NBER WORKING PAPER SERIES

DOES UNIVERSAL COVERAGE IMPROVE HEALTH? THE MASSACHUSETTS EXPERIENCE

Charles J. Courtemanche
Daniela Zapata

Working Paper 17893
http://www.nber.org/papers/w17893

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2012

We thank James Cunningham, David Frisvold, Michael Grossman, Jon Gruber, Stephen Holland, Ted Joyce, Dave Ribar, Chris Ruhm, Ken Snowden, Chris Swann, Rusty Tchernis, Joe Terza, Nicolas Ziebarth, and seminar participants at Cornell University, Georgia State University, the University of Georgia, the University of North Carolina at Greensboro, Yale University, the National Bureau of Economic Research Spring Health Care Meeting, and the Southern Economic Association Annual Meeting for valuable comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Does Universal Coverage Improve Health? The Massachusetts Experience
Charles J. Courtemanche and Daniela Zapata
NBER Working Paper No. 17893
March 2012, Revised January 2013
JEL No. I12,I13,I18

ABSTRACT

In 2006, Massachusetts passed health care reform legislation designed to achieve nearly universal coverage through a combination of insurance market reforms, mandates, and subsidies that later served as the model for national reform. Using data from the Behavioral Risk Factor Surveillance System, we provide evidence that health care reform in Massachusetts led to better overall self-assessed health. Various robustness checks and placebo tests support a causal interpretation of the results. We also document improvements in several determinants of overall health: physical health, mental health, functional limitations, joint disorders, and body mass index. Next, we show that the effects on overall health were strongest among those with low incomes, non-whites, near-elderly adults, and women. Finally, we use the reform to instrument for health insurance and estimate a sizeable impact of coverage on health.

Charles J. Courtemanche
Georgia State University
Andrew Young School of Policy Studies
Department of Economics
P.O. Box 3992
Atlanta, GA 30302-3992
and NBER
ccourtemanche@gsu.edu

Daniela Zapata
University of North Carolina at Greensboro
Department of Economics
P.O. Box 26170
Greensboro, NC 27402
d_zapata@uncg.edu

I. Introduction

A major objective of the Patient Protection and Affordable Care Act (ACA) signed into law in March of 2010 is to increase health insurance coverage in the United States to nearly universal levels through a combination of insurance market reforms, mandates, subsidies, and Medicaid expansions. Although the mandates survived constitutional challenges and the 2012 election outcomes make full repeal unlikely, the implementation of the law still faces challenges, such as the denial of funds during the budgetary process or some states opting not to expand Medicaid. This ongoing debate highlights the need for projections of the law's impacts on health, health care utilization, and state and federal budgets. The multi-faceted nature of the reform and breadth of the population affected suggests that evidence from coverage expansions in other contexts will be of only limited usefulness.

The most similar intervention to date to the ACA is the Massachusetts health care reform of April 2006, entitled "An Act Providing Access to Affordable, Quality, Accountable Health Care" and commonly called "Chapter 58" (Long, 2008).[1] The law enabled Massachusetts to lower its uninsurance rate to 2% by 2010 through a strategy called "incremental universalism," or "filling the gaps in the existing system … rather than ripping up the system and starting over" (Massachusetts' Division of Health Care, Finance and Policy, 2010; Gruber, 2008a:52). Gruber (2010) describes Massachusetts' approach to incremental universalism as a "three legged stool" of insurance market reforms, mandates, and subsidies/Medicaid expansions (Gruber, 2010).

The first leg of the stool reforms non-group insurance markets in an effort to ensure the availability of coverage for those without access to employer-provided or public insurance. Insurers are not allowed to deny or drop coverage based on pre-existing conditions (guaranteed

[1] For a more detailed description of the law, see Long (2008), McDonough et al. (2006) and Gruber (2008a, 2008b).

issue) or vary premiums to reflect health status aside for limited adjustments for age and smoking status (community rating) (Kirk, 2000; McDonough et al., 2006). A health insurance exchange, the Commonwealth Health Insurance Connector Authority, offers plans developed by licensed health insurance companies for those without access to group markets. Enrollment on the Connector began in October 2006 for those with incomes below 100% of the federal poverty line (FPL), in January 2007 for those up to 300% FPL, and in May 2007 for everyone else. Additionally, private health insurance plans are required to provide coverage for young adults on their parents' plans for up to two years after they are no longer dependents or until their 26th birthday (McDonough et al., 2006).[2]

This first leg alone would likely lead to adverse selection and a "death spiral" with rising premiums gradually driving healthy individuals out of the non-group market. The second leg of the three-legged stool therefore involves mandates requiring adults to be covered by health insurance and employers to provide health insurance. Individuals without adequate coverage face a penalty of half of the lowest premium they would have paid in a Health Connector-certified plan. Employers with more than 10 employees must make a "fair and reasonable" contribution toward an employer health insurance plan or pay a state assessment of up to $295 per full-time equivalent worker per year (Massachusetts Health Insurance Connector Authority, 2008).[3] The mandates took effect in July 2007.

To help low- and middle-income households be financially able to comply with the mandate, the third leg of the Massachusetts reform provides subsidies and Medicaid expansions.

[2] Guaranteed issue and community rating have been in place in Massachusetts since 1996. The 1996 law only allowed premiums to vary with age and geography; Chapter 58 further allowed them to vary with tobacco use. The insurance exchange and the requirement regarding young adults on their parents' plans both started with Chapter 58.
[3] Minimum requirements plans must meet to satisfy the mandates include coverage for prescription drugs and preventive and primary care, as well as maximums on deductibles and out-of-pocket spending.

Chapter 58 specifies that health insurance be free for people below 150% FPL and that premiums be subsidized on a sliding scale between 150% and 300% FPL with no deductibles.[4] The reform also expands Medicaid to cover children below 300% FPL (McDonough et al., 2006).

Taking into account the costs of the subsidies and Medicaid expansions as well as the savings from reduced safety net payments, Raymond (2009) estimates the annual fiscal cost of the reform to be $707 million. Through a waiver allowing for a more flexible use of federal Medicaid matching money, half of this amount comes from the federal government, leaving the state government's share at $353 million.

Table 1 compares Massachusetts' approach to incremental universalism with that of the Affordable Care Act. Though there are differences in some of the details, both the Massachusetts and national reforms were clearly motivated by the same "three-legged stool" approach to incremental universalism. Both featured guaranteed issue, community rating, insurance exchanges, mandates, Medicaid expansions, and subsidies. For these reasons, analyzing the effects of health care reform in Massachusetts provides the best available predictor to date of the implications of the Affordable Care Act.

Given that recent nature of the Massachusetts reform, researchers are only beginning to understand its impacts. Long et al. (2009) find that by 2008 the uninsured rate decreased by 6.6 percentage points for the overall nonelderly population and 17.3 percentage points for lower-income adults.[5] Long and Stockely (2011) find a decrease in unmet medical needs because of cost among lower income adults but also some evidence of delays in care from being unable to find a provider. Yelowitz and Cannon (2010) show that Chapter 58's impact on coverage was

[4] For instance, in 2008 a family with an income between 150% and 200% of the poverty line paid a premium of $35 per adult, while a family with an income in the 250% to 300% range paid $105 per adult.
[5] These results support preliminary evidence found by Long (2008) using information from 2006 and 2007.

mitigated by the crowding out of private insurance. They also investigate the effect on self-assessed health, finding mixed results: an increase in the probability of individuals reporting at least good health but a decrease in the probability of reporting at least very good health. Cogan et al. (2010) estimate that the reform increased employer-sponsored insurance premiums by about 6%. Kolstad and Kowalski (2010) show that the reform reduced levels of uninsurance by 36% among the population of hospital discharges. Length of stay and the number of inpatient admissions originating from the emergency room both decreased, with some evidence also suggesting an increase in the utilization of preventive services, a decline in hospitalizations for preventable conditions, and an improvement in quality of care. Miller (2011a) finds a reduction in non-urgent emergency room visits, consistent with the newly-insured having access to such care in other settings. Miller (2011b) focuses on children's outcomes, finding a substitution from emergency room care to office visits, a reduction in medical needs unmet because of cost, and an increase in the probability of reporting excellent health. Kowalski and Kolstad (2012) exploit the reform's effect on employer-provided health insurance to show that wage reductions almost completely offset the cost of health insurance benefits.

We contribute to this growing literature by conducting the most comprehensive analysis to date of Chapter 58's effect on the self-assessed health of adults. Though many open questions remain about the reform's effectiveness, as Gruber (2011b:190) writes, "the most significant of these is the impact of reform on the health of citizens." We utilize individual-level data from the Behavioral Risk Factor Surveillance System (BRFSS), which allows for the use of longer preand post-treatment periods, a much larger sample, and a broader range of health-related questions than Yelowitz and Cannon (2010), enabling us to obtain clearer results.[6]

First, an ordered probit difference-in-differences analysis shows that the reform increased the probability of individuals reporting excellent or very good health. A variety of robustness checks and placebo tests support a causal interpretation of the results. The estimates imply that annual government spending per adult transitioned into excellent or very good health is $11,465, split evenly between the Massachusetts and federal governments. We then provide evidence that the reform improved several determinants of overall health: physical health, mental health, functional limitations, joint disorders, and body mass index. Next, we examine heterogeneity and find that the reform's effect on overall health was strongest for low-income people, racial minorities, near-elderly adults, and women. Finally, we use the reform to instrument for insurance coverage and estimate a large effect of coverage on health.

II. Health Insurance and Health

An important part of the argument for universal coverage is the assumption that health insurance improves health. As quoted by Yelowitz and Cannon (2010), Levy and Meltzer (2008) write,

> The central question of how health insurance affects health, for whom it matters, and how much, remains largely unanswered at the level of detail needed to inform policy decisions. … Understanding the magnitude of health benefits associated with insurance is not just an academic exercise …, it is crucial to ensuring that the benefits of a given amount of public spending on health are maximized (p. 400).

[6] Specifically, Yelowitz and Cannon (2010) use Current Population Survey supplements and compare a pre-treatment period of 2005-2006 with a post-treatment period of 2008. They conduct a difference-in-differences analysis with other New England states as controls. Their sample size is 41,873. In contrast, we utilize data from 2001-2010 and have a sample size of 2,879,296 in our main analysis and 340,592 using only New England.

This section provides a brief summary of theoretical and empirical research on the topic and summarizes our contribution to this broader literature.

Grossman (1972) models health as a durable capital stock that is also an input in the production of healthy time. Health capital depends on the initial endowment of health, past period health, and past period investments made to preserve it. Medical care and time spent in health producing activities are the main forms of health investment. Every period people face uncertainty as to whether they will be affected by a negative health shock, so they buy health insurance to protect themselves against unexpected medical costs. Because health insurance reduces the price of care faced by the consumer it increases the demand for medical care (Arrow, 1963; Pauly, 1968). This increase in consumption of care could result in better health, but if the additional medical care is redundant health outcomes may remain the same or even deteriorate. This effect is sometimes known as “flat of the curve” medical care, because diminishing returns in the health production function imply that at some point the health gains associated with more medical care may be very small (Doyle, 2005).

The majority of empirical investigations into the relationship between health insurance and health are observational studies that use multivariate regression analysis. A review of these studies by Hadley (2003) shows that 15 out of the 20 published between 1991 and 2001 found a positive association between health insurance coverage and recovery from health conditions such as cancer, trauma, and appendicitis. Health insurance was also associated with better overall health status and lower mortality risk in all of the studies that examined these outcomes. However, these relationships cannot be interpreted as causal because the research designs did not address the potential for unobserved heterogeneity and reverse causality.

During the 1970's the RAND Health insurance experiment randomly assigned families to health insurance plans with coinsurance rates ranging from 0% to 95%, with all medical expenses covered over a threshold. Medical care use increased among people assigned to plans with lower coinsurance rates, but health outcomes only improved among the poor (Manning et al., 1987). However, this experiment only shows the impact of health insurance along the intensive margin from less to more generous coverage, not the extensive margin of no coverage to any coverage. It is also unclear to what extent findings from the 1970s are applicable today.

Some studies have taken advantage of the plausibly exogenous variation provided by public insurance programs like Medicaid and Medicare in order to address the endogeneity of coverage. Currie and Gruber (1996a, 1996b) find that Medicaid expansions decrease infant mortality and low birth weight, while Dafny and Gruber (2005) show that they also reduce avoidable hospitalizations among children. Most recently, Finkelstein et al. (2011) exploit a 2008 Oregon lottery in which winners were given the chance to apply for Medicaid to show that coverage improves self-reported physical and mental health. The randomization allows for clean identification of the causal effects of Medicaid eligibility, at least among the low-income uninsured lottery participants.

Evidence on the effect of Medicare on the health of seniors is mixed. Card et al. (2004) find that obtaining Medicare coverage at age 65 improves the self-assessed health of Hispanics and people with low levels of education; however, the effect for the whole sample is smaller and insignificant. Finkelstein and McKnight (2008) show that 10 years after the introduction of Medicare there was not a statistically significant impact on mortality rates for people older than 65. Card et al. (2009) find more favorable results: a reduction in the 7-day mortality rate among emergency room patients older than 65 compared to those right below that cutoff.

A few studies attempt to estimate the causal effect of insurance on health in contexts other than public programs, again finding mixed results. Pauly (2005) uses marital status and firm size as instruments for private insurance coverage and finds a positive but insignificant effect of insurance on self-reported health and a negative but insignificant effect on the probability of having a chronic condition. Doyle (2005) shows that uninsured patients receive less medical care and have higher mortality rates than insured patients after a random health shock (a car accident).

To summarize, the extant literature suggests that health insurance coverage appears to improve health in some contexts but not others. The uninsured in the U.S. consist of a number of groups, including those too sick to obtain coverage, those too healthy to feel insurance is necessary, and those too poor to afford private coverage but not poor enough to qualify for public insurance programs. Obtaining universal coverage in the U.S. therefore requires coverage expansions across a highly heterogeneous group, making it unclear the extent to which these prior findings are applicable. The Massachusetts health care reform provides a unique opportunity to examine an intervention that affects a variety of different categories of uninsured.

III. Data

Health summarizes a combination of factors that reflect physical and mental well-being. Among the usual indicators used to measure health in empirical investigations are mortality rates, hospitalization rates, and self-assessments of overall health. Our study focuses on self-assessments. At the time this paper was written, state-level mortality information was not publically available for a long enough time after the reform to construct an adequate post-treatment period. Regardless, examining mortality rates alone would not capture incremental improvements in health resulting from, for instance, better treatment for chronic but non-life threatening conditions. Hospitalizations are not an appropriate measure of overall health in this context since, to the extent that hospitalizations are price sensitive, changes in hospitalizations after the reform might simply be due to the lower price faced by the newly insured rather than changes in health.

We estimate the effect of the reform on health self-assessments using the BRFSS, a telephone survey of health and health behaviors conducted by state health departments in collaboration with the Centers for Disease Control and Prevention. The BRFSS, which consists of repeated annual cross sections of randomly-sampled adults, is well suited for our analysis for several reasons. First, the dataset contains the necessary variables, including multiple self-reported health measures, demographic characteristics, and state, month, and year identifiers. Second, the BRFSS spans a long enough time period to examine both post-reform outcomes and pre-reform trends. Third, the BRFSS contains an unusually large number of observations for survey data, and a large sample is critical to obtaining meaningful precision when examining the impact of a state-level program with effects that might be concentrated amongst only a fraction of the population.

We construct the sample as follows. We begin with ten-year window surrounding the reform, 2001 to 2010. The treatment group consists of individuals living in Massachusetts, while the initial control group spans most of the rest of the country. We exclude only California, Maine, Vermont, and Oregon – states that also had (less extensive) health care reforms during the sample period and could therefore be considered "treated" to some extent – as well as Hawaii, the only state not included in the BRFSS in every survey year during the sample period.[7]

[7] Finkelstein et al. (2012) show that the concurrent large-scale Medicaid expansion in Oregon improved self-assessed adult health, while Kolstad and Kowalski (2010) mention the reforms in California, Vermont, and Maine. Results including the five omitted states in the control group are very similar and are available upon request.

Following much of the literature (e.g. Long, 2008; Miller, 2011a; Yelowitz and Cannon, 2010), we focus on the non-elderly adult population ages 18-64. The reform had little effect on health insurance coverage for the elderly, as it specifically prevented individuals eligible for Medicare from purchasing insurance through the Connector (Blue Cross Blue Shield of Massachusetts, 2006). Moreover, the pre-treatment trends in many of our health measures for the elderly are substantially different for Massachusetts and the control states, suggesting that our difference-in-differences identification strategy would not be appropriate for this age group.[8] As we will show, the pre-treatment trends are much more similar for the non-elderly population.

Our main dependent variable is a self-reported health index asking respondents to rate their overall health as poor (0), fair (1), good (2), very good (3), or excellent (4). This index has been previously used by other studies analyzing the impact of health insurance on health (Card et al., 2004; Pauly, 2005; Yelowitz and Cannon, 2010) and has been repeatedly shown to be correlated with objective measures of health such as mortality (e.g. Idler and Benyamini, 1997; DeSalvo et al., 2006; Phillips et al., 2010). According to Idler and Benyamini, another advantage of the index is that it is a global measure of health that captures the full range of diseases and limitations a person may have.

The primary concern with the self-reported health index is its subjective nature. We will be able to flexibly control for the sources of reporting heterogeneity identified in the literature, such as age, income, and gender (Ziebarth, 2010). Nonetheless, the estimated effect of the reform on self-assessed health could still reflect factors beyond objective health. For instance, improved access to medical care might increase awareness about medical conditions, causing one to self-report a lower health status after obtaining insurance coverage, ceteris paribus (Strauss and

[8] Miller (2011a) makes the same observation about pre-treatment trends in the context of the reform's effect on emergency room utilization.

