Title

Health Care Cost Savings Attributable to the California Tobacco Control Program, 1989 to 2018

Authors
Lightwood, James, PhD
Anderson, Steve

Publication Date
2020-07-26

Health Care Cost Savings Attributable to the California Tobacco Control Program, 1989 to 2018

Final Report[*]

James Lightwood[†] and Steve Anderson[‡•]

Department of Clinical Pharmacy
School of Pharmacy
University of California, San Francisco

July 26, 2020

[*] Supported by California Department of Health and Human Services contract CTCP 19-10620
[†] Department of Clinical Pharmacy, University of California, San Francisco
[‡] JPMorgan Chase & Co., San Francisco, California, United States of America
[•] Analytical Steve Consulting, San Francisco, California, United States of America

This report is available for free download at the University of California eScholarship collection at https://escholarship.org/uc/item/53b9b8fz.

Table of Contents

EXECUTIVE SUMMARY

Background

The California Tobacco Control Program (CTCP) was established in 1989. The program adopted a comprehensive approach designed to change social norms that were previously established which tolerated and encouraged cigarette smoking and other tobacco use. A previous paper published in 2013, 'The Effect of the California Tobacco Control Program on Smoking Prevalence, Cigarette Consumption, and Healthcare Costs: 1985–2008' (Lightwood and Glantz 2013) (ECTCP) estimated the public health impact of the CTCP from 1985 to 2008. This research updates that analysis to the most recent calendar year data available.

Objectives

This research had four objectives

1) Update data used for the 2013 ECTCP estimates to the most recent data available
2) Evaluate the performance of the previously published analysis for the years following 2008
3) Update statistical estimates of the impact of the CTCP using the same, and similar, models using the most recently available data
4) Compare and contrast the 2013 to updated estimates and predictions
5) Predict CTCP program effect up to 2019 (the different years depend the most recent data available).

The measures of impact of CTCP used in this report are

1) effect of a dollar of tobacco control educational expenditure on prevalence of current adult smoking
2) effect of a dollar of tobacco control educational expenditure on mean cigarette consumption per smoker
3) effect of history of CTCP program on annual and cumulative change in
    a. prevalence of current adult current smoking prevalence
    b. mean cigarette consumption per current adult smoker
    c. cumulative person-years of smoking
    d. cumulative value of cigarette sales prevented
    e. real per capita health care expenditure.

Monetary outcomes are in 2019 dollars.

Methods

As in the 2013 research, the CTCP intervention was measured by real per capita tobacco educational expenditure in California compared to 38 control states that have not had a sustained tobacco control program or tobacco control education expenditures.

The four regression equation estimates from the 2013 publication were re-estimated using updated data to 2014 to 2018, depending on the most recent data available. The most recent data available for adult smoking prevalence and mean consumption per smoker is 2018 (for Equations 1 and 2 in the 2013 publication), for the NIPA measure of real per capita health care expenditure is 2017 (Equation 3), and for the CMS measure of health care expenditure is 2014 (Equation 4 in this report).

Data used in the 2013 paper were updated using the same data sources whenever possible. Where updated data were no longer available from the original sources, new sources were found that were as consistent as possible with the data used in the 2013 publication, and consistent over the whole sample period.

The regression coefficients from the 2013 publication were applied to the updated data in order to produce updated forecasts of CTCP program effects. The same regression models from the 2013 publication, and closely related variations, were re-estimated using the updated data, and updated estimates of CTCP program effects were calculated using these regression results. The predictions of the models were compared. The estimates that most accurately reproduced the updated data were used for the final estimates of program effect to 2019.

Results

The data used for the 2013 publication could not be exactly reproduced due to changes in data availability. For this update, we used data that we could find that were most consistent with data used in the 2013 publication. The new estimates were statistically and substantively consistent with those on the 2013 publication. Monetary values are reported in year 2019 dollars.

An additional dollar per year of real per capita education expenditure by CTCP reduces

- current adult smoking prevalence by 0.0591 (SE 0.0114) percentage points.
- mean annual consumption by 2.61 (SE 0.314) packs.

A reduction of adult smoking prevalence of one percentage point reduces real per capita NIPA and CMS health care expenditure by $41.5 (SE $18.3) and $68.5 ($16.7),respectively.

A reduction of mean consumption per smoker of one pack reduces real per capita NIPA and CMS health care expenditure by $2.30 (SE $0.470) and $3.32 (SE $0.508), respectively.

The effect of the CTCP tobacco control program from 1989 to 2018/2019 is summarized below:

- Current adult smoking prevalence was reduced by 2.70 (SE 0.524) percentage points in 2018
- Mean cigarette consumption per current adult smoker reduced by 119 (SE 14.4) packs/year in 2018
- Real per capita NIPA healthcare expenditures were reduced by $632 (SE $164) in 2019
- Real per capita CMS healthcare expenditures were reduced by $955 (SE $147) in 2019

Over the life of the CTCP program, California tobacco educational expenditures are associated with a cumulative reduction in

- 9.45 (SE 1.14) million person-years of smoking to 2018
- 15.7 (SE 3.04) billion packs of cigarette consumption worth $51.4 (SE $6.16) billion in pre-tax sales to the cigarette companies to 2018
- Cumulative savings in the NIPA measure of healthcare expenditures of $500 (SE $129) billion to 2019
- Cumulative savings in the CMS measure of healthcare expenditures of $737 (SE $120) billion to 2019

The NIPA measure is the top-line result for consistency with the previously published estimate; it is highly correlated with the CMS measure.

Conclusions

The original ECTCP model published in 2013 with data through 2008 is stable and predicts the observed data well through 2014 through 2018. CTCP continues to be effective in reducing smoking behavior and attributable health care expenditure in California.

