Imbens GW. Identification and inference in nonlinear difference-in-differences models. Econometrica 2006;74(2):431–497.

10. King G, Tomz M, Wittenbert J. Making the most of statistical analyses: Improving interpretation and presentation. American Journal of Political Science 2000;44:341–55.

11. Stuart EA, Huskamp HA, Duckworth K, et al. Using propensity scores in difference-in-differences models to estimate the effects of a policy change. Health Serv Outcomes Res Methodol 2014;14(4):166–82.

12. Daw JR, Hatfield LA. Matching and Regression to the Mean in Difference-in-Differences Analysis. Health services research 2018;

13. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS®. 2012;6.

14. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci 2010;25(1):1–21.

15. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res 2011;46(3):399–424.

16. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician 1985;39(1):33–38.

17. Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death." Journal of Educational and Behavioral Statistics 2003;28(4):353–368.

18. Sommers BD, Baicker K, Epstein AM. Mortality and Access to Care among Adults after State Medicaid Expansions. New England Journal of Medicine 2012;367(11):1025–34.

19. Sommers BD, Long SK, Baicker K. Changes in Mortality After Massachusetts Health Care Reform: A Quasi-experimental Study. Annals of Internal Medicine 2014;160(9):585.

20. Polsky D, Doshi JA, Escarce J, et al. The health effects of Medicare for the near-elderly uninsured. Health services research 2009;44(3):926–945.

21. Polsky D, Doshi JA, Manning WG, et al. Response to McWilliams commentary:"Assessing the health effects of medicare coverage for previously uninsured adults: A matter of life and death?" Health services research 2010;45(5p1):1423–1429.

22. Hays RD, Sherbourne CD, Mazel RM. The rand 36-item health survey 1.0. Health economics 1993;2(3):217–227.

23. McWilliams JM, Meara E, Zaslavsky AM, Ayanian JZ. Commentary: assessing the health effects of Medicare coverage for previously uninsured adults: a matter of life and death? Health services research 2010;45(5p1):1407–1422.

24. Ding P, Lu J. Principal stratification analysis using principal scores. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2017;79(3):757–777.

25. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics 2002;58(1):21–29.

26. Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death." Journal of Educational and Behavioral Statistics 2003;28(4):353–368.

27. Chaix B, Evans D, Merlo J, Suzuki E. Commentary: Weighing up the Dead and Missing. Epidemiology 2012;23(1):129–31.

28. Cameron AC, Miller DL. A practitioner's guide to cluster-robust inference. Journal of Human Resources 2015;50(2):317–372.

29. Cameron AC, Miller DL. Robust inference with clustered data. Working Papers, University of California, Department of Economics; 2010.

30. Rosenbaum PR. Observational studies [Internet]. In: Observational Studies. Springer; 2002 [cited 2016 Oct 11]. p. 1–17.Available from: http://link.springer.com/chapter/10.1007/978-1-4757-3692-2_1

31. Garthwaite C, Gross T, Notowidigdo M, Graves JA. Insurance expansion and hospital emergency department access: evidence from the Affordable Care Act. Annals of internal medicine 2017;166(3):172–179.

32. Conley TG, Taber CR. Inference with "difference in differences" with a small number of policy changes. The Review of Economics and Statistics 2011;93(1):113–125.

33. Rokicki S, Cohen J, Fink G, Salomon JA, Landrum MB. Inference With Difference-in-Differences With a Small Number of Groups. Medical care 2018;56(1):97–105.

34. Lehmann EL, D'Abrera HJ. Nonparametrics: statistical methods based on ranks. 1975;