Thomas, 2007). In this case, the reform's effect on self-assessed health would be smaller than its effect on objective health. Alternatively, if the peace of mind from having health insurance influences one's answers to subjective health-related questions, the reform could lead to larger improvements in self-assessed health than objective health.

Consequently, we also utilize a number of other health-related dependent variables in an attempt to verify that the results for the overall self-reported health index are not driven merely by subjectivity. First, we consider number of days out of the past 30 not in good physical health and number of days out of the past 30 not in good mental health. These variables are somewhat less subjective than the overall health index because the respondents are specifically asked to consider a particular component of health. Even less subjective is the next health measure: number of days out of the past 30 with health-related functional limitations. Our last four health-related dependent variables – an indicator for the presence of activity-limiting joint pain, body mass index (BMI), minutes per week of physical activity, and an indicator for whether the individual currently smokes – are quite specific and therefore the least open to subjective interpretation.[9] [10]

We measure coverage with a binary variable reflecting whether or not the individual has "any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare." The BRFSS does not indicate the source of coverage or

[9] BMI=weight in kilograms divided by height in squared meters. Self-reported weight and height are potentially susceptible to biases. Some researchers utilize an adjustment developed by Cawley (2004) that predicts actual height and weight based on self-reported height and weight using the National Health and Nutrition Examination Survey, and then applies the prediction equation to other datasets that only include the self-reported measures. However, studies with BMI as the dependent variable have repeatedly found that applying this adjustment has little influence on the results, so we do not use it here (e.g. Courtemanche et al., 2011).

[10] Minutes per week of physical activity is the sum of minutes of moderate and vigorous physical activity. The BRFSS gives respondents guidance for how to define moderate and vigorous physical activity, reducing the subjectivity of these variables. Moderate activities include "brisk walking, bicycling, vacuuming, gardening, or anything else that causes small increases in breathing or heart rate." Vigorous activities include "running, aerobics, heavy yard work, or anything else that causes large ncreases in breathing or heart rate."

provide any information on premiums, deductibles, or copayments. Finally, we utilize as control variables the BRFSS' information on age, marital status, race, income, education, marital status, and current pregnancy status.

We also include four state-level variables as controls in a robustness check. The first is monthly state unemployment rate, obtained from the Bureau of Labor Statistics. Next, monthly state cigarette excise tax rates come from The Tax Burden on Tobacco (Orzechowski and Walker, 2010) and are adjusted for inflation using the Consumer Price Index for all urban consumers from the Bureau of Labor Statistics. Finally, we use annual state hospital and physician data from the Census Bureau to impute monthly estimates of numbers of hospitals and physicians per 100,000 residents.[11]

Tables 2 and 3 compare the descriptive statistics for Massachusetts and the 45 control states (44 states plus the District of Columbia) in the pre-treatment period of January 2001 through March 2006. Prior to the reform, Massachusetts was already healthier than the rest of the country along most dimensions and had a higher coverage rate. Massachusetts residents averaged higher income and more education than those in other states, and were more likely to be single and white. Massachusetts also had a relatively low unemployment rate, high cigarette tax, high physician density, and low hospital density. These baseline differences illustrate the difficulty in isolating the causal impact of Massachusetts' health care reform. A naïve estimator using only a post-treatment cross section would attribute the entire difference in health between Massachusetts and other states to the reform, including the part of the difference that was already present prior to its enactment. Our empirical analysis will therefore rely on a difference-in-

[11] Monthly estimates were calculated using the formula: $X_{estimate} = X_1 + \frac{n}{12}(X_2 - X_1)$, where X_1 and X_2 are annual estimates, and n is number of months from X_1 to $X_{estimate}$.

differences estimator that controls for pre-treatment differences in state health as well as a number of time-varying observable characteristics.

As a precursor to the regression analysis, Figure 1 plots the average values of the health status index in Massachusetts and the control states every year from 2001 to 2010, along with their 95% confidence intervals. The graph also shows linear pre-treatment trends for Massachusetts and the other states, computed by regressing the mean health index on year plus a constant term. Consistent with the summary statistics from Table 2, Massachusetts residents had better average self-assessed health than those in the control states even before the reform. Despite this difference in baseline levels, the pre-treatment trends in both Massachusetts and the other states were both downward sloping and – critically for the validity of the difference-in-differences approach – close to parallel. The year-to-year fluctuations in the control states in the pre-treatment period are estimated very precisely and lie almost exactly on top of the trend line, while the year-to-year fluctuations in Massachusetts are estimated much less precisely and deviate more substantially. This underscores the importance of utilizing a sufficiently long pre-treatment period in the regression analysis. If, for instance, 2005 – a year in which health in Massachusetts appears to have been below trend – was the only pre-treatment year, a difference-in-differences estimate might capture mean reversion in addition to the causal effect.

Over the period after the reform was passed, 2006-2010, health in the control states remained relatively stable (perhaps dropping slightly) while health in Massachusetts improved. In 2006, when the subsidies and Medicaid expansions took effect in the early stages of the reform's implementation, health increased in both Massachusetts and the control states, but the gain in Massachusetts was larger. In 2007, health dropped in the control states but increased slightly in Massachusetts. Health improved slightly in the control states in 2008 while remaining steady in Massachusetts. In 2009, health improved in both groups but by a larger amount in Massachusetts. Health in 2010 dropped in the control states but held constant in Massachusetts. To more formally investigate whether the health improvement in Massachsuetts relative to the control states was a causal response to health care reform, we next turn to regression analysis. The regression results will broadly support the preliminary findings from Figure 1, although we will see that in a regression context the health benefits did not emerge until 2007.

IV. Regression Analysis

IVa. Baseline Model

We estimate the impact of Massachusetts health care reform on overall self-assessed health status using an ordered probit difference-in-differences model.[12] Suppose the underlying relationship between the covariates and a latent variable representing health (y^*) is given by

$$y^*_{ist} = \beta_0 + \beta_1(MA_s * During_t) + \beta_2(MA_s * After_t) + \boldsymbol{X'_{ist}\beta_3} + \sigma_s + \varphi_t + \varepsilon_{ist} \tag{1}$$

where i, s, and t are indices for individual, state, and month/year combination (e.g. January 2001). MA_s is a dummy variable for whether the respondent lives in Massachusetts. Following Kolstad and Kowalski (2010), we define $During_t$ as a dummy variable equal to 1 from April 2006 to June 2007, the time period after the law had been passed but before all the key provisions had been implemented. $After_t$ is a dummy variable equal to 1 starting in July of 2007, when the final major component of the reform – the individual mandate – took effect. $\boldsymbol{X'_{ist}}$

[12] Given the strong distributional assumptions made by the ordered probit model, we also considered two more flexible approaches to modeling the impact of the reform on health. The first estimates a series of four probits with the dependent variables being indicators for fair or better, good or better, very good or better, and excellent health. The second uses the same dependent variables but estimates linear probability models. The conclusions reached are the same; the results are shown in Appendix Tables A1 and A2.

consists of the age, marital status, race, income, education, and pregnancy variables listed in Table 3.[13] σ_s and φ_t are state and month fixed effects, while ε_{ist} is the error term.

We do not observe y^*_{ist} and instead observe an ordinal health measure y_{ist} such that

$$y_{ist} = \begin{cases} 0 \text{ if } y^*_{ist} \le \kappa_1 \\ 1 \text{ if } \kappa_1 < y^*_{ist} \le \kappa_2 \\ 2 \text{ if } \kappa_2 < y^*_{ist} \le \kappa_3 \\ 3 \text{ if } \kappa_3 < y^*_{ist} \le \kappa_4 \\ 4 \text{ if } y^*_{ist} > \kappa_4 \end{cases} \tag{2}$$

where κ_1 through κ_4 are constants that represent the cut-off points. An ordered probit regression of y_{ist} on the covariates from (1) computes the following probabilities of being in each of the five health states:

$$\Pr(y_{ist} = 0) = \Phi(\lambda_1 - \beta_1(MA_s * During_t) - \beta_2(MA_s * After_t) - \boldsymbol{X'_{ist}\beta_3} - \sigma_s - \varphi_t) \tag{3}$$

$$\begin{aligned} &\Pr(y_{ist} = k) \\ &= \Phi\big(\lambda_j - \beta_1(MA_s * During_t) - \beta_2(MA_s * After_t) - \boldsymbol{X'_{ist}\beta_3} - \sigma_s - \varphi_t\big) \\ &- \Phi\big(\lambda_{j-1} - \beta_1(MA_s * During_t) - \beta_2(MA_s * After_t) - \boldsymbol{X'_{ist}\beta_3} - \sigma_s - \varphi_t\big)\ \forall\, j \in (2,3,4) \end{aligned} \tag{4}$$

$$\Pr(y_{ist} = 4) = 1 - \Phi(\lambda_4 - \beta_1(MA_s * During_t) - \beta_2(MA_s * After_t) - \boldsymbol{X'_{ist}\beta_3} - \sigma_s - \varphi_t) \tag{5}$$

where $\lambda_j = \kappa_j - \beta_0$, the cutoff points adjusted for the constant term. The coefficient of interest is β_2, which captures the difference between the change in Massachusetts from the "before" to the "after" period and the change in the control states from the "before" to the "after" period – in other words, the "difference in differences."

Computing treatment effects in non-linear models has been the source of confusion in the literature. Ai and Norton (2003) showed that the cross difference in a nonlinear model is different from the marginal effect on the interaction term, and could even be the opposite sign.

[13] The controls are especially important in our context, as the Massachusetts reform could have caused uninsured individuals to move to the state, changing the underlying demographic composition of the state's residents and potentially confounding the estimation of the reform's causal effect. (The pooled cross-sectional nature of the BRFSS does not allow us to exclude Massachusetts residents who moved there after the reform.) Indeed, using the control variables as dependent variables shows that after the reform Massachusetts' population became older, less white, and poorer, relative to the control states. Accordingly, excluding the control variables from equation (1) attenuates the estimated effect of the reform somewhat (although the sign and significance level remains the same).

However, Puhani (2008) showed that the cross difference identified by Ai and Norton (2003) is not the same as the treatment effect, and that when the treatment effect is the parameter of interest it is appropriate to focus on the coefficient of the interaction term. A similar observation has been made by Terza (2012). Following Puhani (2008), our "treatment effect on the treated" is given by

$$\tau(After = 1, MA = 1) = E[Y^1|After = 1, MA = 1, \boldsymbol{X}, \varphi] - E[Y^0|After = 1, MA = 1, \boldsymbol{X}, \varphi] \quad (6)$$

where Y^1 and Y^0 are potential outcomes with and without treatment. The "average treatment effect on the treated" is the mean of this treatment effect across those individuals living in Massachusetts in the "after" period (July 2007 through December 2009).

Because of the nonlinearity of the model, the treatment effect depends on the value of the other covariates. The effects of the reform on the probabilities of being in each of the five health states among the treated are

$$\tau_{\boldsymbol{i,MA,t}}(y = 0) = \Phi\left(\lambda_1 - \beta_2 - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right) - \Phi\left(\lambda_1 - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right) \quad (7)$$

$$\begin{aligned} &\tau_{\boldsymbol{i,MA,t}}(y = j) \\ &= \left[\Phi\left(\lambda_j - \beta_2 - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right) - \Phi\left(\lambda_{j-1} - \beta_2 - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right)\right] \\ &- \left[\Phi\left(\lambda_j - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right) - \Phi\left(\lambda_{j-1} - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right)\right] \\ &\forall\, j \in (2,3,4) \end{aligned} \quad (8)$$

$$\begin{aligned} &\tau_{\boldsymbol{i,MA,t}}(y = 4) \\ &= 1 - \Phi\left(\lambda_4 - \beta_2 - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right) - \left[1 - \Phi\left(\lambda_4 - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right)\right] \\ &= \Phi\left(\lambda_4 - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right) - \Phi\left(\lambda_4 - \beta_2 - \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\beta_3} - \sigma_{MA} - \varphi_t\right) \end{aligned} \quad (9)$$

where the state subscript s has been replaced by MA for Massachusetts, and t is restricted to the "after" period.

The key identifying assumption in the difference-in-differences model is that $MA_s * During_t$ and $MA_s * After_t$ are uncorrelated with the error term. In other words, the estimates can be interpreted as causal effects of the reform if we assume that in the absence of the reform changes over time in health would have been the same in Massachusetts and the control states, conditional on the control variables. The similarity of Massachusetts' pre-treatment trend in health to that of the other states shown in Figure 1 provides preliminary support for this assumption. We therefore use all 45 other states as the control group in the baseline regression, and consider several alternatives in Section IVb.

Our standard errors in the baseline regression are heteroskedasticity-robust and clustered by state. As shown by Bertrand et al. (2004), conventional difference-in-differences methods can over-reject the null hypothesis because of serial correlation even when standard errors are clustered. We therefore use more stringent standards for statistical significance than usual: 0.1%, 1%, and 5% significance levels. In Section IVd we will more formally investigate whether underestimated standard errors could be driving our conclusions.

The first column of Table 4 reports the coefficient estimates for $MA * During$ and $MA * After$ from the ordered probit regression, along with the average treatment effects on the treated in the after period.[14] $MA * During$ is insignificant and its coefficient estimate is small, while $MA * After$ is significant at the 0.1% level and its coefficient estimate is considerably larger. In other words, the reform's effect on health did not emerge until it was fully implemented. The t-statistic for $MA * After$ is 5, meaning that our clustered standard errors would have to be underestimated by a factor of more than 2.5 for the statistical significance to be driven by autocorrelation.

The estimated average treatment effects show that the Massachusetts health care reform decreased the probabilities of being in poor, fair and good health and increased the probabilities of being in very good and excellent health. The drops in the probabilities of being in poor, fair,

[14] Coefficient estimates for the other covariates are available upon request.

and good health are 0.2, 0.4, and 0.7 percentage points, respectively, while the increases in the probabilities of being in very good and excellent health are 0.1 and 1.1 percentage points.

We next conduct two back-of-the-envelope calculations to help assess the economic significance of these estimates. The first consolidates the five treatment effects into a single measure that attempts to quantify the overall increase in health. We multiply each of the treatment effects by the value of the health status index associated with the corresponding category (0 for poor, 1 for fair, 2 for good, 3 for very good, and 4 for excellent), and then divide by the sample standard deviation. This result is an overall effect on health of 0.029 standard deviations, shown in the third-to-last row of Table 4.[15] The magnitude of the impact therefore appears modest across the entire population, but perhaps large amongst the small fraction of the population who experienced a change in coverage as a result of the reform and is likely responsible for most of the effect.

The second calculation combines the estimated treatment effects with the information on the reform's costs from the introduction to compute the annual fiscal cost for each adult transitioned from poor, fair, or good health to very good or excellent health. We do this first considering total government spending (federal and state), and then using only Massachusetts' share of that spending. The former provides a more relevant projection for national health care reform, while the latter is more relevant for evaluations of the Massachusetts reform. 1.2% of the adult population transitioned into very good or excellent health. The adult population in Massachusetts was 5,138,919 in July 2010 according to the Census, so 1.2% translates to 61,667 individuals. Since the reform cost an estimated $707 million in FY2010, total government

[15] This calculation should be interpreted with caution, as it relies on the strong assumption that each incremental increase in the health index represents the same improvement in health.

spending is an estimated $11,465 per year for every adult whose health improves from poor, fair, or good to very good or excellent. Since Massachusetts splits the costs evenly with the federal government, the state spends approximately $5,732 annually per adult transitioned into very good or excellent health. These calculations are far from complete cost-effectiveness analyses, as they ignore costs to patients and private insurers as well as benefits from consumption smoothing/risk protection and improvements in children's health. They do, however, provide some information about the returns to government spending while underscoring the point that financing universal coverage at the federal level is likely to be more difficult than in Massachusetts, as matching money is not available.

IVb. Robustness Checks

This section further examines the validity of the identifying assumption of common counterfactual health trends between Massachusetts and the rest of the country by considering a number of alternative control groups and adding state-level covariates. First, we use as the control group the ten states with the most similar pre-treatment average health status indices to Massachusetts ("match on pre-treatment levels"). Second, we "match on pre-treatment trends" by running regressions of health on time (measured in months) plus a constant term for each state from 2001-2005 and then choosing as the comparison group the ten states with the most similar slopes to Massachusetts. Next, we use a control group of the ten states with the most similar pre-reform health insurance coverage rates ("match on pre-treatment coverage").[16] We then consider

[16] When matching on pre-treatment levels, the control states are Colorado, Connecticut, District of Columbia, Maryland, Minnesota, Nebraska, New Hampshire, Utah, South Dakota and Virginia. When matching on pre-treatment trends, the control states are Alabama, Arkansas, Florida, Nebraska, New Mexico, North Carolina, Ohio, Oklahoma, Pennsylvania and Washington. When matching on pre-treatment coverage, the control states are Connecticut, Delaware, District of Columbia, Iowa, Maryland, Michigan, Pennsylvania, Rhode Island, Virginia and Wisconsin.

a control group consisting of other New England states because of their geographic proximity to Massachusetts.[17]

The fifth robustness check constructs a "synthetic control group" for Massachusetts, as described by Abadie et al. (2010). We first aggregate to the state-by-year level and allow the data to select the combination of the other 50 states that best matches Massachusetts on health status and the control variables during the pre-treatment years 2001-2005.[18] The resulting control group is 70.5% Connecticut, 13.7% Washington, D.C., 0.3% New Hampshire, and 15.5% Washington state. Following Fitzpatrick's (2008) application of this method to individual data, we then multiply the weights for the individual-level observations by these shares, leaving Massachusetts fully weighted and dropping the states that received a zero weight.[19]

The next regression uses the original control group of 45 states but excludes the year 2005. Recall from Figure 1 that in 2005 health in Massachusetts was below the trend line, raising the question of whether the improvement in health from 2005 to 2006 could be due to a temporary negative shock in 2005 rather than the reform in 2006. The long pre-treatment period mitigates this concern by tempering 2005's influence, but dropping 2005 addresses it directly.[20]

Finally, we return to the full sample but control for the potential time-varying state-level confounders unemployment rate, cigarette tax rate, physician density, and hospital density, along

[17] Recall that the New England states Vermont and Maine were excluded from the main control group because they had their own (less intensive) health care reforms during the sample period. We include them in the New England control group, but we have verified in an unreported regression that the results are similar if we exclude them and just use Connecticut, Rhode Island, and New Hampshire.