The estimated effect on smoking behavior of an additional dollar spent (adjusted for inflation) on education is the same now has been constant since 2008. The lower estimated total effect of the CTCP program on smoking behavior per year has declined since 2008, but that is due to reductions in real expenditure devoted to the program. If funding were increased, the total effect of the CTCP program should return to its previous levels.

BACKGROUND

The California Tobacco Control Program (CTCP) was established in 1988. The program adopted a comprehensive and integrated approach to tobacco use and prevention aimed at changing social norms that previously tolerated and encouraged tobacco use. A paper published in 2013, 'The Effect of the California Tobacco Control Program on Smoking Prevalence, Cigarette Consumption, and Healthcare Costs: 1989–2008' (Lightwood and Glantz 2013) (ECTCP), estimated that from 1989 through 2008 the CTCP cost $2.81 billion and had a healthcare expenditure savings of $134 billion, which is equivalent to $215 billion in 2019 dollars.

There are two reasons updating the existing estimates are important. The first is that it is of inherent interest to evaluate the continued effect of the CTCP from 2008 to 2019. The second reason is that updating the estimate of the CTCP effect provides an opportunity to validate the previous model and estimate.

The sample data for the main analysis in in ECTCP ended in 2004. ECTCP did include a small out-of-sample forecast validation of part of the model, from the years 2005 to 2008 (which were available at the time the earlier analysis was conducted), because new data was released towards the end of the research. So, the existing out of sample validation for the model is limited to a short-term evaluation immediately after the sample used for estimation. Successfully validating the previous estimate using a new time series sample that extends up to ten years after 2008 would continue to support the efforts of the CTCP. Evaluation of the effect of large public health programs over several decades must rely on observational data, and therefore any estimates are subject to the associated criticism of the quality of that evidence. Successful validation of a model with data not used for the original estimates greatly increases the persuasiveness of the argument that the CTCP expenditure has a stable causal effect on cigarette smoking behavior in California. This report is an update of all the equations in the model that predict adult smoking prevalence, cigarette consumption per adult current smoker, and per capital health care expenditure as a function of the CTCP expenditures and other explanatory variables.

METHODS

The methods of the paper consisted of the following main steps:

1. Update the data used in ECTCP to the most recently available.

2. Re-estimate the identical models used in ECTCP and a related model designed to improve causal inference for the following sample periods:

    a. from 1985 to 2008 for comparison to the published estimates

    b. from 1985 to the most recent data available to update forecasts.

3. Evaluate the estimates and model predictions from Step 2

4. Choose the model which fitted the data best to estimate the impact of the CTCP expenditure from 1989 to 2019, compared to the counterfactual of no CTCP expenditure.

Two technical issues had to be solved in the course of the research. First, some data used in ECTCP were not available until the end of the sample period used for this research. Second, the Behavioral Risk Factor Surveillance Survey (BRFSS) was used for several data series in the model, and its survey methodology underwent a major change in 2011. This change in methodology produced a break in the time series of prevalence of adult smoking and measures of racial and ethnic composition of the population.

Details of the methods are below. Technical details are in the Supplemental Text.

Models

Three models were estimated for this report. The first model is identical to the one published in ECTCP and will be referred to as the 'published' model (Eqs. 1a to 4a), and is used to validate the published results with data not used in the original publication. The second model, called the 'forecast' model (Eqs. 1b to 4b), is a variation of the published model that includes only lagged explanatory variables, specifically including the control state smoking prevalence, cigarette consumption per smoker, and per capita health care expenditures. If the forecast model performed well, it would be preferred for estimating the CTCP impact. Lagging all explanatory variables produces a reduced form regression that reduces the possibility of endogeneity between the California and control state variables, which will provide more convincing argument for a causal relationship between CTCP expenditure and effects on California smoking behavior and health care expenditures. The third model is called the 'final forecast' model (Eqs. 1c to 4c), which was used for estimation of CTCP effect. The final forecast model is a variation on the forecast model, and developed after finding specification problems with the model, namely the possibility of several omitted variables.

All of the models consist of four equations. The first (Eqs. 1a, 1b, 1c) predicts prevalence of smoking, the second predicts consumption per smoker (Eqs. 2a, 2b, 2c) and the third predicts the National Product and Income Accounts (NIPA) measure per capita health care expenditures (Eqs. 3a, 3b, 3c). The NIPA measure of health care expenditures is used for national product and income accounting. It focuses on functional type of good or service produced, and is more strictly focused on medical health care goods and services. The fourth predicts the Centers for Medical and Medicaid Services (CMS) measure per capita health care expenditures (Eqs. 4a, 4b, 4c). The CMS measure of health care expenditures is industry based, and classifies most expenditures by health care provider organizations as health care related. One example of the difference between NIPA and CMS measures is stricter rules to avoid double counting sales of intermediate goods and services in the NIPA measure. The two measures differ in level, but are highly correlated across time.

ECTCP estimated equations for real per capita health care expenditures using both the NIPA and CMS measures. However, the estimate for the CMS measure was conducted as a sensitivity analysis and did not have its own numbered equation in the main text of ECTCP. For clarity, in this report, the CMS and NIPA measures each will have its own equation in this report.