[18] We do this using the Stata module "synth" (Abadie et al., 2011). For health status, we include average health during all five pre-treatment years, as well as average health in the first and last pre-treatment years 2001 and 2005. This accounts for both pre-treatment levels (the overall average) and trends (the first and last pre-treatment years).

[19] In the "matching on pre-treatment levels," "matching on pre-treatment trends," New England, and synthetic control regressions, the number of states is 11 or fewer. Angrist and Pischke (2008) note that standard errors clustered by state are unreliable when the number of states is small. As they recommend, we instead cluster standard errors at the state-by-year level in these four regressions.

[20] Other unreported robustness checks experimented with the use of shorter pre-treatment periods beginning in 2002, 2003, or 2004. The results remained similar.

with linear state-specific time trends to allow for differential trends in health along unobservable dimensions.[21] Controlling for unemployment rate and cigarette tax could be especially important given the differential impacts of the recession across states and the large cigarette tax increase passed in Massachusetts in 2009.

We present the results of these robustness checks in Columns 2 through 8 of Table 4. The coefficient of the interaction term $MA * During$ remains insignificant in almost all specifications, while the interaction term $MA * After$ remains highly significant in all specifications. The magnitude of the coefficient estimate for $MA * After$ ranges from 0.026 to 0.053, so the baseline estimate is toward the conservative end of that range. The treatment effects are also generally similar across specifications. The effect on: P(Poor) ranges from -0.001 to -0.003, P(Fair) ranges from -0.003 to -0.006, P(Good) ranges from -0.005 to -0.010, P(Very Good) ranges from 0.001 to 0.002, and P(Excellent) ranges from 0.008 to 0.017.

IVc. Testing for Differential Pre-Treatment Trends and Delayed Effects

This section simultaneously addresses two possible issues with the estimates from Table 4. First, the difference-in-differences approach assumes common counterfactual health trends between Massachusetts and the rest of the country. The robustness of the estimates to different constructions of the control group is consistent with this assumption, but conceivably health trends in Massachusetts could be so unique that no appropriate comparison group of states exists. Second, the preceding regressions do not differentiate between the short- and long-run health effects of the reform following full implementation. Since health is a capital stock accumulated

[21] We relegate the state-level control variables to a robustness check rather than using them in the main analysis because of concerns that some of them – in particular physician density and hospital density –could be endogenous to health care reform. Moreover, the four state-level controls are all individually and jointly insignificant, so the state fixed effects appear to sufficiently capture their influence on health, at least conditional on the individual controls.

through repeated investments, the improvements in health resulting from the reform could increase over time. Alternatively, the long-term uninsured might experience a pent-up demand for medical services after obtaining coverage, in which case the entire improvement in health could be reached quickly or even be temporary.

We address these issues by re-estimating equation (1) with a broader set of interaction terms. First, we split the "before" and "after" periods into halves, resulting in five periods called "early before," "late before," "during," "early after," and "late after."[22] We then include interactions of the Massachusetts dummy with indicators for each one of these periods, except "early before," which is considered the reference period. A second regression interacts Massachusetts with a full set of year dummies. These models test the common trends assumption by testing for differential trends between Massachusetts and other states in the pre-treatment period. If the treatment and control groups were trending similarly before the reform, then they likely would have continued to trend similarly from 2006-2010 if the reform had not occurred. The models also distinguish between short- and long-run effects by including multiple interactions from the post-reform period.

Table 5 displays the coefficient estimates for the interaction terms. The regression splitting the before and after periods into halves shows that health trends in Massachusetts and other states were similar across the "early before," "late before," and "during" periods, with a sizeable gap emerging in the "early after" period that grew somewhat in the "late after" period. The difference between the coefficient estimates for $MA * Early\ After$ and $MA * Late\ After$ is not significant, though, so the evidence that the effect strengthened over time is not conclusive.

[22] Specifically, "early before" consists of months 1-31 of the pre-treatment period while "late before" consists of months 32-63. "Early after" consists of the first 21 months of the after period and "late after" the last 21.

Turning to the regression with one-year splits, there is again little evidence of differential pre-treatment trends, with the one exception being the temporary negative health shock in Massachusetts in 2005 that we observed earlier in Figure 1. (Recall that the regression excluding 2005 from Table 4 provides evidence that this shock is not meaningfully influencing our conclusions.) A positive interaction effect first emerges in 2007 – the first year in which the reform was fully implemented – and the magnitude of the effect remains roughly similar through 2010. In sum, the results from Table 5 do not provide clear evidence of differential pre-treatment trends between Massachusetts and the control states, or the effects being either delayed or temporary.

IVd. Tests Related to Inference

This section conducts tests to help rule out the possibility that the statistical significance observed in the baseline regression is merely an artifact of underestimated standard errors. First, following Bertrand et al.'s (2004) suggestion, we compress all the available data into a state-level panel with three time periods – "before", "during", and "after" – and regress state average health index on $MA * During$, $MA * After$, and state and time period fixed effects. Next, we compress the data into only two cross-sectional units – Massachusetts and other states – and ten years, defining 2006 and 2007 as the "during" period and 2008 to 2010 as the "after" period. We then regress average health index on $MA * During$, $MA * After$, a Massachusetts dummy, and year fixed effects.[23] As shown in Table 6, $MA * After$ remains statistically significant in both

[23] The small sample sizes preclude the inclusion of the full set of control variables. In lieu of these, we include in both regressions a single control variable that summarizes the influence of all the controls at once. This variable, which we call "predicted health status," is computed by regressing the health index on the controls in individual-level regressions using the pre-treatment data, then using the coefficient estimates to predict the health index for the whole sample, then aggregating to the appropriate level.

regressions despite the small sample, and the effect sizes in standard deviations (of the individual-level health index) are similar to those from Table 4.

In the spirit of Abadie et al. (2010), we also consider a different approach to inference and ask how likely it would be to estimate similarly large health improvements simply by picking any state at random. We re-estimate the baseline ordered probit regression with each of the 45 control states as the "treated" unit, considering the p-value for Massachusetts to be the proportion of these states that exhibited as large a health improvement as that estimated for Massachusetts (i.e. the probability a health improvement as large as the one seen in Massachusetts would be observed in a random draw). Since Abaide et al. (2010) specifically recommended this method for use with the synthetic control approach, we also repeat the analysis but using a synthetic control group for each state rather than the full control group. In the full sample regressions, four states – Florida, New Jersey, New York, and Tennessee – had larger health improvements in the after period than Massachusetts, for a p-value of 0.089. Using synthetic controls, only two states – Florida and Tennessee – had larger gains than Massachusetts, for a p-value of 0.044.[24]

IVe. Other Health Outcomes

This section moves beyond the overall health index and explores the effect of the reform on a variety of additional health outcomes: number of days out of the past 30 not in good physical health, not in good mental health, and with health-related functional limitations; activity-limiting joint pain; BMI; minutes per week of physical activity; and smoking status. These variables were chosen because they satisfy two conditions: 1) they are strongly and

[24] We do not report the full set of results for all 45 states due to space considerations; they are available upon request.

significantly correlated with the overall health index in the expected direction (as shown in Appendix Table A3), and 2) they do not rely on a doctor's diagnosis, since a diagnosis requires medical access which is endogenous to the reform.[25]

Analyzing health outcomes beyond the overall self-assessed health index serves three purposes. First, verifying that we also observe improvements in health using more specific (and therefore less subjective) questions increases our confidence that the reform did in fact improve objective – and not merely subjective – health. Second, examining additional outcomes sheds light on the mechanisms through which the effect on overall health occurred. For instance, obtaining health insurance can improve physical (or mental) health through increased utilization of medical services, mental health through lower stress from reduced financial risk, or health behaviors through expanded access to advice and information. Third, including the health behavior-related variables BMI and smoking tests a separate prediction of economic theory: insurance coverage reduces financial vulnerability to health shocks, which could cause people to take more health risks, a phenomenon known as "ex ante moral hazard" (e.g. Dave and Kaestner, 2009; Bhattacharya et al., 2011).

Days not in good physical and mental health, days with health-related limitations, and minutes of exercise per week are non-negative count variables with variances higher than the means. We therefore estimate negative binomial models for these outcomes. The conditional expectation is given by

$$E[num_{ist}|\mu_{ist},\alpha] = \mu_{ist} \tag{11}$$

[25] The first condition excludes, for instance, alcoholic drinks per month, which is only weakly correlated with health and in the opposite of the expected direction. The second condition excludes BRFSS questions that ask whether a respondent has ever been diagnosed with a particular chronic condition, such as diabetes and asthma.

where num is the number of days or minutes, α is the over-dispersion coefficient, and μ is defined by

$$\mu_{ist} = \exp\left(\gamma_0 + \gamma_1(MA_s * During_t) + \gamma_2(MA_s * After_t) + \boldsymbol{X}'_{\boldsymbol{ist}}\boldsymbol{\gamma_3} + \theta_s + \rho_t\right) \tag{12}$$

The treatment effect on the treated is defined as

$$\tau_{i,MA,t} = \exp\left(\gamma_0 + \gamma_2 + \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\gamma_3} + \theta_{MA} + \rho_t\right) - \exp\left(\gamma_0 + \boldsymbol{X}'_{\boldsymbol{i,MA,t}}\boldsymbol{\gamma_3} + \theta_{MA} + \rho_t\right) \tag{13}$$

while the average treatment effect on the treated is the mean of τ among Massachusetts residents in the "after" period.

For the binary outcome variables (activity-limiting joint pain and smoking status), we estimate probit models of the form

$$\begin{aligned} &\Pr\left(y_{ist} = 1\right) \\ &= \Phi\left(\delta_0 + \delta_1(MA_s * During_t) + \delta_2(MA_s * After_t) + \boldsymbol{X}'_{\boldsymbol{ist}}\boldsymbol{\delta_3} + \omega_s + \upsilon_t\right. \end{aligned} \tag{14}$$

with the treatment effect on the treated being

$$\tau_{i,MA,t} = \Phi(\delta_0 + \delta_2 + \boldsymbol{X}'_{\boldsymbol{ist}}\boldsymbol{\delta_3} + \sigma_{MA} + \upsilon_t) - \Phi(\delta_0 + \boldsymbol{X}'_{\boldsymbol{ist}}\boldsymbol{\delta_3} + \omega_{MA} + \upsilon_t)\,. \tag{15}$$ [26]

Body mass index is continuous, so we estimate a linear regression in which the treatment effect is simply the coefficient estimate for $MA_s * After_t$.

Some of the health-related questions were not asked in Massachusetts in certain years, necessitating restrictions to the sample. Activity-limiting joint pain and the two measures of exercise are only available in odd-numbered survey years, meaning that the "during" period spans only six months (January 2007 to June 2007). We therefore combine those six months with the rest of 2007 and 2009 and classify the two years as the "after" period, dropping the $MA * During$ interaction from those regressions. Additionally, the physical health, mental health, and health limitations variables are not available in 2002.

[26] We include cigarette tax as an additional covariate in the smoking regression.

Table 7 presents the results. We use the full control group because, as shown in Appendix Figures A1-A7, the pre-treatment trends for these other health outcomes are similar for the full control group and Massachusetts.[27] Health care reform in Massachusetts is associated with reductions in the number of days not in good physical health, not in good mental health, and with health-related functional limitations, as well as a lower probability of having activity-limiting joint pain. The magnitudes of these reductions range from 0.013 to 0.041 standard deviations, roughly similar to the size of the effect for the overall health status index. It therefore seems unlikely that the observed effect on the health index is driven purely by the subjectivity of the question. Moreover, these results suggest that the reform improved health more broadly than merely by reducing stress from lower financial risk.

Turning to the health-behavior related variables, the reform is associated with a 0.040 standard deviation reduction in BMI, but no statistically detectable effect on exercise or smoking. The drop in BMI suggests that expanded access to medical care improves at least some weight-related behaviors, perhaps through information or accountability. Moreover, the lack of an effect on exercise suggests that dietary changes are responsible for the weight loss, although the BRFSS does not contain sufficient information during our sample period to test this directly. The non-effect on smoking is consistent with evidence that smoking habits respond only gradually to external factors (e.g. Courtemanche, 2009), but could also reflect the health consequences of smoking already being widely-known even without physician access. Importantly, none of the regressions provide any evidence of ex-ante moral hazard causing individuals to take more health risks after obtaining insurance.

[27] Specifically, the pre-treatment trend lines for Massachusetts and the control states are very similar for physical health, health limitations, joint pain, BMI, exercise, and smoking. The pre-treatment trends are somewhat different for mental health, but by an amount that could easily be attributed to sampling error.

The final column of Table 7 presents the results using as the dependent variable a "cardinalized overall health status index" equal to the predicted outcome from a regression of the health index on the most plausibly objective health outcomes: functional limitations, joint pain, BMI, exercise, and smoking ($R^2 = 0.15$).[28] This approach is advocated by Ziebarth et al. (2010) and others as a way to handle reporting heterogeneity in self-assessed health. The impact of $MA * After$ remains positive and significant, and the effect size in standard deviations is similar to those from Table 4. This provides further evidence that our conclusions are not merely driven by subjectivity.

IVf. Heterogeneity

We next return to the actual overall health status index and examine heterogeneity in the effect of Massachusetts health care reform on the bases of gender, age, race, and income. Kolstad and Kowalski (2010) found the largest coverage expansions among men, young adults, minorities, and those with low incomes. However, different effects on coverage do not necessarily translate to different effects on health, as the impacts of coverage on health could also be heterogeneous. We consider the following subsamples: women; men; ages 18-34, 35-44, 45-54, and 55-64; non-Hispanic whites; those of another race/ethnicity; and household incomes below $25,000, between $25,000 and $75,000, and above $75,000.[29] We choose these income splits in order to loosely align with the provisions of the reform, which specify that health insurance be free up to 150% FPL ($23,050 for a family of four) and subsidized up to 300%

[28] We also considered using all seven alternate health outcomes to make the prediction. The results were similar.
[29] We also ran regressions further dividing the category for those of another race/ethnicity into blacks, Hispanics, and others, but the sample sizes became too small for meaningful differences to emerge.

($69,150 for a family of four).[30] We estimate the baseline ordered probit model with the full control group, as the pre-treatment trends for Massachusetts and the control states (shown in Appendix Figures A8-A19) are similar for these subsamples.[31]

Table 8 reports the results. The impact on health is positive and significant for both women and men but stronger for women. Stratifying by age, the effect is largest among the near-elderly aged 55-64, then about equal among those 18-34 and 45-54, and smallest (but still positive and significant) among those 35-44. The reform improved health across both racial groups, but the effect was largest for non-whites. A back-of-the-envelope calculation suggests that the reform reduced health disparities between non-whites and whites by 5.3%.[32] Stratifying by income, the reform had by far the largest effect among those with incomes below $25,000, the group eligible for the most generous subsidies. We observe moderate health improvements among the middle income group, and smaller and statistically insignificant improvements for those with high incomes.

IVg. Instrumental Variables

We close the empirical analysis by using $MA * During$ and $MA * After$ as instruments to estimate the impact of having insurance coverage on health. This instrumental variables approach requires stricter assumptions than the reduced-form model, as the reform must only

---

[30] 2012 federal poverty lines are available at coverageforall.org/pdf/FHCE_FedPovertyLevel.pdf, accessed 6/26/12. Since the BRFSS only reports income categories and lacks comprehensive information about household size, lining up the categories to exactly match 150% and 300% of the poverty line is not possible.

[31] Specifically, the pre-treatment trend lines are almost identical for the subsamples of women, men, ages 35-44, ages 45-54, ages 55-64, and those with low incomes. The pre-treatment trend lines look slightly different for the race groups and those with middle or high incomes, but the differences are small enough that they could be due merely to sampling error.

[32] In Massachusetts in the "before" period, the mean health status indices of non-whites and whites were 2.559 and 2.898, respectively, for a difference of -0.339. The treatment effects imply changes in the health status indices of non-whites and whites of 0.054 and 0.036, for a difference of 0.018. We therefore estimate that the reform reduced racial health disparities by 5.3%.

impact health along the extensive margin of insurance coverage, conditional on the controls. This assumption would be violated if the reform also influenced health through the intensive margin of coverage, for instance by causing some individuals to switch from high-deductible catastrophic coverage to more comprehensive coverage available through the Connector. This assumption would also be violated if the reform affected the health of those who did not switch insurance plans through system-wide changes to health care delivery or peer effects. Despite these caveats, the instrumental variables analysis is useful because it estimates the magnitude of the impact of insurance on health that would be necessary for the extensive margin to be the only channel through which the reform influenced health. If the magnitude is implausibly large, then other mechanisms must play a role as well.

The first stage predicts insurance coverage using the following linear probability model:

$$ins_{ist} = \alpha_0 + \alpha_1(MA_s * During_t) + \alpha_2(MA_s * After_t) + \boldsymbol{X'_{ist}\alpha_3} + \zeta_s + \eta_t + u_{ist} \quad (15)$$

where ins is a dummy variable equal to 1 if the person reported having any health insurance coverage. Because of the non-linearity of the second stage, we utilize a two-stage residual inclusion (2SRI) approach in which the residual from the first-stage regression is included as an additional regressor in the second stage. Terza et al. (2008) show that in non-linear contexts 2SRI gives consistent coefficient estimates, while traditional two stage least squares does not. The second stage is modeled as an ordered probit and the probabilities of being in each of the five health states are given by,

$$\Pr(y_{ist} = 0) = \Phi(\lambda_1 - \pi_1 ins_{ist} - \boldsymbol{X'_{ist}\pi_2} - \pi_3\hat{u}_{ist} - \sigma_s - \varphi_t) \quad (16)$$

$$\Pr(y_{ist} = k) = \Phi\big(\lambda_j - \pi_1 ins_{ist} - \boldsymbol{X'_{ist}\pi_2} - \pi_3\hat{u}_{ist} - \sigma_s - \varphi_t\big) - \Phi\big(\lambda_{j-1} - \pi_1 ins_{ist} - \boldsymbol{X'_{ist}\pi_2} - \pi_3\hat{u}_{ist} - \sigma_s - \varphi_t), \forall\, j \in (2,3,4) \quad (17)$$

$$\Pr(y_{ist} = 4) = 1 - \Phi(\lambda_4 - \pi_1 ins_{ist} - \boldsymbol{X'_{ist}\pi_2} - \pi_3\hat{u}_{ist} - \sigma_s - \varphi_t) \quad (18)$$

where $\hat{u}_{ist}$ is the first-stage residual. The effect of health insurance on the probability of being in health state j is

$$\Delta p_j = \Pr(y_{ist} = j | ins_{ist} = 1) - \Pr(y_{ist} = j | ins_{ist} = 0) \tag{19}$$

The asymptotic standard errors of these probabilities and the standard errors for the second stage estimates were calculated following Terza (2011).[33] Equation (19) represents the "local average treatment effect" of insurance among those who obtained coverage as a result of the reform, and is subject to the usual caveat regarding generalizability.

Table 9 reports the coefficient estimates of interest from the first and second stage regressions for the full sample, along with the estimated impacts of insurance on the health state probabilities. The first stage estimates an increase in the coverage rate of 2.1 percentage points in the "during" period and 5.6 percentage points in the "after" period. The F statistic from a test of the joint significance of $MA * During$ and $MA * After$ is large, suggesting the instruments are sufficiently strong. Turning to the second stage, obtaining insurance leads to a positive and statistically significant improvement in health. The first-stage residual is significant and negatively associated with health, providing evidence that an OLS estimator would suffer from a downward bias. Insurance is estimated to reduce the probabilities of being in poor, fair, and good health by 5.8, 9.2, and 8.5 percentage points, while increasing the probabilities of being in very good and excellent health by 8.1 and 15.3 percentage points. The overall effect of insurance on the health status index, encompassing changes in all five probabilities, is 0.594 of the sample standard deviation.

These effects are strikingly large, but assessing their plausibility requires a comparison to other estimates from the literature. Finkelstein et al. (2011) employ the cleanest research design

[33] Mata code is available upon request.

to date among studies of the impact of insurance on self-assessed health: a randomized intervention in Oregon granting Medicaid eligibility to a subset of the uninsured. They estimate that Medicaid enrollment increases the probability of being in good, very good, or excellent health by 13.3 percentage points. The sum of our estimated effects on the probabilities of being in those three health states is a similar 14.9 percentage points. The results from the two papers are not directly comparable given the differences in interventions and populations, but this similarity suggests that it is at least conceivable that the reform's entire effect on the self-assessed health of the non-elderly could have occurred through the extensive margin of coverage. Future research should more directly investigate the roles of other potential channels, in particular the intensive margin of coverage.

We also conduct instrumental variables analyses for the gender, race, age, and income subgroups, allowing us to assess whether the heterogeneity in the reform's effect on health observed in Section IVf comes from heterogeneity in the effect *on* coverage or the effect *of* coverage. Table 10 reports the results. The coverage expansions are larger for men than women, but women have greater health gains from coverage, explaining the greater net effect of the reform for women. Among the age subsamples, those under 35 years old have the largest gains in coverage, but also the smallest health improvements from obtaining coverage. 55-64 year olds have the smallest effect of the reform on coverage but the largest effect of coverage on health. Stratifying by race shows that coverage rates increase the most for non-whites but that the health effects of coverage are the largest for whites. Finally, the coverage expansions are by far the largest for the low-income group, second largest for the middle-income group, and relatively small for those with high incomes. However, the effect of coverage on health is the strongest for the high income group.

V. Conclusion

This paper examined the effect of health care reform in Massachusetts on self-assessed health using data from the Behavioral Risk Factor Surveillance System (BRFSS). An ordered probit difference-in-differences analysis showed that the reform increased the probability of individuals reporting excellent or very good health while reducing their probability of reporting good, fair, or poor health. These results were robust to alternative constructions of the control group and the addition of state-level covariates. We did not find evidence that the estimates were meaningfully impacted by differential pre-treatment trends, or that the short- and longer-run effects were statistically different. Next, we examined a number of more specific health outcomes and found improvements in physical health, mental health, functional limitations, joint disorders, and body mass index. Evaluating heterogeneity revealed that those with low incomes, minorities, near-elderly adults, and women experienced the largest gains in health as a result of the reform. Finally, we used the reform as an instrument for health insurance coverage and estimated a large positive impact of coverage on health.

Perhaps the clearest limitation of our analysis is that all our health outcomes were self-reported. Our finding of similar results across a range of health outcomes with varying degrees of subjectivity increases our confidence that our findings largely represent "real" changes in physical/mental health. However, we cannot rule out the possibility that some of the observed improvement in health could merely be due to a "warm glow" from acquiring health insurance. To underscore this point, recall that our estimated effects of insurance on self-assessed health are a similar magnitude to those of Finkelstein et al. (2011), and they found that a sizeable portion of the reported health improvements following the Oregon experiment occurred prior to measurable changes in overall health care utilization. Obtaining insurance coverage can reduce stress, which can directly improve numerous aspects of health even without any additional medical care being utilized, but Finkelstein et al. (2011) do raise the question of what exactly the estimated improvements in self-assessed health are capturing.

We argue that Finkelstein et al.'s (2011) finding regarding timing does not automatically apply to our context for several reasons. First, their data only tracked individuals for a year after the intervention, while we have 4½ years of data after first of the newly-insured in Massachusetts obtained coverage and 3½ years after all major facets of the reform took effect. If a "warm glow" from acquiring insurance was driving the effect, we would have expected the reported health benefits from the reform to diminish over time, but as Table 5 showed this was not the case. Second, other studies have documented changes in health care utilization in Massachusetts either before or around the same time as we observed health improvements (Kowalski and Kolstand, 2010; Miller, 2011a). Next, the newly insured in the Oregon experiment were winners of a random lottery, which could lead to a stronger "warm glow" than simply acquiring health insurance from a statewide intervention like the reform in Massachusetts. Accordingly, we consistently find that the effects on health were statistically zero in the "during" period, which includes nine months after those with incomes below 100% FPL became eligible for free coverage. We therefore do not observe the immediate spike in self-assessed health seen in the Oregon experiment. Nonetheless, as the necessary data become available it will be important to evaluate the impact of the Massachusetts reform on unambiguously objective measures of health such as mortality.

Another natural question is the degree to which our results from Massachusetts can serve as projections for the Affordable Care Act. The general strategies for obtaining nearly universal coverage in both the Massachusetts and federal laws involved the same three-pronged approach of non-group insurance market reforms, mandates, and subsidies, suggesting that the health effects should be broadly similar. However, the federal legislation included additional cost-cutting measures such as Medicare cuts that could potentially mitigate the gains in health from the coverage expansions. On the other hand, baseline uninsured rates were unusually low in Massachusetts, so the coverage expansions – and corresponding health improvements – from the Affordable Care Act could potentially be greater. Of course, larger coverage expansions may mean higher costs, and costs should be weighed against benefits when evaluating the welfare implications of reform.

References

Abadie A., Diamond A., and Hainmueller J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), 493-505.

Abadie A., Diamond A., and Hainmueller J. (2011). SYNTH: Stata Module to Implement Synthetic Control Methods for Comparative Case Studies. Software item, available http://econpapers.repec.org/software/bocbocode/s457334.htm.

Ai C. and Norton E. (2003). Interaction Terms in Logit and Probit Models. *Economics Letters,* 80(1), 123-129.

Angrist J. and Pischke J. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Arrow K. (1963). Uncertainty and the Demand for Medical Care. *American Economic Review*, 53(5), 941-73.

Bertrand M., Duflo E., and Mullainathan S. (2004). How Much Should We Trust Difference-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1), 249-275.

Bhattacharya J., Bundorf M.K., Pace N., and Sood N. (2011). Does Health Insurance Make You Fat? In Grossman M. and Mocan N. (eds.), *Economic Aspects of Obesity*, Chicago, IL: University of Chicago Press, 35-64.

Blue Cross Blue Shield of Massachusetts (2006). Massachusetts Health Care Reform Bill Summary. Available at http://tinyurl.com/bn5yepz.

Card D., Dobkin C., and Maestas N. (2009). Does Medicare Saves Lives? *Quarterly Journal of Economics*, 124(2), 597-636,

Card D., Dobkin C., and Maestas N. (2004). The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare. National Bureau of Economic Research Working Paper #10365.

Cawley J. (2004). The Impact of Obesity on Wages. *Journal of Human Resources*, 39(2): 451-474.

Cogan J., Hubbard R., and Kessler D. (2010). The Effect of Massachusetts' Health Reform on Employer-Sponsored Insurance Premiums. *Forum for Health Economics and Policy*, 13(2), Article 5.

Courtemanche C. (2009). Rising Cigarette Prices and Rising Obesity: Coincidence or Unintended Consequence? *Journal of Health Economics*, 28(4), 781-798.

Courtemanche C., Heutel G., and McAlvanah P. (2011). Impatience, Incentives, and Obesity. National Bureau of Economic Research Working Paper #17483.

Currie, J. and Gruber, J. (1996a). Health Insurance Eligibility, Utilization of Medical Care, and Child Health. *Quarterly Journal of Economics*, 111(2), 431-466.

Currie, J. and Gruber, J. (1996b). Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. *Journal of Political Economy*, 104(6), 1263–1296.

Dafny L. and Gruber J. (2005). Public Insurance and Child Hospitalizations: Access and Efficiency Effects. *Journal of Public Economics*, 89, 109-129.

Dave D. and Kaestner R. (2009). Health Insurance and Ex Ante Moral Hazard: Evidence from Medicare. International Journal of Health Care Finance and Economics, 9, 367-390.

DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., and Muntner, P. (2006). Mortality prediction with a single general self-rated health question. A meta-analysis. *Journal of General Internal Medicine*, 21(3), 267–275.

Doyle J. (2005). Health Insurance, Treatment and Outcomes: Using Auto Accidents as Health Shocks. *Review of Economics and Statistics*, 87(2), 256-270.

Finkelstein A. and McKnight R. (2008). What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending. *Journal of Public Economics*, 92, 1644-1669.

Finkelstein A., Taubman S., Wright B., Bernstein M., Gruber J., Newhouse J., Allen H., Baicker K., and The Oregon Health Study Group (2011). The Oregon Health Insurance Experiment: Evidence from the First Year. *Quarterly Journal of Economics*, forthcoming.

Fitzpatrick M. (2008). Starting School at Four: The Effect of Universal Pre-Kindergarten on Children's Academic Achievement. *B.E. Journal of Economic Analysis and Policy*, 8(1), Article 46.

Grossman M. (1972). *The Demand for Health: A Theoretical and Empirical Investigation*, Columbia University Press: New York, NY.

Gruber J. (2008a). Incremental Universalism for the United States: The States Move First? *Journal of Economic Perspectives*, 22(4), 51-68.

Gruber J. (2008b). Massachusetts Health Care Reform: The View From One Year Out. *Risk Management and Insurance Review*, 11(1), 51-63.

Gruber J. (2010). Health Care Reform is a "Three-Legged Stool": The Costs of Partially Repealing the Affordable Care Act. Center for American Progress. Available http://www.americanprogress.org/issues/2010/08/pdf/repealing_reform.pdf.

Gruber J. (2011a). The impact of the Affordable care Act: How reasonable are the projections? National Bureau of Economic Research Working Paper #17168.

Gruber J. (2011b). Massachusetts Points the Way to Successful Health Care Reform. *Journal of Policy Analysis and Management*, 30(1), 184-192.

Hadley J. (2003). Sicker and Poorer—The Consequences of Being Uninsured: A Review of the Research on the Relationship between Health Insurance, Medical Care Use, Health, Work, and Income. *Medical Care Research and Review*, 60(2), 3-75.

Harrington S. (2010). U.S. Health Care Reform: The Patient Protection and Affordable Act. *Journal of Risk and Insurance*, 77(3), 703-708.

Idler E. and Benyamini Y. (1997). Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *Journal of Health and Social Behavior*, 38 (1), 21-37.

Johnson-Lans S. (2005). *A Health Economics Primer*. Boston, MA: Addison Wesley/Pearson.

Kolstad J. and Kowalski A. (2010). The Impact of Health Care Reform on Hospital and Preventive Care: Evidence from Massachusetts. National Bureau of Economic Research Working Paper #16012.

Kolstad J. and Kowalski A. (2012). Mandate-Based Health Reform and the Labor Market: Evidence from the Massachusetts Reform. National Bureau of Economic Research Working Paper #17933.

Kirk A. (2000). Riding the Bull: Experience with Individual Market Reform in Washington, Kentucky, and Massachusetts. *Journal of Health Politics, Policy and Law*, 25(1), 133-173.

Levy H. and Meltzer D. (2008). The Impact of Health Insurance on Health. *Annual Review of Public Health*, 29, 399–409.

Long, S. (2008). On the Road to Universal Coverage: Impacts of Reform in Massachusetts at One Year. *Health Affairs*, 27(4), w270-84.

Long, S. and Stockely K. (2011). The Impacts of State Health Reform Initiatives on Adults in New York and Massachusetts. *Health Services Research*, 46(1,II), 365-428.

Long S., Stockely K. and Yemane A. (2009). Another Look at the Impacts of Health Reform in Massachusetts: Evidence Using New Data and a Stronger Model. *American Economic Review Papers and Proceedings*, 99(2), 508-511.

Manning W., Newhouse J., Duan N., Keeler E., Leibowitz A., and Marquis M. (1987). Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *American Economic Review*, 77(3), 251-77.

Massachusetts Division of Health Care, Finance, and Policy (2010). Health Care in Massachusetts: Key Indicators. Available http://archives.lib.state.ma.us/handle/2452/69933.

Massachusetts Governor's Budget FY2010 (2009). Health Care. Retrieved from http://www.mass.gov/bb/h1/fy10h1/prnt10/exec10/pbuddevhc.htm.

Massachusetts Health Insurance Connector Authority (2008). Report to the Massachusetts Legislature Implementation of the Health Care Reform Law, Chapter 58 2006-2008. Available http://tinyurl.com/74osvco.

McDonough J., Rosman B., Phelps F., and Shannon M. (2006) The Third Wave of Massachusetts Health Care Access Reform. *Health Affairs*, 25(6), w420-31.

Miller S. (2011a). The Effect of Insurance on Outpatient Emergency Room Visits: An Analysis of the 2006 Massachusetts Health Reform. Working paper, University of Illinois.

Miller S. (2011b). The Impact of the Massachusetts Health Care Reform on Health Care Use Among Children. Forthcoming, *American Economic Review Papers and Proceedings*.

Orzechowski W., Walker R. (2010). *The Tax Burden on Tobacco: Historical Compilation, Volume 45*. Arlington, VA: Orzechowski and Walker.

Patel K., McDonough J. (2010). From Massachusetts to 1600 Pennsylvania Avenue: Aboard the Health Reform Express. *Health Affairs*, 29(6), 1106-1110.

Pauly M. (1968). The Economics of Moral Hazard: Comment. *American Economic Review,* 58(3), 531-36.

Pauly M. (2005). Effects of Insurance Coverage on Use of Care and Health Outcomes for Non-poor Young Women. *American Economic Review, Papers and Proceedings*, 95(2), 219-223.

Phillips A., Der G., and Carroll D. (2010). Self-reported health, self-reported fitness, and all-cause mortality: Prospective cohort study. *British Journal of Health Psychology*, *15*(2), 337-346.

Puhani P. (2008). The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear "Difference-in-Differences" Models. IZA Discussion Paper Series, 3478.

Raymond A. (2009). Massachusetts Health Reform: The Myth of Uncontrolled Costs. Massachusetts Taxpayer Foundation. Retrieved from http://www.masstaxpayers.org/files/Health%20care-NT.pdf

Strauss J., Thomas D. (2007). Health over the life course. *Handbook of Development Economics*, Volume 4, Elsevier.