<u>Published Model</u>

Current adult smoking prevalence in California:

$$(prev_{c,t} - prev_{CA,t}) = \alpha_1(EC_{CA,t-1} - EC_{c,t-1}) + \alpha_2(p_{c,t-1} - p_{CA,t-1}) + \alpha_3(y_{c,t-1} - y_{CA,t-1}) + \varepsilon_{1a,t} \quad (1a)$$

Cigarette Consumption per Smoker:

$$(cpsa_{c,t} - cpsa_{CA,t}) = \beta_0 + \beta_1(EC_{CA,t-1} - EC_{c,t-1}) + \beta_2(p_{c,t-1} - p_{CA,t-1}) + \beta_3(y_{c,t-1} - y_{CA,t-1}) + \varepsilon_{2a,t} \quad (2a)$$

NIPA Healthcare Expenditures:

$$n_{CA,t} = \gamma_0 + \gamma_1 n_{c,t} + \gamma_2(prev_{c,t} - prev_{CA,t}) + \gamma_3(cpsa_{c,t} - cpsa_{CA,t}) + \gamma_4(y_{c,t-1} - y_{CA,t-1}) + \varepsilon_{3a,t} \quad (3a)$$

CMS Healthcare Expenditures:

$$h_{CA,t} = \delta_0 + \delta_1 h_{c,t} + \delta_2(prev_{c,t} - prev_{CA,t}) + \delta_3(cpsa_{c,t} - cpsa_{CA,t}) + \delta_4(y_{c,t-1} - y_{CA,t-1}) + \varepsilon_{4a,t} \quad (4a)$$

Where $prev_{j,t}$: Prevalence of current smoking in population *j*, for California and control states in year *t*, in percentage points, $cpsa_{j,t}$: Cigarette consumption per current adult smoker in population *j*, for California and control states in year *t*, in packs/year per smoker, $EC_{j,t}$: Cumulative per capita tobacco control funding in population *j*, for California and control states in year *t*, in dollars, $p_{j,t}$: Price per pack of cigarettes in population *j*, for California and control states in year *t*, in dollars, $y_{j,t}$: Per capita personal income in population *j*, for California and control states in year *t*, in thousands of dollars, $n_{j,t}$: per capita NIPA healthcare expenditures in population *j*, for California and control states in year *t*, in thousands of dollars, $h_{j,t}$: per capita CMS healthcare expenditures in population *j*, for California and control states in year *t*, in thousands of dollars, $\varepsilon_{ka,t}$: Stationary error terms for equation *k*=1 to 3, in year *t*, *j*: Index for population *j* =*CA* for California (intervention), *j* = *c* for control state population, *t*: Time index, *t* =1985 to *T*, and *T* is the most recent observation available.

<u>Forecast Model</u>

Current adult smoking prevalence in California:

$$(prev_{c,t} - prev_{CA,t-1}) = \alpha_1(EC_{CA,t-1} - EC_{c,t-1}) + \alpha_2(p_{c,t-1} - p_{CA,t-1}) + \alpha_3(y_{c,t-1} - y_{CA,t-1}) + \varepsilon_{1b,t} \quad (1b)$$

Cigarette Consumption per Smoker:

$$(cpsa_{c,t} - cpsa_{CA,t-1}) = \beta_0 + \beta_1(EC_{CA,t-1} - EC_{c,t-1}) + \beta_2(p_{c,t-1} - p_{CA,t-1}) + \beta_3(y_{c,t-1} - y_{CA,t-1}) + \varepsilon_{2b,t} \quad (2b)$$

NIPA Healthcare Expenditures:

$$n_{CA,t} = \gamma_0 + \gamma_1 n_{c,t-1} + \gamma_2(prev_{c,t} - prev_{CA,t}) + \gamma_3(cpsa_{c,t} - cpsa_{CA,t}) + \gamma_4(y_{c,t-1} - y_{CA,t-1}) + \varepsilon_{3b,t} \quad (3b)$$

CMS Healthcare Expenditures:

$$h_{CA,t} = \delta_0 + \delta_1 h_{c,t-1} + \delta_2(prev_{c,t} - prev_{CA,t}) + \delta_3(cpsa_{c,t} - cpsa_{CA,t}) + \delta_4(y_{c,t-1} - y_{CA,t-1}) + \varepsilon_{4b,t} \quad (4b)$$

<u>Final Forecast Model</u>:

Current adult smoking prevalence in California:

$$(prev_{c,t} - prev_{CA,t-1}) = \alpha_1(EC_{CA,t-1} - EC_{c,t-1}) + \alpha_2(p_{c,t-1} - p_{CA,t-1}) + \alpha_3(y_{c,t-1} - y_{CA,t-1}) + \epsilon_{1c,t} \quad (1c)$$

Cigarette Consumption per Smoker:

$$(cps_{c,t} - cps_{CA,t-1}) = \beta_0 + \beta_1(EC_{CA,t-1} - EC_{c,t-1}) + \beta_2(p_{c,t-1} - p_{CA,t-1}) + \beta_3(y_{c,t-1} - y_{CA,t-1}) + \epsilon_{2c,t} \quad (2c)$$

Healthcare Expenditures, NIPA measure of health care expenditure:

$$n_{CA,t} = \gamma_0 + \gamma_1 n_{c,t-1} + \gamma_2(prev_{c,t-1} - prev_{CA,t-1}) + \gamma_3(cpsa_{c,t-1} - cpsa_{CA,t-1}) + \gamma_4(y_{c,t-1} - y_{CA,t-1}) + \gamma_5(of_{c,t-1} - of_{CA,t-1}) + \gamma_5(om_{c,t-1} - om_{CA,t-1}) + \gamma_6(w_{c,t-1} - w_{CA,t-1}) + \epsilon_{3c,t} \quad (3c)$$

Healthcare Expenditures, CMS measure of health care expenditure:

$$h_{CA,t} = \delta_0 + \delta_1 n_{c,t-1} + \delta_2(prev_{c,t-1} - prev_{CA,t-1}) + \delta_3(cpsa_{c,t} - cpsa_{CA,t}) + \delta_4(y_{c,t-1} - y_{CA,t-1}) + \delta_5(a_{c,t-1} - a_{CA,t-1}) + \delta_5(hs_{c,t-1} - hs_{CA,t-1}) + \varepsilon_{4c,t} \quad (4c)$$

Where $of_{j,\,t}$: Prevalence of older women (age 45-64) in population *j*, in year *t*, in percentage points for California and control states in year *t*, in packs/year per smoker, $om_{j,\,t}$: Prevalence of older men (age 45-64) in population *j*, in year *t*, $w_{j,\,t}$: Prevalence of women of child-bearing age (age 48-44) in population *j*, in year *t*, $a_{j,\,t}$: Prevalence of elderly (age >= 65) in population *j*, in year *t*, $hs_{j,\,t}$: Prevalence Hispanic ethnicity in population *j*, in year *t*.