Terza J., Basu A., and Rathouz P. (2008). Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling. *Journal of Health Economics*, 27, 531–543.

Terza, J.V. (2011). Correct Standard Errors for Multi-Stage Causal Effect Estimation in the Context of Empirical Health Policy Analysis: A Practitioner's Guide. Unpublished Manuscript, University of North Carolina at Greensboro.

Terza, J.V. (2012). Correct Standard Errors for Multi-Stage Regression-Based Estimators: A Practitioner's Guide with Illustrations. Unpublished Manuscript, University of North Carolina at Greensboro.

Weissman J. and Bigby J. (2009). Massachusetts Health Care Reform — Near-Universal Coverage at What Cost? *New England Journal of Medicine*, 361(21), 2012-2015.

Yelowitz A. and Cannon M. (2010). The Massachusetts Health Plan Much Pain, Little Gain. *Policy Analysis*, 657.

Ziebarth N. (2010). Measurement of Health, Health Inequality, and Reporting Heterogeneity. *Social Science and Medicine*, 71, 116-124.

Figure 1 – Changes in Health Status Index 2001-2010

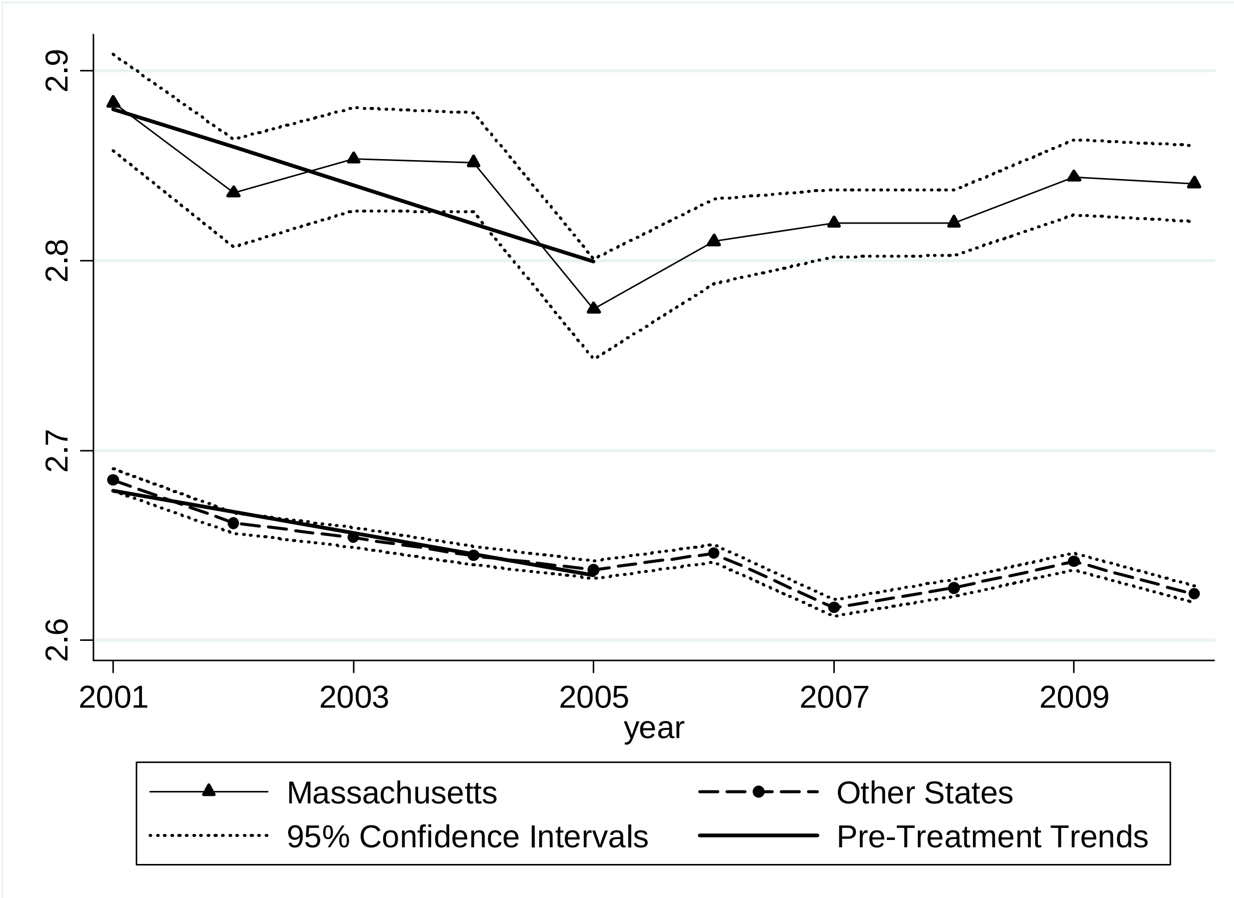

Table 1 – Similarities and Differences between the Massachusetts Reform and the National Reform (ACA)

Domain	Massachusetts reform	National reform (ACA)
Modification of existing insurance markets		
	- No pre-existing condition exclusions (since 1996). - Community rated premiums that can only vary by age and smoking status (in place since 1996). - Minimum standards for policies, including essential benefits and maximum out of pocket expenditures. - Creation of a state health insurance exchange where insurance companies compete to offer three regulated levels of coverage to small employers and individuals. - Young adults must be allowed coverage on their parents' plans for up to two years after they are no longer dependents or until their 26th birthday.	- No pre-existing condition exclusions. - Community rated premiums that can only vary by age and smoking status. - Minimum standards for policies, including essential benefits and maximum out of pocket expenditures. - States must create a health insurance exchange where insurance companies compete to offer four regulated levels of coverage to small employers and individuals. States are able to join multistate exchanges. - Young adults must be allowed coverage on their parents' plans until their 26th birthday.
Mandates	- Individuals are required to purchase coverage if affordable, (based on income and family size) or pay a penalty of no more than 50% of the insurance premium of the lowest-cost insurance exchange plan for which they are eligible. - Employers with more than 10 full time employees (FTE) are required to offer policies with minimum standard or pay a penalty of up to $295 annually per FTE.	- Individuals are required to purchase coverage if it costs no more than 8% of income, or pay a penalty of the greater of 2.5 percent of taxable income or $695. - Employers with 50 employees or more are required to offer policies with minimum standard or pay penalties that range from $2,000-$3,000 per FTE.
Medicaid expansions and subsidies		
	- Medicaid expansions for children with household incomes up to 300% of the poverty line (FPL), for long-term unemployed up to 100% FPL, and for people with HIV up to 200% FPL. - Free coverage for all adults below 150% FPL. Sliding scale of subsidies for adults up to 300% FPL.	- Medicaid expansions to all individuals with incomes below 133% FPL. - Sliding scale of tax credits for people up to 400% FPL. - Tax credits for employers with 25 or fewer employees and average annual wages less than $50,000 for offering coverage.
Financing	- Redirection of federal funding to safety net providers. - Redirection of the state uncompensated care pool, a mechanism through which hospitals were able to bill the state the cost of treating low-income patients. - Individual and employer penalties. - One-time assessment to health care providers and insurers. - Since 2009, a $1 per pack cigarette tax.	- Reduction of Medicare reimbursements. - Increase in the Medicare payroll tax and extension of this tax to capital income for singles (families) with incomes more than $200,000 ($250,000). - Individual and employer penalties. - Taxes on insurers, pharmaceutical companies, and medical device manufactures. - Excise taxes on high-cost insurance plans ("Cadillac tax").

Sources: Gruber (2011a, 2008b) and Harrington (2010).

Table 2 – Pre-Treatment Means of Health Variables

Variable	MA (n=29,868)	Other States (n=894,027)	Difference
Any health insurance coverage	0.898	0.826	0.072***
Overall health; 0 (poor) to 4 (excellent)	2.840	2.656	0.184***
Poor health	0.022	0.032	-0.011***
Fair health	0.072	0.095	-0.023***
Good health	0.244	0.288	-0.044***
Very good health	0.367	0.352	0.015***
Excellent health	0.294	0.232	0.062***
Days not in good physical health (of last 30)[++]	2.884	3.087	-0.203***
Days not in good mental health (of last 30)[++]	3.552	3.639	-0.087
Days with health limitations (of last 30)[++]	1.790	1.946	-0.156***
Activity-limiting joint problems[+]	0.106	0.116	-0.010**
Body mass index	26.305	27.992	-0.767***
Minutes of exercise per day[+]	97.268	93.201	4.068***
Currently smokes cigarettes	0.212	0.256	-0.045***

Notes: *** indicates difference between Massachusetts and other states is significant at the 0.1% level; ** 1% level; * 5% level. Observations are weighted using the BRFSS sampling weights. + indicates variable from only odd-numbered survey years. ++ indicates variable from all years except 2002. Standard errors are available on request.

Table 3 – Pre-Treatment Means of Control Variables

Variable	Massachusetts (n=29,868)	Other States (n=894,027)	Difference
Age 18 to 24	0.135	0.141	-0.006
Age 25 to 29	0.099	0.103	-0.004
Age 30 to 34	0.127	0.123	0.002
Age 35 to 39	0.125	0.121	0.004
Age 40 to 44	0.137	0.132	0.005*
Age 45 to 49	0.118	0.118	0.000
Age 50 to 54	0.105	0.108	-0.003
Age 55 to 59	0.087	0.086	0.001
Age 60 to 64	0.066	0.067	0.000
Female	0.502	0.496	0.006
Married	0.575	0.609	-0.034***
Race is non-Hispanic white	0.828	0.722	0.107***
Race is non-Hispanic black	0.037	0.112	-0.075***
Race is Hispanic	0.128	0.152	-0.024***
Race is neither black nor white nor Hispanic	0.007	0.014	-0.007***
Income less than $10,000	0.033	0.046	-0.013***
Income $10,000 to $15,000	0.029	0.044	-0.014***
Income $15,000 to $20,000	0.046	0.070	-0.024***
Income $20,000 to $25,000	0.062	0.089	-0.027***
Income $25,000 to $35,000	0.097	0.134	-0.027***
Income $35,000 to $50,000	0.147	0.180	-0.033***
Income $50,000 to $75,000	0.200	0.190	0.010***
Income $75,000 or more	0.386	0.248	0.138***
Less than a high school degree	0.063	0.094	-0.031***
High school degree but no college	0.234	0.299	-0.065***
Some college but not four-year degree	0.247	0.280	-0.033***
College graduate	0.456	0.326	0.130***
Currently pregnant	0.013	0.014	-0.001
State unemployment rate	4.984	5.334	-0.350***
State cigarette tax (2010 $)	1.485	0.774	0.711***
State physician density (per 10,000 residents)	436.259	256.672	179.587***
State hospital density (per 10,000 residents)	1.208	1.791	-0.583***

Notes: *** indicates difference between Massachusetts and other states is significant at the 0.1% level; ** 1% level; * 5% level. Observations are weighted using the BRFSS sampling weights. Standard errors are available on request.

Table 4 – Difference-in-Differences Ordered Probit Regressions

Dependent Variable: Overall Health								
	Full Sample	Match on Pre-Tx. Level	Match on Pre-Tx. Trend	Match on Pre-Tx. Coverage	New England	Synthetic Control Group	Drop 2005	Add State Controls/ Trends
Coefficient Estimates of Interest								
MA*During	0.005 (0.008)	0.001 (0.021)	0.014 (0.023)	0.007 (0.021)	0.006 (0.011)	0.008 (0.008)	-0.004 (0.003)	0.022 (0.010)*
MA*After	0.035 (0.007)***	0.051 (0.011)***	0.045 (0.015)***	0.046 (0.011)***	0.052 (0.008)***	0.053 (0.009)***	0.026 (0.008)***	0.049 (0.010)***
Average Treatment Effects on Treated (After Period)								
P(Poor)	-0.002 (0.0003)***	-0.002 (0.0005)***	-0.002 (0.0007)***	-0.002 (0.0005)***	-0.002 (0.0004)***	-0.002 (0.0004)***	-0.001 (0.0003)***	-0.003 (0.0006)***
P(Fair)	-0.004 (0.0007)***	-0.005 (0.001)***	-0.005 (0.002)***	-0.005 (0.001)***	-0.006 (0.001)***	-0.006 (0.001)***	-0.003 (0.0007)***	-0.006 (0.001)***
P(Good)	-0.007 (0.001)***	-0.010 (0.002)***	-0.008 (0.003)***	-0.009 (0.002)***	-0.010 (0.002)***	-0.010 (0.002)***	-0.005 (0.001)***	-0.008 (0.002)***
P(Very Good)	0.001 (0.0002)***	0.001 (0.0004)***	0.001 (0.0004)**	0.001 (0.0004)***	0.001 (0.0003)***	0.001 (0.0003)***	0.001 (0.0002)***	0.002 (0.0006)***
P(Excellent)	0.011 (0.002)***	0.016 (0.003)***	0.014 (0.005)***	0.014 (0.003)***	0.016 (0.003)***	0.017 (0.003)***	0.008 (0.002)***	0.015 (0.003)***
Overall Effect in Std. Dev.	0.029	0.042	0.038	0.036	0.041	0.045	0.022	0.041
# Control States	45	10	10	10	5	4	45	50
Observations	1,979,383	486,449	675,076	453,695	260,893	299,668	1,771,837	2,879,296

Notes: Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. In columns 2-6, standard errors are clustered at the state*year level rather than state because of the small number of states. *** indicates statistically significant at the 0.1% level; ** 1% level; * 5 % level. All regressions include the individual-level control variables, state fixed effects, and fixed effects for each month in each year. Observations are weighted using the BRFSS sampling weights.

Table 5 – Testing for Differential Pre-Treatment Trends and Delayed Effects

Dependent Variable: Overall Health		
	Split Before and After Periods	1-Year Splits
MA*Late Before	-0.008 (0.005)	--
MA*During	0.002 (0.008)	--
MA*Early After	0.027 (0.007)***	--
MA*Late After	0.037 (0.008)***	--
MA*2002	--	-0.008 (0.008)
MA*2003	--	-0.001 (0.010)
MA*2004	--	0.007 (0.008)
MA*2005	--	-0.047 (0.009)***
MA*2006	--	-0.021 (0.011)
MA*2007	--	0.030 (0.009)***
MA*2008	--	0.020 (0.010)*
MA*2009	--	0.035 (0.013)***
MA*2010	--	0.029 (0.010)**
Observations	1,979,383	1,979,383

Notes: Coefficient estimates are shown; treatment effects are available upon request. Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at the 0.1% level; ** 1% level; * 5 % level. All regressions include the individual-level control variables, state fixed effects, and fixed effects for each month in each year. The full control group is used. Observations are weighted using the BRFSS sampling weights.

Table 6 – Regressions with Aggregated Data

Dependent Variable: Average Health Status		
	State-Level with Three Time Periods	Annual with Two Cross-Sectional Units (MA and not MA)
MA*During	0.009 (0.006)	0.027 (0.026)
MA*After	0.039 (0.006)***	0.053 (0.018)**
Effect in Standard Deviations (After Period)	0.039	0.053
Observations	138	20

Notes: The coefficient estimates are equal to the treatment effects because the model is linear. Heteroskedasticity-robust standard errors (clustered by state in the first column) are in parentheses. *** indicates statistically significant at the 0.1% level; ** 1% level; * 5 % level. The first regression includes state fixed effects and dummies for the during and after periods; the second regression includes year fixed effects and a dummy for MA. Both regressions also control for "predicted health," which is based on regressing health on the control variables in an individual-level regression. The control group consists of all 45 states from the full control group in the first regression, and one cross-sectional unit collapsing all individuals from these 45 states in the second regression. Observations are weighted using the BRFSS sampling weights when aggregating.

Table 7 – Regression Results for Other Health Outcomes

Dependent Variable:	Days not in Good Physical Health	Days not in Good Mental Health	Days with Health Limitations	Activity-Limiting Joint Pain	BMI	Minutes of Vigorous Exercise	Smoker	Cardinal-ized Overall Health
MA*During	-0.065 (0.014)***	-0.018 (0.015)	-0.029 (0.018)	--	-0.084 (0.021)***	--	-0.004 (0.008)	--
MA*After	-0.098 (0.011)***	-0.059 (0.012)***	-0.076 (0.014)***	-0.060 (0.010)***	-0.220 (0.023)***	-0.013 (0.013)	0.001 (0.009)	0.024 (0.011)*
ATE on Treated	-0.281 (0.033)***	-0.208 (0.042)***	-0.142 (0.027)***	-0.009 (0.002)***	-0.220 (0.023)***+	-1.152 (1.169)	0.000 (0.002)	0.024 (0.011)*+
Effect in Std. Deviations	-0.041	-0.027	-0.013	-0.019	-0.040	-0.020	0.000	0.047
Observations	1,813,024	1,811,980	1,821,481	920,878	1,912,757	863,751	1,979,503	790,899

Notes: + indicates the treatment effect and coefficient estimate are equal because the model is linear. Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at the 0.1% level; ** 1% level; * 5 % level. All regressions include the individual-level control variables, state fixed effects, and fixed effects for each month in each year. MA*During is not included for joint pain, exercise, and cadrinalized health, since these variables are only available in odd-numbered survey years. The full control group is used in all regressions. Observations are weighted using the BRFSS sampling weights.

Table 8 – Heterogeneity in the Effect on Health by Gender, Age, Race, and Income

Dependent Variable: Overall Health

	Gender		Age				Race		Income		
	Women	Men	18-34	35-44	45-54	55-64	White Non-Hispanic	Other	<$25K	$25K-$75K	>$75K
MA*After	0.047 (0.007)***	0.023 (0.006)***	0.030 (0.010)**	0.018 (0.008)**	0.031 (0.007)***	0.050 (0.007)***	0.036 (0.006)***	0.054 (0.014)***	0.069 (0.011)***	0.024 (0.007)***	0.015 (0.008)
Average Treatment Effects on Treated											
P(Poor)	-0.002 (0.000)***	-0.001 (0.000)***	-0.0007 (0.000)**	-0.0006 (0.000)**	-0.002 (0.000)***	-0.004 (0.001)***	-0.001 (0.000)***	-0.003 (0.001)***	-0.009 (0.002)***	-0.001 (0.000)***	-0.0002 (0.000)*
P(Fair)	-0.005 (0.001)***	-0.003 (0.001)***	-0.003 (0.011)**	-0.002 (0.001)**	-0.003 (0.001)***	-0.006 (0.001)***	-0.003 (0.001)***	-0.008 (0.002)***	-0.011 (0.002)***	-0.003 (0.001)***	-0.001 (0.0005)
P(Good)	-0.009 (0.001)***	-0.005 (0.002)**	-0.006 (0.002)**	-0.004 (0.002)**	-0.006 (0.001)***	-0.008 (0.001)***	-0.007 (0.001)***	-0.008 (0.002)***	-0.003 (0.000)***	-0.005 (0.001)***	-0.003 (0.002)
P(V. Good)	0.001 (0.000)***	0.001 (0.000)***	0.001 (-0.004)	0.00003 (-0.000)	0.0008 (0.000)***	0.004 (0.001)***	0.002 (0.001)***	0.004 (0.001)***	0.009 (0.002)***	0.002 (0.001)***	-0.001 (0.0005)*
P(Excel.)	0.015 (0.002)***	0.007 (0.002)**	0.009 (0.003)*	0.006 (0.003)**	0.010 (0.002)***	0.014 (0.002)***	0.012 (0.002)***	0.015 (0.004)***	0.014 (0.002)***	0.007 (0.002)***	0.006 (0.003)
Overall Eff. in Std. Dev.	0.039	0.024	0.026	0.014	0.026	0.042	0.032	0.044	0.058	0.022	0.014
Observations	1,186,329	793,054	443,359	467,050	558,475	563,405	1,564,481	414,902	478,276	945,561	555,546

Notes: Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at the 0.1% level; ** 1% level; * 5 % level. All regressions include MA*During, the individual-level control variables, state fixed effects, and fixed effects for each month in each year. The full control group is used in all regressions. Observations are weighted using the BRFSS sampling weights.

Table 9 – Instrumental Variables

First Stage: Any Insurance Coverage	
Coefficient Estimates	
MA*During	0.021 (0.002)***
MA*After	0.056 (0.003)***
1st Stage F Statistic	179.41
Second Stage: Overall Health	
Coefficient Estimates	
Insurance	0.655 (0.154)***
1st Stage Residual	-0.634 (0.124)***
Local Average Treatment Effects	
P(Poor)	-0.058 (0.012)***
P(Fair)	-0.092 (0.016)***
P(Good)	-0.085 (0.015)***
P(Very Good)	0.081 (0.014)***
P(Excellent)	0.153 (0.029)***
Overall Effect in Standard Deviations	0.594
Observations	1,976,564

Notes: A linear probability model is estimated in the first stage so the coefficient estimate equals the treatment effect. Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at the 0.1% level; ** 1% level; * 5 % level. All regressions include the individual-level control variables, state fixed effects, and fixed effects for each month in each year. The full control group is used. Observations are weighted using the BRFSS sampling weights.

Table 10 – Instrumental Variables: Stratified by Gender, Age, Race, and Income

	Gender		Age				Race		Income		
	Women	Men	18-34	35-44	45-54	55-64	White	Other	<$25K	$25K-$75K	>$75K
First Stage Coefficient Estimates: Any Insurance Coverage											
MA*During	0.023 (0.002)***	0.019 (0.003)***	0.030 (0.005)***	0.020 (0.003)***	0.012 (0.002)***	0.020 (0.002)***	0.016 (0.001)***	0.035 (0.006)**	0.073 (0.006)***	0.016 (0.004)***	0.007 (0.002)**
MA*After	0.044 (0.002)***	0.068 (0.004)***	0.085 (0.005)***	0.050 (0.003)***	0.045 (0.002)***	0.041 (0.002)***	0.042 (0.002)***	0.092 (0.006)***	0.145 (0.008)***	0.070 (0.003)***	0.015 (0.002)***
F Statistic	169.15	163.52	137.16	123.80	520.33	237.28	228.37	126.35	169.24	321.62	26.61
Second Stage: Overall Health											
Coefficient Estimates											
Insurance	1.059 (0.171)***	0.392 (0.119)***	0.370 (0.122)**	0.379 (0.172)*	0.722 (0.152)***	1.231 (0.174)***	0.861 (0.140)***	0.619 (0.148)***	0.073 (0.006)***	0.016 (0.004)***	0.007 (0.002)**
First Stage Residual	-1.024 (0.170)***	-0.375 (0.123)**	-0.227 (-0.125)	-0.378 (0.175)*	-0.820 (0.149)***	-1.361 (0.176)***	-0.859 (0.141)***	-0.555 (0.148)***	-0.496 (0.084)***	-0.258 (0.019)**	-1.010 (0.512)**
Local Average Treatment Effects											
P(Poor)	-0.125 (0.020)***	-0.028 (0.010)**	-0.014 (0.004)***	-0.023 (0.009)*	-0.085 (0.017)***	-0.247 (0.026)***	-0.087 (0.015)***	-0.058 (0.018)***	-0.071 (0.012)***	-0.019 (0.005)***	-0.082 (0.021)***
P(Fair)	-0.148 (0.019)***	-0.052 (0.016)**	-0.054 (0.015)***	-0.049 (0.021)*	-0.101 (0.019)***	-0.139 (0.018)***	-0.110 (0.015)***	-0.105 (0.023)***	-0.070 (0.011)***	-0.045 (0.013)***	-0.145 (0.042)***
P(Good)	-0.088 (0.017)***	-0.062 (0.017)***	-0.087 (0.026)***	-0.065 (0.031)*	-0.068 (0.018)***	-0.015 (0.018)	-0.108 (0.016)***	-0.054 (0.010)*	-0.008 (0.005)	-0.059 (0.019)**	-0.190 (0.010)***
P(V. Good)	0.145 (0.014)***	0.042 (0.013)**	0.035 (0.009)***	0.037 (0.014)**	0.105 (0.017)***	0.190 (0.014)***	0.112 (0.013)***	0.078 (0.018)***	0.061 (0.009)***	0.040 (0.010)***	0.144 (0.014)***
P(Excel.)	0.215 (0.040)***	0.100 (0.035)**	0.120 (0.037)**	0.100 (0.048)*	0.149 (0.036)***	0.181 (0.041)***	0.193 (0.033)***	0.138 (0.030)***	0.087 (0.016)***	0.083 (0.027)**	0.273 (0.166)
Overall Eff. in Std. Dev.	0.961	0.352	0.389	0.337	0.647	1.074	0.804	0.533	0.393	0.302	1.159
Observations	1,185,172	791,392	441,873	466,669	557,974	510,048	1,562,540	414,024	476,970	944,457	555,137

See notes for Table 9.