Data

All data series used for the new estimates were obtained from the original data sources used in ECTCP, from 1989 to 2008, whenever possible, and updated to the most recent year of data available. This procedure was followed in order to ensure a consistent data set over the whole sample period. All data were available except for two variables as described in the Supplemental text: some price indices were no longer available over the entire period following 2008, and small differences in the CMS measure of health care expenditure over the whole sample.

The intervention population was the aggregate population of California that was exposed to the CTCP. The control population is the aggregate population of 38 control states that had not had a continuous tobacco control program (Abadie A, Diamond A, Hainmueller J 2010), including Alabama, Arkansas, Colorado, Connecticut, Delaware, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, West Virginia, Wisconsin, and Wyoming. Estimates of smoking prevalence are not available for all of the 38 control states starting in 1989; data from 13 states were available as of 1984 and all were available by 1994. As a result, each of the 38 control states contributed to the control population as annual estimates of state smoking prevalence became available.

Mean consumption per smoker was calculated by dividing per capita cigarette consumption for the respective populations by current adult smoking prevalence. The definition of tobacco control funding used for the main analysis included state and federal funding; private funding was omitted, though including or excluding private and federal funding did not make a difference in the results. Per capita healthcare expenditures were calculated by dividing state totals by the state resident populations. Cumulative real per capita tobacco control funding was constructed by summing annual real per capita funding. The values for all variables expressed as dollar amounts were converted to real 2019 dollars using Bureau of Labor Statistics Census Region price indices and the California state price index. The regression data were kept in 2010 dollars in order to facilitate comparison to the ECTCP results. Because of lack of data availability to update the price indices in a consistent way, we felt it would be more reliable to keep the regression data in 2010 dollars and then convert results to 2019 dollars.

The only change in the definition of a variable between ECTCP and the current models are for mean consumption per smoker, $cpsa_{j,t}$, which in this analysis is defined in terms of the adult population, but in the ECTCP was defined in terms of the total population of all ages. This change makes no difference in the effect estimation of the CTCP impact. The change in the denominator only concerns at what stage of the analysis the adjustment for proportion of the population that is adult is made. The change should increase the regression coefficient for CTCP expenditure in Equations 2a-c by the factor (1/0.74), where 0.74 is the average proportion of the California population who are adults over the sample period.

The time spans of the data for estimation were determined by data availability for the dependent variables. The first year of the times series samples was 1989 for all equations. Data for smoking prevalence and cigarette consumption per smoker was available to 2018 (34 observations); CMS and NIPA measures of health care expenditure were available up to 2014 (30 observations) and 2017 (33 observations), respectively. Forecasts, conditional on the observed predictor variables, were calculated to 2018 and 2019 using the last year of available data for the predictor variables. The year 2017 was the last year of the sample period available for regression estimates because the last year of available data for state level tobacco control expenditures is 2016.

See the Data Appendix to this report and Supplemental Text for more details, all data sources, and additional details of variable construction.

Statistical Analysis and Estimate of Program Effect

The estimation methods and the estimation of program effect were the same as in ECTCP. Irrelevant instrumental variables were used to estimate each regression separately as a first order autoregression.

Several variables are affected by a break in BRFSS survey methodology in 2011, which produced state-specific shifts in the state-specific control variables, and changes in measured trends, in the prevalence of adult smoking time series. This level change also affected cigarette consumption per smoker, which is a function of smoking prevalence. The break was modeled using the same techniques used for previously published national panel data estimates (Lightwood and Glantz 2016, Lightwood, Anderson et al. 2020) of the effect of state level smoking behavior on per capita health care expenditure. Following the previous research, a measurement adjustment model was added to the regression to model the effect of the break in survey methodology. A time series break indicator was added to the explanatory variables of regression equations for samples that included observations after 2011. The indicator variables was equal to zero before 2011 and one in 2011 and afterward, in order to model the level shift. Two interaction terms were also added: the prevalence of smoking times the indicator variable, and consumption per current smoker times the indicator variable, in order to model possible changes in trends. Therefore, all estimated regression slope coefficients, forecasts and program effects are adjusted for the change in BRFSS methodology. See Supplemental Text for the regression specification with the measurement adjustment model variables.

Several diagnostic tests were conducted to evaluate regression specifications and assumptions. Parameter stability was assessed using CUSUM tests on recursive residuals of the regression. Ordinary least squares regression residuals were used for the CUSUM tests when recursive residuals could not be calculated over the whole sample, or resulted in a sample of time series of residuals that were too small for a valid test. The stability of the regression coefficients was assessed using recursive regression estimates. Multicollinearity in the explanatory variables was assessed using Variance Inflation Factors (VIFs). Heteroskedasticity was assessed using the Cook-Weisberg test. The presence of a unit root in the residuals was assessed using the Phillips-Perron unit root test. Bartlett's White Noise Test was used to test for independent identically distributed (iid) residuals. Autocorrelation and partial autocorrelation statistics were calculated for the residuals. The five percent significance level was used for each test. The presence of potentially influential observations was assessed using leverage versus squared residuals plots.