Appendix

Figure A1 – Changes in Days not in Good Physical Health 2001-2010

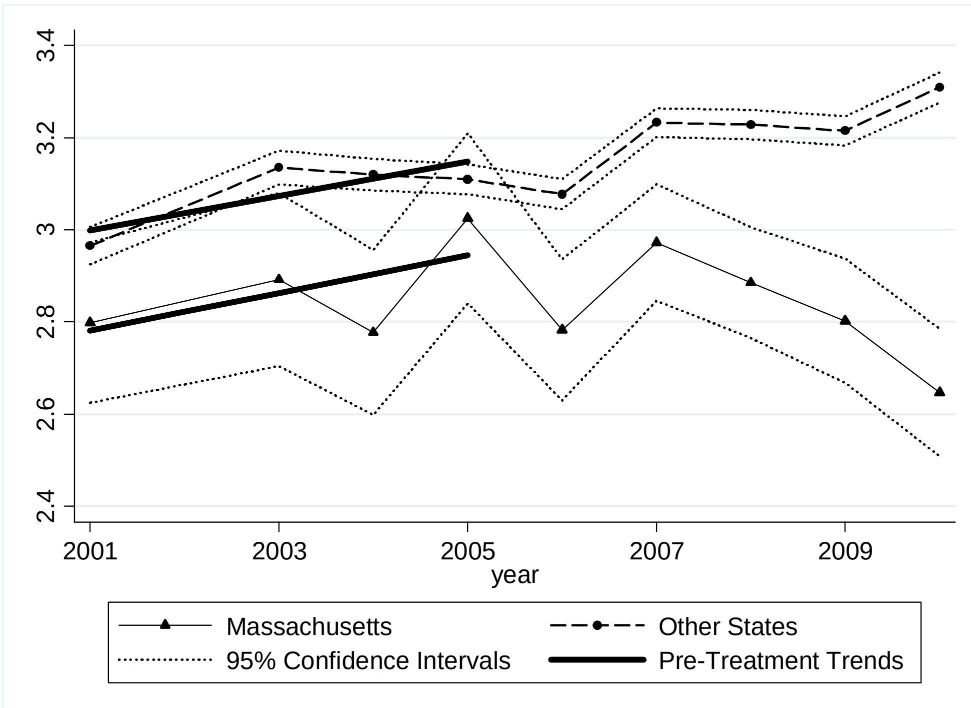

Figure A2 – Changes in Days not in Good Mental Health 2001-2010

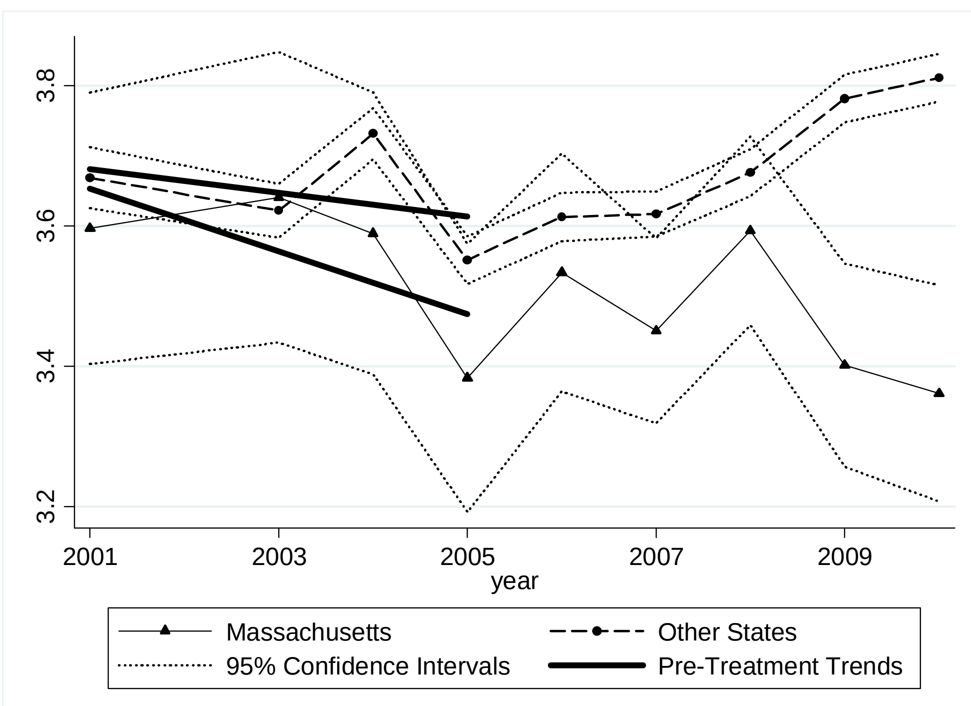

Figure A3 – Changes in Days with Health Limitations 2001-2010

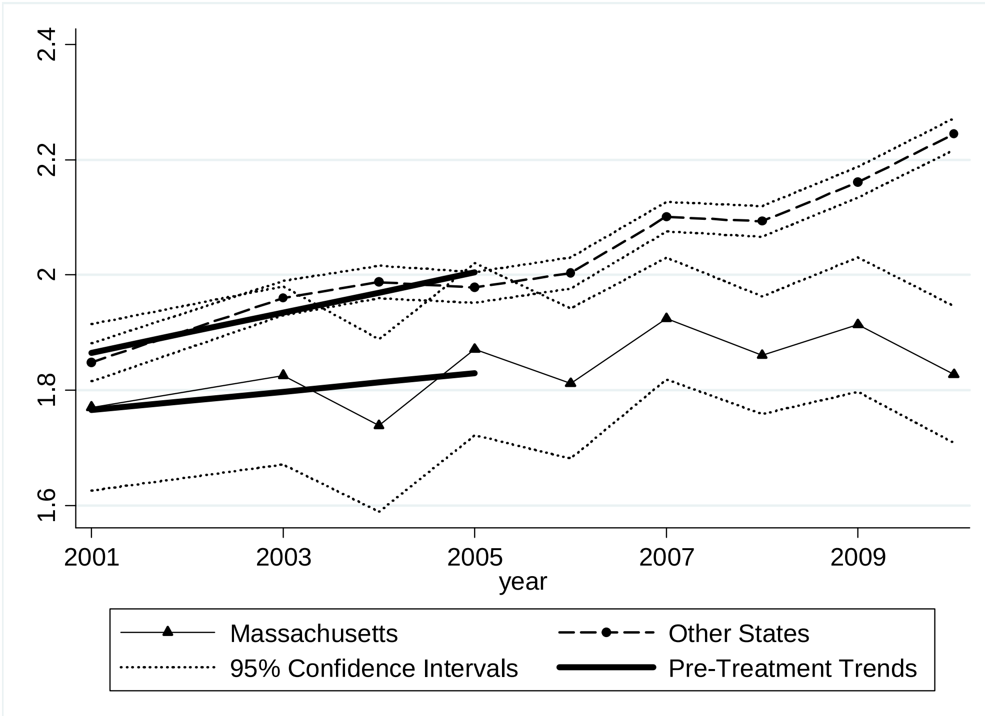

Figure A4 – Changes in Activity-Limiting Joint Pain 2001-2010

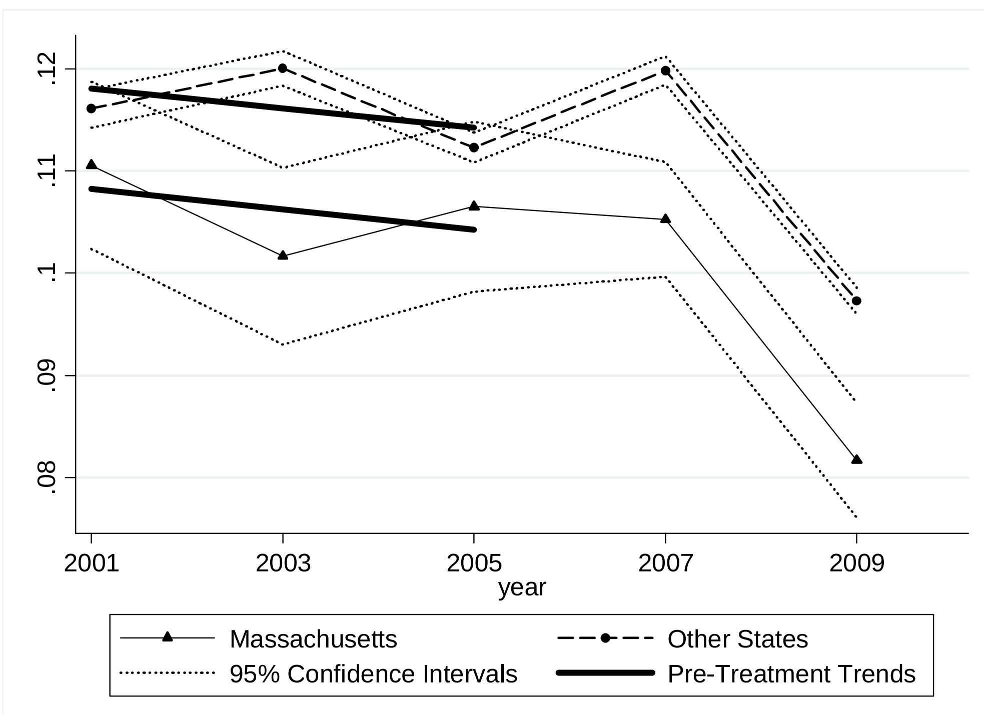

Figure A5 – Changes in BMI 2001-2010

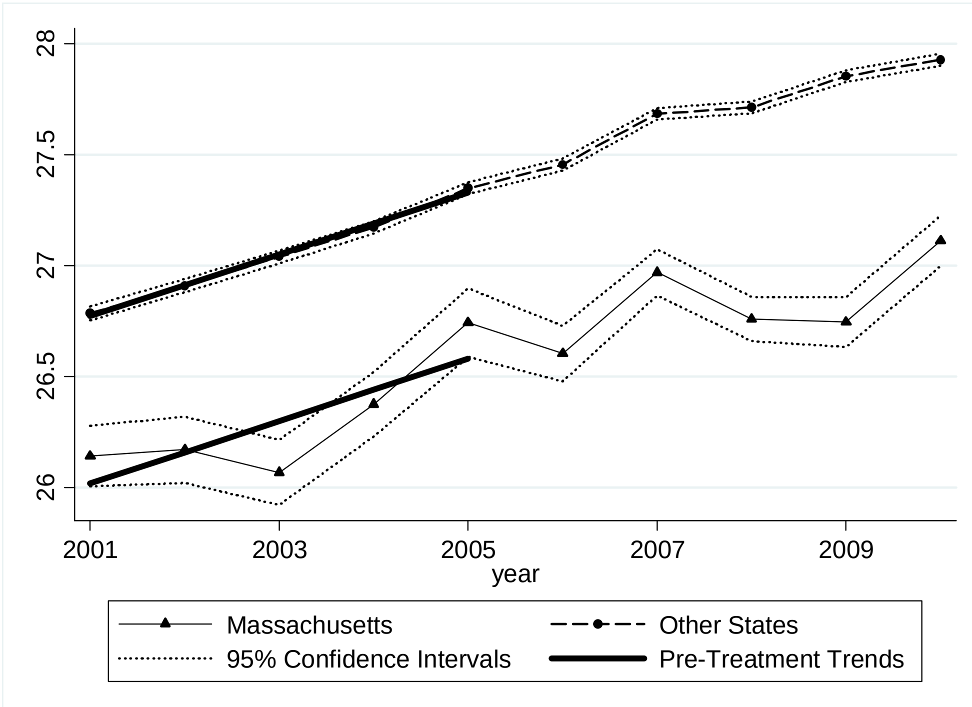

Figure A6 – Changes in Minutes of Exercise 2001-2010

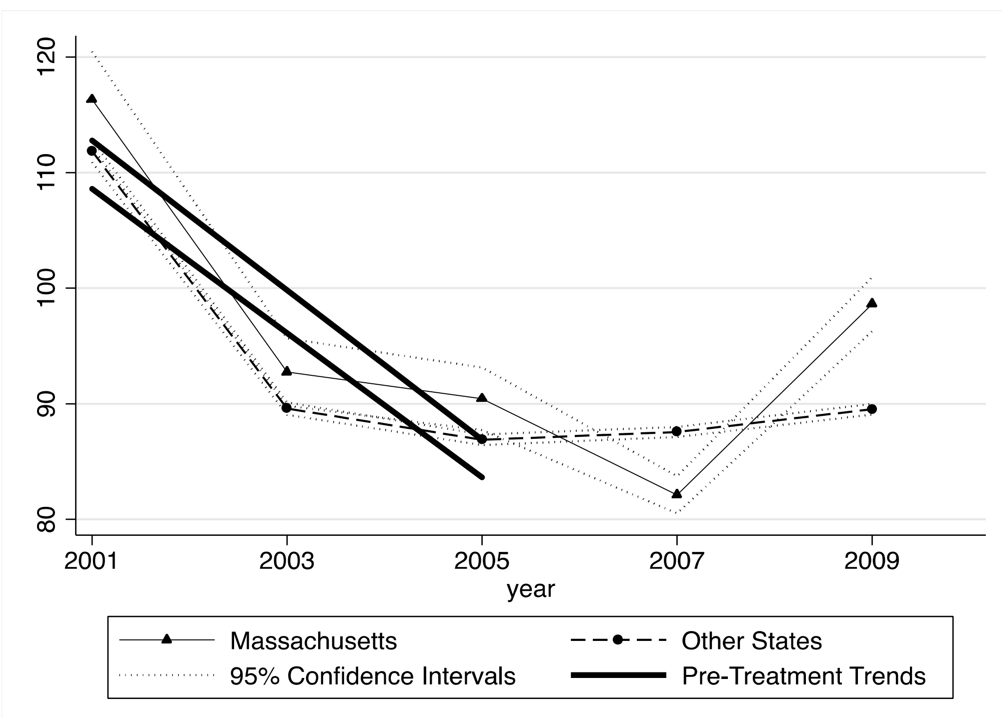

Figure A7 – Changes in Smoking Rate 2001-2010

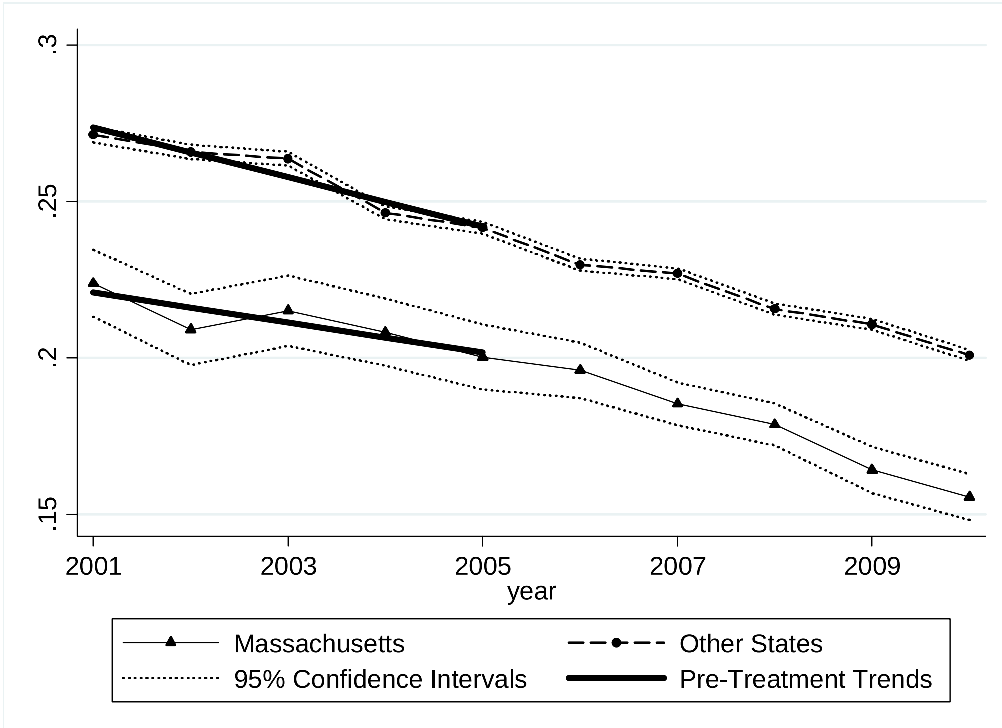

Figure A8 – Changes in Health Status Index for Women

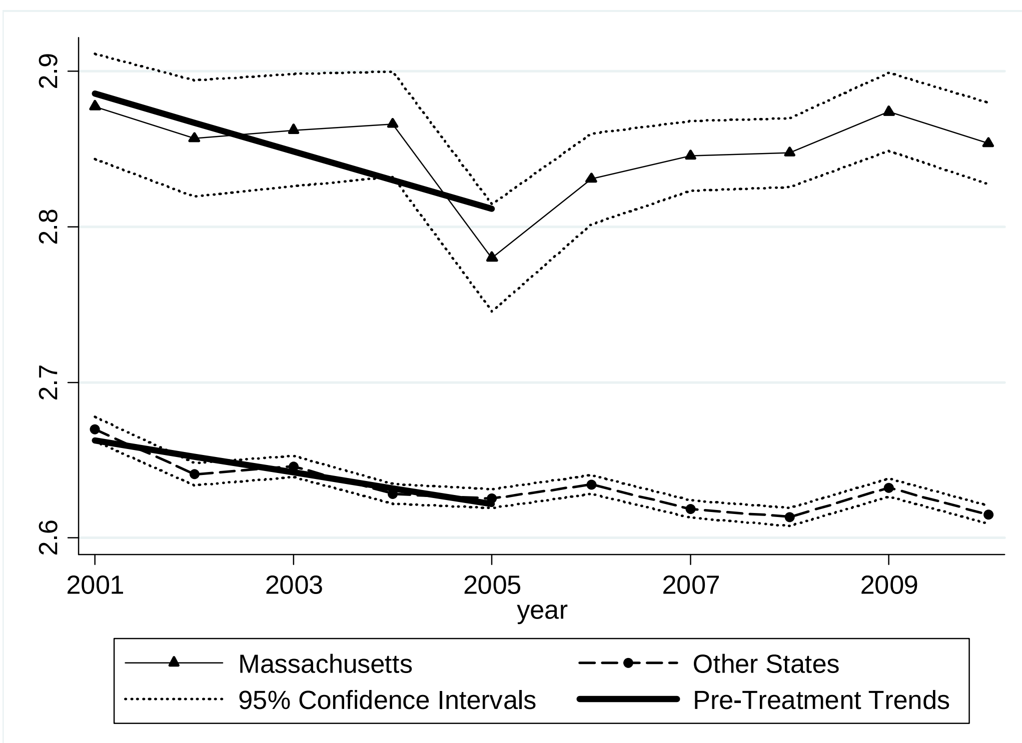

Figure A9 – Changes in Health Status Index for Men

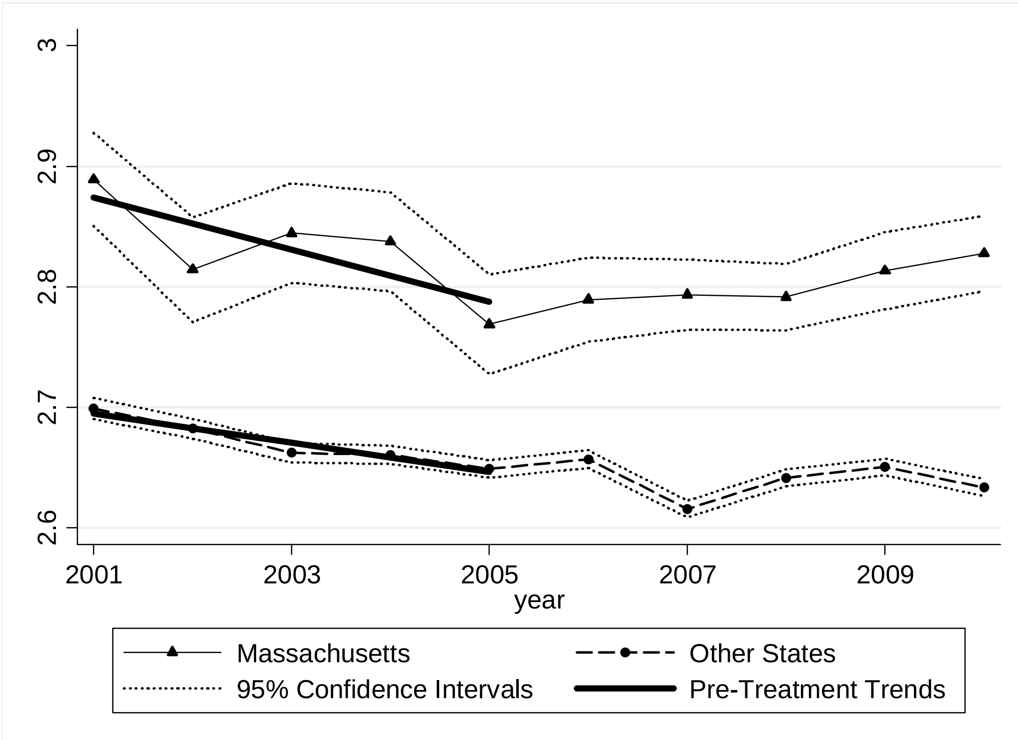

Figure A10 – Changes in Health Status Index for Ages 18-34

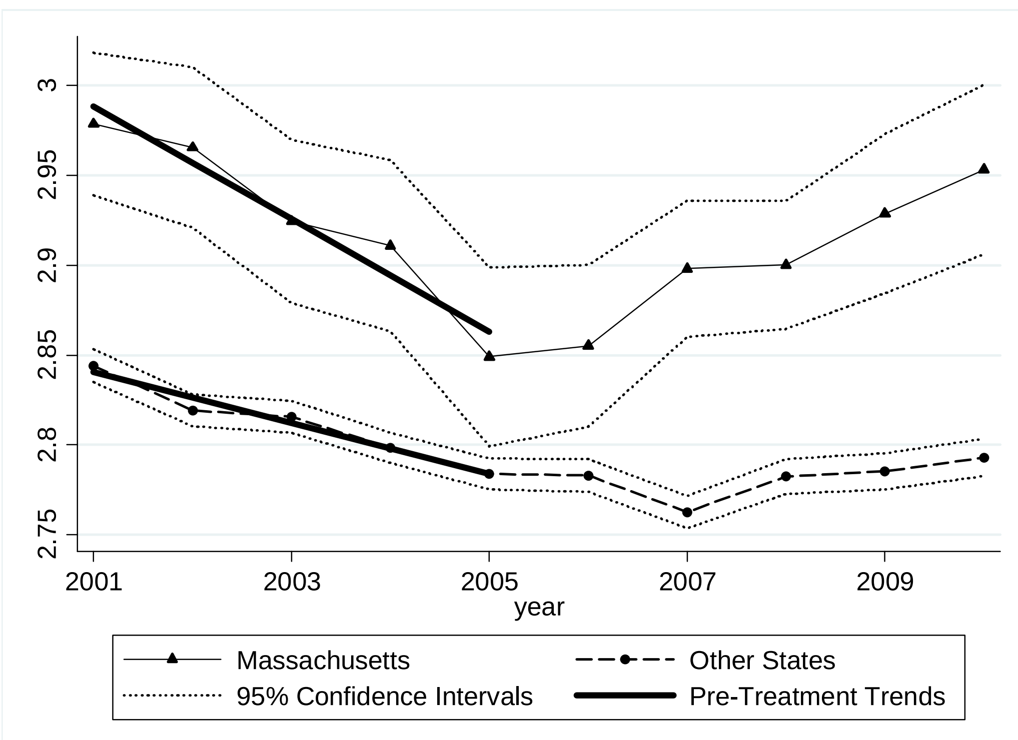

Figure A11 – Changes in Health Status Index for Ages 35-44

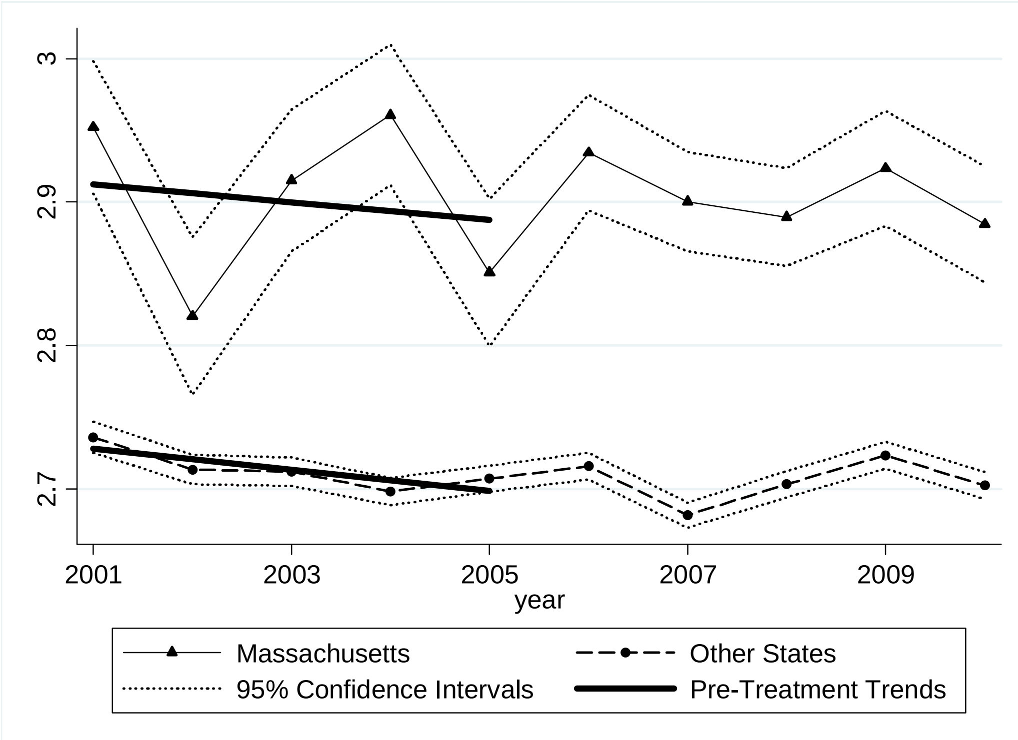

Figure A12 – Changes in Health Status Index for Ages 45-54

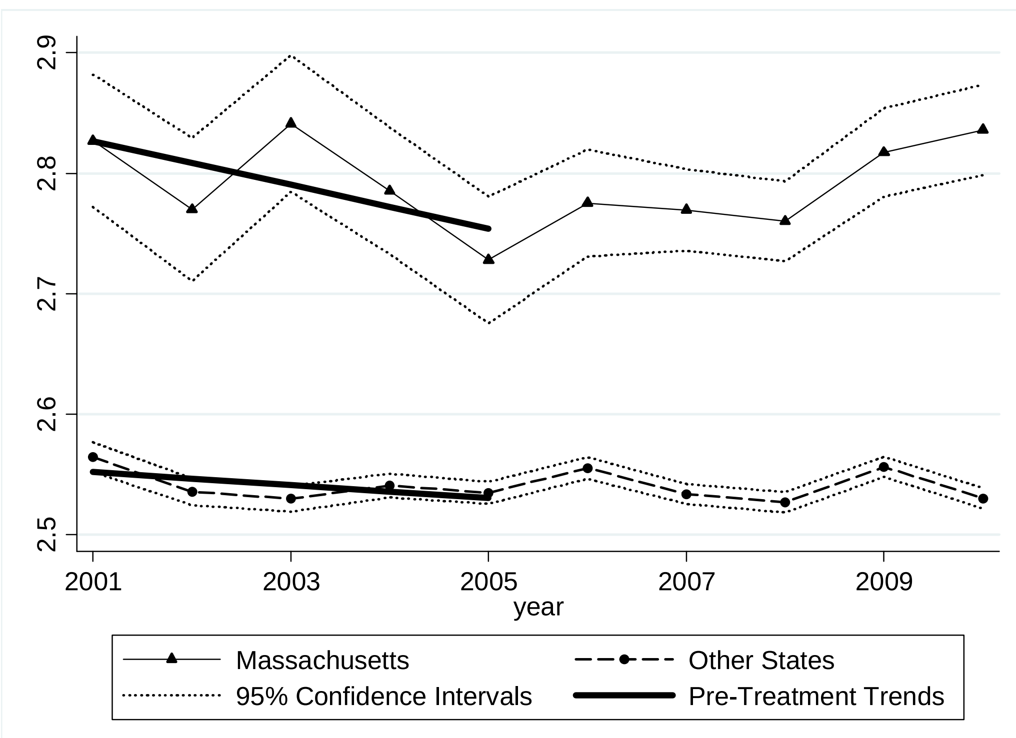

Figure A13 – Changes in Health Status Index for Ages 55-64

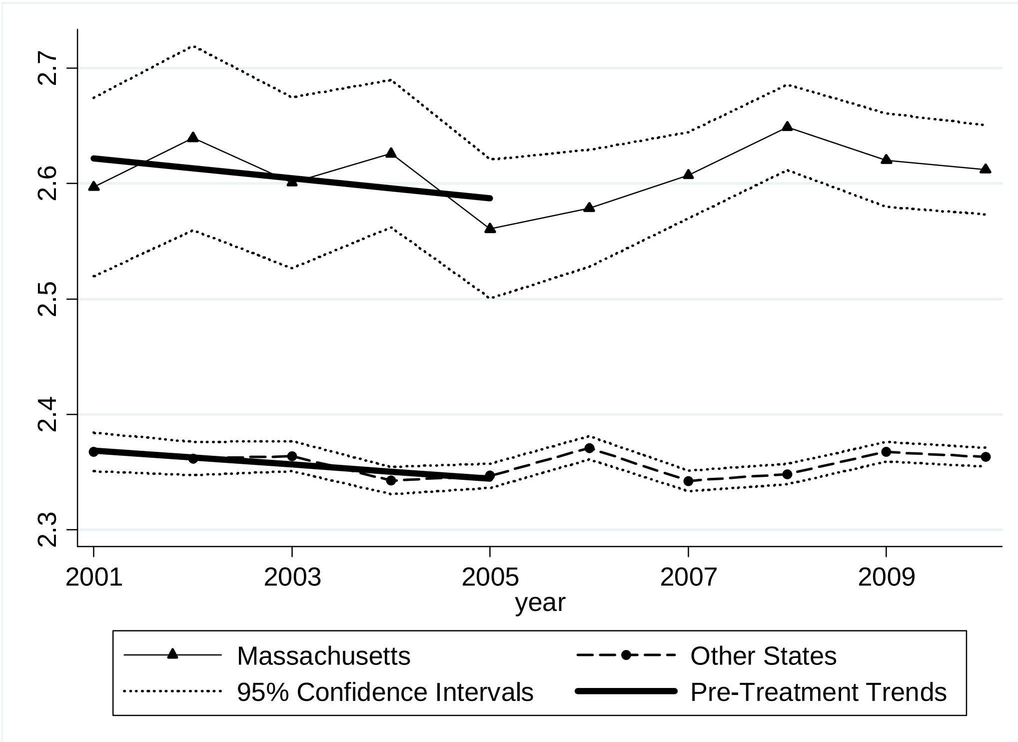

Figure A14 – Changes in Health Status Index for Non-Hispanic Whites

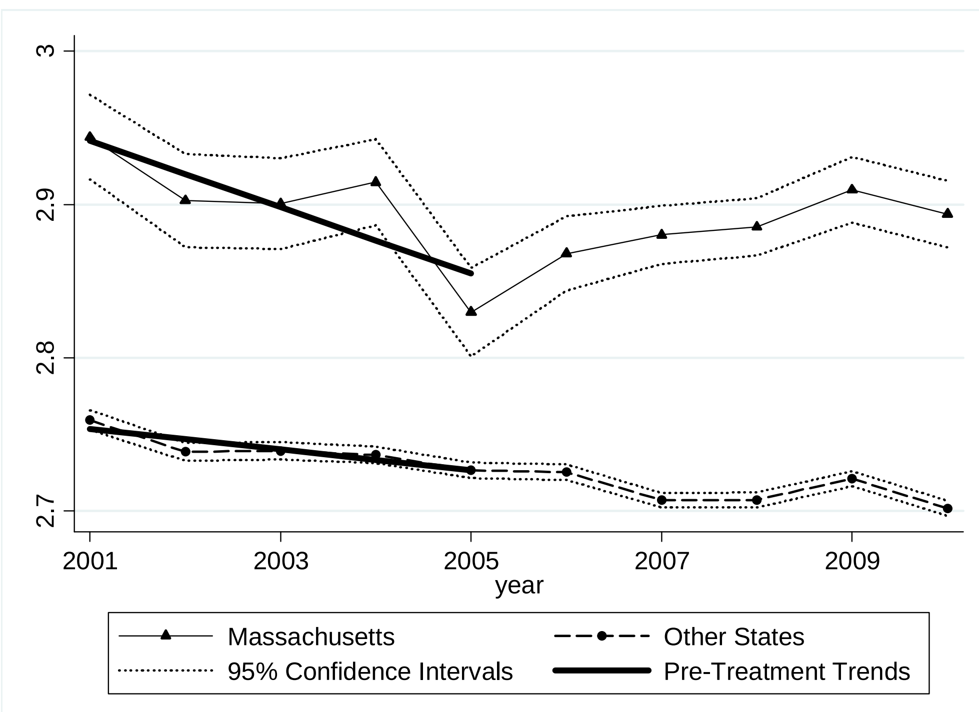

Figure A15 – Changes in Health Status Index for those of a Race/Ethnicity Other than Non-Hispanic White

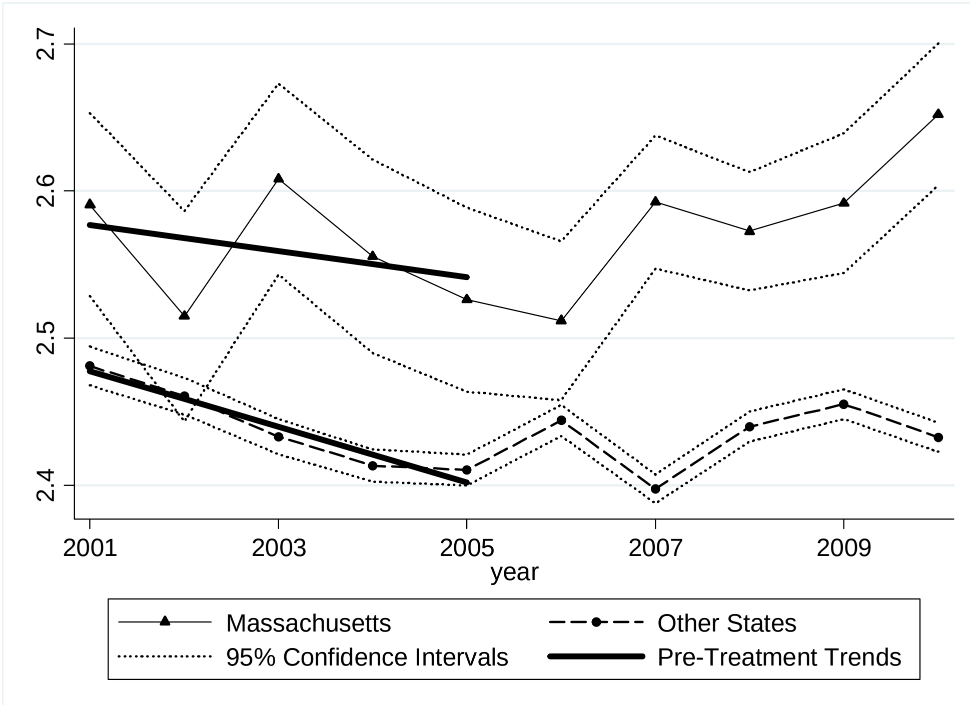

Figure A16 – Changes in Health Status Index for those with Household Incomes Below $25,000

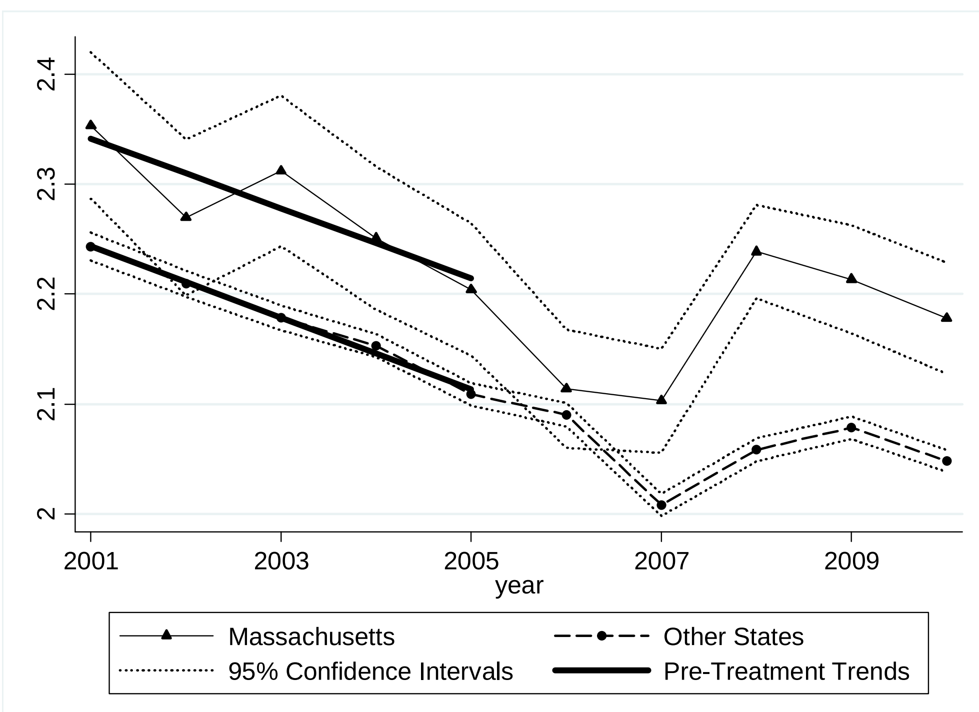

Figure A17 – Changes in Health Status Index for those with Household Incomes Between $25,000 and $75,000

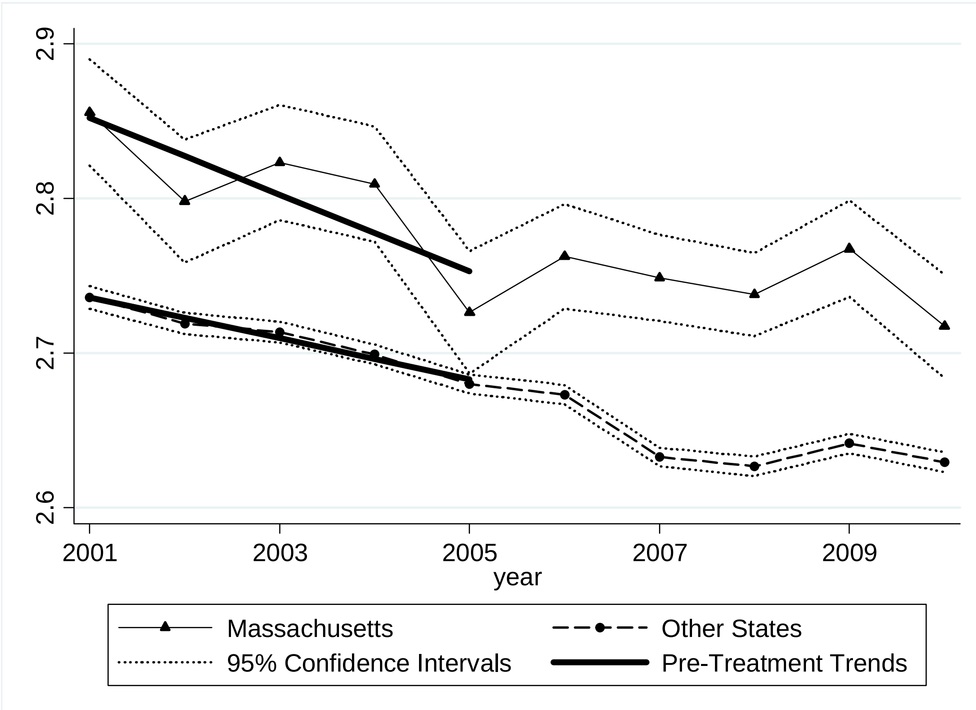

Figure A18 – Changes in Health Status Index for those with Household Incomes Above $75,000

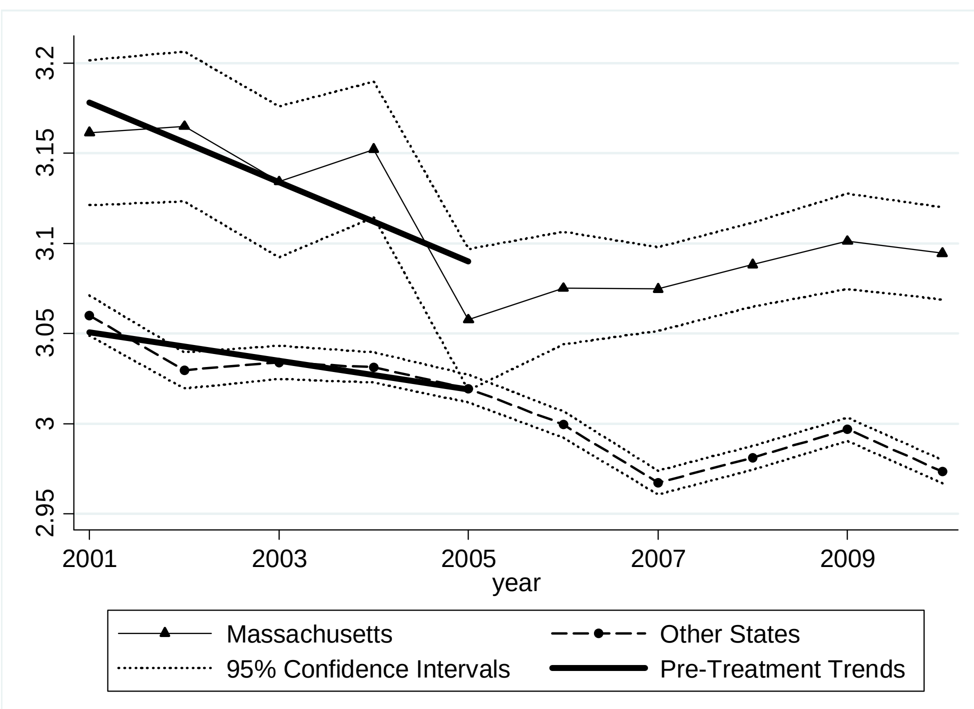

Table A1 – Estimating Effect on Health Using Series of Probits

Dependent Variable:	P(Fair or Better)	P(Good or Better)	P(Very Good or Better)	P(Excellent)