The original and updated data sets were compared using the correlation coefficient. The root mean square error (RMSE) of model predictions was used as the primary criterion to evaluate model and estimate forecast adequacy. The correlation coefficient was used as the secondary criterion. The relative RMSE was also calculated. Two RMSE statistics were estimated for the NIPA and CMS per capita health care expenditure. One RMSE was estimated for the health care equation predictions using the observed time series for smoking prevalence and cigarette consumption per smoker used as explanatory variables. A second RMSE was estimated for the health care equation predictions using the predicted time series for smoking prevalence and cigarette consumption per smoker used as explanatory variables using the model estimates. The second RMSE was more important for estimation of CTCP program effect because those estimate are needed to evaluate the counterfactual of no program.

Data processing used the SAS statistical analysis package and R. Statistical analysis used Stata version 16. Simulations used the Yasai excel add-n.

Data Validation and Model Selection

Data and model validation to 2008

The updated data was compared those used in ECTCP from 1989 to 2008. The published model (equations 1a to 3a) was re-estimated using the updated data set to 2008 and compared to the published estimates.

Estimation of published and forecast models using updated data

The published model (Eq. 1a to 3a) was re-estimated using the updated data set from 1989 to the latest year of data available. Regression diagnostics were used to evaluate the quality of the estimates. If the regression estimates failed any of the diagnostic tests, the model or estimation techniques were modified to address any problems that were found.

Evaluation of model predictions

Predictions for the following variables were used to evaluate the models:

1) prevalence of adult current smoking for California

2) mean cigarette consumption per current smoker for California

3) real per capita health care expenditure for California, NIPA measure

4) real per capita health care expenditure for California, CMS measure

The predictions were compared using the following regression estimates:

1) the regression estimates to 2008 in ECTCP published in 2013

2) the published model (Equations 1a to 4a) estimated using the updated data

3) the forecast model (Equations 1b to 4b) estimated using the updated data.

Estimation of program effect

The model that judged to produce the predictions was used to estimate program effect using the same methods as in ECTCP.

RESULTS

Data and model validation to 2008

The correlations between the data used in ECTCP and the updated data set from 1989 to 2009 exceeded 97 percent for all variables in equations 1a to 3a, except for personal income, $(y_{c,t-1} - y_{CA,t-1})$, for which the correlation was only 87 percent. The low correlation in the last variable was due to changes in the BLS price index methodology. After using the California Department of Finance price indices for California, the correlations between the old and updated variables exceeded 97 percent for $(y_{c,t-1} - y_{CA,t-1})$. The correlation was 97 percent for the CMS measure of the health care expenditure for California and control states, and 99 percent for the other variables. The relatively low correlation for the CMS health care expenditure could not be corrected because successive updates of that series do not preserve the exact data values of previous versions.

The top panel of Table 1 shows the estimates of equations 1a to 4a for 1989 to 2008. The updated data passed the regression diagnostic tests, except for finding heteroscedasticity for the equation for NIPA health care expenditure (eq. 3a). The estimated regression coefficients of the ECTCP and updated data for the effect of tobacco educational expenditure on prevalence of smoking (equation 1a) are almost identical for: 0.0497 (SE 0.0035) and 0.0489 (SE 0.0101). One dollar of the CTCP per capita of expenditure is estimated to reduce prevalence by about 5 percentage points in both data sets. The corresponding estimated effect of expenditure on consumption per smoker are statistically significantly different: 1.39 (SE 0.132) and 2.11 (SE 0.271), respectively (P-value for difference = 0.017). One dollar of the CTCP per capita of expenditure reduces consumption per current smoker by 2.11 packs, rather than 1.39 packs estimated in ECTCP. However, adjusting for the change in the denominator of mean consumption per smoker from total population to adults, the coefficient 2.11 is equivalent to 0.73*2.11 = 1.56, which is inside the 95 percent confidence interval for the ECTCP estimate, so the two estimates are consistent.

The main difference between the ECTCP and updated estimates is a much lower coefficient for the effect of smoking prevalence on CMA health care expenditure in ECTCP, -67.8 (SE 7.310), compared to the updated data set, -144 (SE 31.2) (P-value for difference = 0.017), though the coefficient for the effect of consumption per smoker is approximately the same. The discrepant results of CMS health care expenditure may be due to the instability in the CMS time series following updates, and the difficulty in reproducing the per capita personal income.

See Table S1 in the Supplemental Text for the full regression estimates.

Estimation of ECTCP model using updated data and model selection

When estimated over the whole sample period from 1989 to the most recent year of data, both the ‘published’ and the ‘forecast’ models showed two problems. The first problem was the existence of several possibly influential observations in each of the regressions. However, robust regression estimates that reduced the influence of large residuals did not change the estimates. Therefore, influential observations do not appear to be a problem.

The second problem was that estimates for health care expenditure also displayed severe problems with autocorrelation in the residuals. The equations for NIPA health care expenditure (Eq. 3a) and CMS health care expenditure (Eq. 4a) produced residuals that displayed severe autocorrelation, and were not cointegrating (the null of an autogressive unit root in the residuals was not rejected at the 5 percent significance level). The conclusion that the regressions for health care expenditure (Eq. 3a and 4a) are not cointegrating because there is a possibility that the regressions may be spurious and the estimated regression coefficients may be seriously biased. The 'forecast' model estimates behaved similarly, and the health care equations also were not cointegrating over the whole sample period.

The lack of cointegration was addressed by a specification search over several possible omitted variables for the health care equations. Backwards stepwise regression was used to select the best set of additional explanatory variables for the 'forecast' model. The variables considered for inclusion were proportion of the population who were elderly (age >= 65), older adult men (age 45 to 64), older adult women (age 45 to 64), adult women of child-bearing age (age 18 to 44), Hispanic, and African-American, proportion of population enrolled in Medicaid, poverty rate, and unemployment rate.

The variables older men, older adult women and women of childbearing age were retained for the equation for the NIPA health care expenditure. Multicollinearity which affected these three variables did not give a clear indication of which ones were needed to produce a cointegrating regression, so all three were retained. The variables elderly and Hispanic were retained for the equation for the CMS health care expenditure. When these additional variables were included in the 'published' and 'forecast' models, both were cointegrating. The recursive estimate of the coefficients used for estimation of CTCP program effect were very stable over time (See Table S1 in the Supplemental Text for the recursive estimates for the final 'forecast' model).

Evaluation of model predictions

The model predictions of the revised 'published' and 'forecast' models were compared to determine which had the best predictive performance (Table 2).

The RMSEs and correlations for forecasts using the ECTCP program performed well, but were somewhat higher than the models estimated with the updated data. Most of the increase in RMSE was due to the fact that the forecasts from this model could not account for the change in BRFSS survey methodology in 2011.

The RMSEs for the 'published' model are slightly smaller for prevalence and cigarette consumption per smoker (0.736 and 21.237, resp.) than those for the 'forecast' model (1.005 and 25.534, resp.). Neither the 'published' nor the 'forecast' model dominated in terms of the RMSE for health care expenditure equations that used observed smoking prevalence and cigarette consumption per smoker. The RMSE for the 'forecast' model was smaller than the 'published' model for both CMS and NIPA health care expenditure equations that used predicted smoking prevalence and cigarette consumption per smoker. For NIPA health care expenditures the RMSEs for the 'forecast' model and 'published' model were 71.610 and 102.815, respectively. For CMS health care expenditures, the RMSEs were 119.582 and 125.382, respectively. The RMSEs for the health care equation models that used predicted smoking prevalence and cigarette smoking are most important for program evaluation, so the 'forecast' model was chosen for the final model used to update the estimates of the impact of the CTCP program.

The correlations between observed and predicted dependent variables are higher for the 'forecast' model than the 'published' model, except for the prevalence of current smoking, where the 'published' model was slightly higher (Table 3.). For all regression estimates, the relative RMSE for prevalence of current smoking was less than 6 percent, for mean cigarette consumption per smoker less than 13 percent, for NIPA health care expenditure less than 6 percent, and for CMS health care less than 4 percent

The 'forecast' model was chosen to estimate the program effect because it was better than the other models by most measures, and superior in model forecasts needed to estimate updated CTCP program effect.

Updated Model Estimates

The estimated regression coefficients of the final 'forecast' model that were used to estimate the CTCP effect are shown in the bottom panel of Table 1, with a comparison to the ECTCP estimates published in 2013. Only two coefficients from the 'forecast' model differ statistically from ECTCP estimates. In the updated estimates of the 'forecast' model, one dollar of CTCP tobacco expenditure reduces cigarette consumption per smoker by 2.22 (SE 0.316) packs, while it reduced 1.39 (SE 0.132) packs (P-value for difference = 0.015) in ECTCP. As with the replication of results discussed above, after adjustment for the change in the denominator used to calculate mean consumption per current smoker from total to adult population results in an estimate for the 'forecast' model of 0.74*2.22 = 1.63, which is just inside the 95 percent confidence interval for the ECTCP estimate, and again there is no statistically significant difference.

In the updated estimates for NIPA and CMS health care expenditures, none of the estimated regression coefficients were statistically or substantively different than the TCTCP estimates.

In 2019 dollars, a reduction of adult smoking prevalence of one percentage point reduces real per capita NIPA and CMS health care expenditure by $41.5 (SE $18.3) and $68.5 ($16.7), respectively. A reduction of mean consumption per smoker of one pack reduces real per capita NIPA and CMS health care expenditure by $2.30 (SE $0.470) and $3.32 (SE $0.508), respectively.

See Table S1 for the full regression results for the ECTCP estimates using the original data, and the models estimated with the updated data in 2010 dollars.

Figures 1 and 2 show the observed and predicted prevalence of smoking, means cigarette consumption per smoker, for California and the difference between California and the control states. Figure 3 shows the NIPA, and CMS per capita health care expenditures for California. There are two predictions for NIPA and CMS per capita health care expenditures. One prediction (black lines in Figure 3) are for the regressions that use the observed prevalence of smoking and mean consumption per smoker as explanatory variables. The other prediction (dashed gray lines in Figure 3) are for predictions that plug in the predictions for prevalence of smoking (from Eq. 1c) and mean consumption per smoker (from Eq. 2c). This second set of predictions is needed to estimate program effects, where the model predictions for the observed data (with historical CTCP program expenditures) must be compared to the model predictions with CTCP expenditures set to zero (for the counterfactual of no CTCP program). The predictions of the estimated models with updated data track the observed data closely. Only one to two observations fall outside the 95 percent forecast interval for individual observations for any of the predicted time series, as expected with the respective sample sizes.

The model predictions fit the data well, except for the year 2011. The discontinuities at 2011 are due to the break in the BRFSS survey methodology in that year. The other explanatory variables in the regression are adjusted for the presence of the break because of the addition of variables designed to model the effect of the break on the observed data. The model predictions that include the variables that adjust for the existence of the break cannot eliminate the break in the model predictions of the observed data series, which must include the effect of the change in BRFSS survey methodology.