Coefficient Estimates				
MA*During	0.040 (0.010)***	0.030 (0.007)***	0.028 (0.009)***	0.001 (0.007)
MA*After	0.070 (0.006)***	0.062 (0.009)***	0.056 (0.008)***	0.018 (0.007)**
Average Treatment Effect on Treated (After Period)				
MA*After	0.004 (0.0004)***	0.010 (0.001)***	0.018 (0.003)***	0.006 (0.002)**
Effect in Std. Deviations	0.018	0.027	0.037	0.014
Observations	1,979,383	1,979,383	1,979,383	1,979,383

Notes: Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at the 0.1% level; ** 1% level; * 5 % level. All regressions include the individual-level control variables, state fixed effects, and fixed effects for each month in each year. The full control group is used Observations are weighted using the BRFSS sampling weights.

Table A2 – Estimating Effect on Health Using Series of Linear Probability Models

Dependent Variable:	P(Fair or Better)	P(Good or Better)	P(Very Good or Better)	P(Excellent)
Coefficient Estimates = Average Treatment Effects				
MA*During	0.010 (0.0007)	0.005 (0.002)*	0.003 (0.003)	-0.001 (0.002)
MA*After	0.004 (0.0004)***	0.011 (0.002)***	0.016 (0.003)***	0.004 (0.002)
Effect in Std. Deviations	0.024	0.033	0.032	0.009
Observations	1,979,383	1,979,383	1,979,383	1,979,383

Notes: Standard errors, heteroskedasticity-robust and clustered by state, are in parentheses. *** indicates statistically significant at the 0.1% level; ** 1% level; * 5 % level. All regressions include the individual-level control variables, state fixed effects, and fixed effects for each month in each year. The full control group is used. Observations are weighted using the BRFSS sampling weights.

Table A3 – Correlations Between Overall Health and Other Health Outcomes

	Correlation with Overall Health
Days not in Good Physical Health	-0.461***
Days not in Good Mental Health	-0.283***
Days with Health Limitations	-0.387***
Activity-Limiting Joint Pain	-0.318***
BMI	-0.257***
Minutes of Exercise	0.098***
Smoker	-0.160***

*** indicates statistically significant at the 0.1% level; ** 1% level; * 5 % level.
Observations are weighted using the BRFSS sampling weights