Estimation of CTCP program effect

The regression estimates of the counterfactual case of no CTCP expenditures are shown in Figures 1 to 3 (solid gray lines), for prevalence of smoking, mean cigarette consumption, NIPA and CMS health care expenditures. The difference between the model predictions of prevalence (Figure 1) and cigarette consumption (Figures 2) for the observed data, and the counterfactual of no program expenditure, grows steadily from the inception of the program until around 2005, then stays approximately constant. This change may reflect the reduction in average annual real per capita CTCP expenditures in the last half of the sample.

From 1989 to 2017, the CTCP program reduced adult current smoking prevalence by 2.70 (SE 0.524) percentage points and mean cigarette consumption per current adult smoker reduced by 119 (SE 14.4) packs/year in 2017 (Figure 4). The program prevented 9.45 (SE 1.14) million person-years of smoking and consumption of 15.7 (SE 3.04) billion packs of cigarettes (Figure 5). The total value of cigarettes sales prevented by the CTCP program is $51.4 (SE $6.16) billion in pre-tax sales (Figure 6).

In 2018 Real per capita NIPA and CMS healthcare expenditures were reduced by $632 (SE $164) and $955 (SE $147), respectively (Figure 7). From 1989 to 2018, the program produced cumulative savings in real per capita health care expenditure of $500 (SE $129) billion using the NIPA measure, and $737 (SE $120) billion using the CMS measure (Figure 8). Figures 4 to 8 also show that there are no statistically significant differences between the estimated program effects between the ECTCP estimates the updated estimates.

DISCUSSION

The estimates from the ECTCP remain very stable over the updated data with extend five and nine years beyond the original sample period from 1989 to 2008. The updated data produce estimates of CTCP effect that are substantively and statistically consistent with the ECTCP estimates published in 2013. Updated estimates of CTCP program that used the ECTCP regression estimates would not be substantially different from those that use the updated regression estimates that are presented here.

The differences between the ECTCP estimates of program effect appear to be driven by changes in the data following 2008. The difference between California and control state population real per capita CTCP program expenditure fell from an average of $2.87 (in 2019 dollars) from 1989 to 2008 by more than half in the following years. The vast majority of the control states continued to have very low, though slowly growing, per capital tobacco control expenditure, so most of the change is due to reductions in annual real per capita expenditure in California. This reduction is only good in the sense that, for purposes of estimation of program effect, it increases the variation in explanatory variables. Therefore, it allows more precise estimates of program effect, other things held constant. It also shows that the regression results and estimates of CTCP program effect are robust to changes in the properties of the time series. However, it also indicates that reductions in the growth in total savings attributable to the program are mostly caused by reduction in program effort.

The limitations of this analysis, like that in ETCTP, is that it uses historical observational data, and therefore has from the limitations of that study design. Particularly with respect to imputation of causal effects. However, the argument for causal effect is stronger in this report than in ECTCP for three reasons. First, the original estimates are stable and robust out of sample over six to nine years, despite the fact that the properties of the time series change out of sample. Second, the final estimates used for estimation of program effect in this report used lagged dependent variables. This change strengthens the argument that the statistics have a causal interpretation because of the time sequencing of all of the explanatory variables, including the variables that reflect the difference between California and control states for the time series of policy interest: California smoking behavior and per capita health care expenditure. Third, the results are robust to changes in regression specification.

CONCLUSION

Since 1989, the CTCP has prevented a total of about 9 million person-years of smoking and 16 packs of cigarette consumption, and between $500 billion and $737 billion in health care expenditure, depending on the measure used.

The model and results of the ECTCP analysis in 2013 are very stable, and generalize to nine years out of the sample used for those estimates. The CTCP continues to have the same effectiveness estimated in 2013 per each program dollar spent. The total impact of the program is growing at a slower rate since 2008 mainly because of lower intensity of program effort, that is, because of lower expenditure on tobacco education and other program activities since 2008 compared to previous years.

Dependent variable Sample period 1985 to 2008	coefficient for:	2013 publication 1985 - 2008		updated data 1985 - 2008		test for difference	
		Point est.	SE	Point est.	SE	z-score	p-value
CA prevalence (Eq. 1a)	tobacco educational expenditure	0.0497	0.0035	0.0489	0.0101	0.076	0.940
CA mean consumption (Eq. 2a)	tobacco educational expenditure	**1.39**	**0.132**	**2.11**	**0.271**	**2.39**	**0.017**
Real per capita NIPA healthcare	Prevalence	-35.4	9.85	-63.7	16.5	1.473	0.141
Expenditure (Eq. 3a)	mean consumption	-3.14	0.786	-3.16	0.401	0.023	0.982
Real per capita CMS healthcare	Prevalence	**-67.8**	**7.31**	**-144**	**31.2**	**2.378**	**0.017**
Expenditure (Eq. 4a)	mean consumption	-5.48	0.928	-5.69	0.848	0.167	0.867
Dependent variable Sample period 1985 to 2014, 2017, and 2018, depending on dependent variable.	coefficient for:	2013 publication 1985 - 2008		Updated data 1985 - most recent		test for difference	
		Point est.	SE	Point est.	SE	z-score	p-value
CA prevalence (to year 2018) (Eq. 1a, c)	tobacco educational expenditure	0.0497	0.0035	0.0503	0.0097	0.058	0.954
CA mean consumption (to year 2018) (Eq. 2a, c)	tobacco educational expenditure	**1.39**	**0.132**	**2.22**	**0.316**	**2.42**	**0.015**
Real per capita NIPA healthcare	Prevalence	-35.4	9.85	-53.2	23.5	0.699	0.485
expenditure (to year 2017) (Eq. 3a, c)	mean consumption	-3.14	0.786	-2.95	0.602	0.192	0.848
Real per capita CMS healthcare	Prevalence	-67.8	7.31	-87.6	21.4	0.876	0.381
Expenditure (to year 2014) (Eq. 4a, c)	mean consumption	-5.48	0.928	-4.26	0.650	1.077	0.282

Table 1.-Comparison of 2013 publication coefficient estimates and estimates with updated data, using ‘forecast’ version of model
Note: Bold indicates statistically significant different between 2013 publication and updated estimates.
Note: expressed in 2010 dollars.

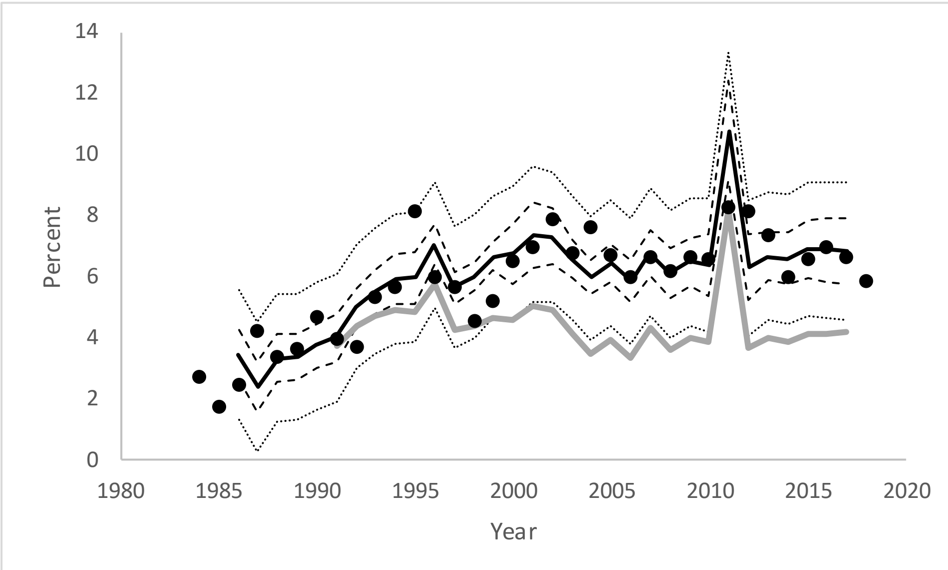
Dependent Variable to be predicted	Root Mean Square Error of Predictions		
	ECTCP estimates applied to updated data to most recent data available	'Published' Model with updated data to most recent data available	'Forecast' model with updated data* to most recent data available
Prevalence (s), control - CA	0.869	0.736	1.005
Consumption (cpsa), control - CA	47.948	21.237	25.534
Per capita real NIPA healthcare expenditure, CA			
with observed s and cpsa	159.210	45.521	59.768
with predicted s and cpsa	--	102.815	71.610
Per capita real CMS healthcare expenditure, CA			
with observed s and cpsa	204.309	109.876	98.428
with predicted s and cpsa	--	125.382	119.582

Table 2.-Root Mean Square Error of predicted versus observed
*'Forecast' model means all explanatory variables are lagged. Yellow highlighting indicates estimates with best results

Dependent Variable to be predicted	Correlations			
	ECTCP estimates applied to old data used in 2013 publication	2013 estimates applied to updated data to most recent data available	Re-estimated 2013 model with updated data to most recent data available	Estimated 'forecast' version 2013 model with updated data* to most recent data available
Prevalence (s), control - CA	0.875	0.860	0.874	0.786
Consumption (cpsa), control - CA	0.891	0.893	0.713	0.851
Per capita real NIPA healthcare expenditure, CA				
with observed s and cpsa	0.816	0.976	0.870	0.959
with predicted s and cpsa	--	--	0.532	0.535
Per capita real CMS healthcare expenditure, CA				
with observed s and cpsa	--	0.970	0.950	0.972
with predicted s and cpsa	--	--	0.956	0.960

Table 3.-Correlations of predicted versus observed
*'Forecast' model means all explanatory variables are lagged. Yellow highlighting indicates estimates with best results.

Panel a.



Panel b.

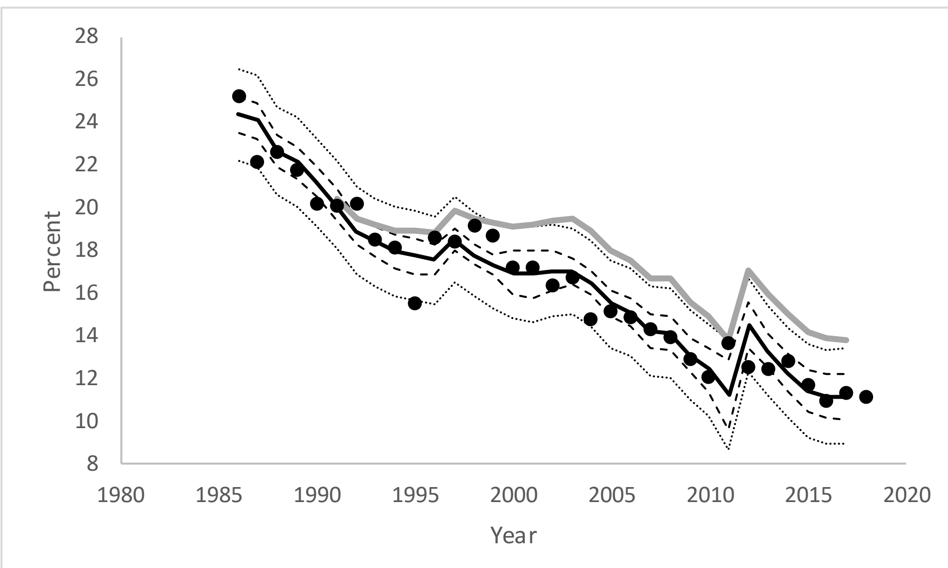
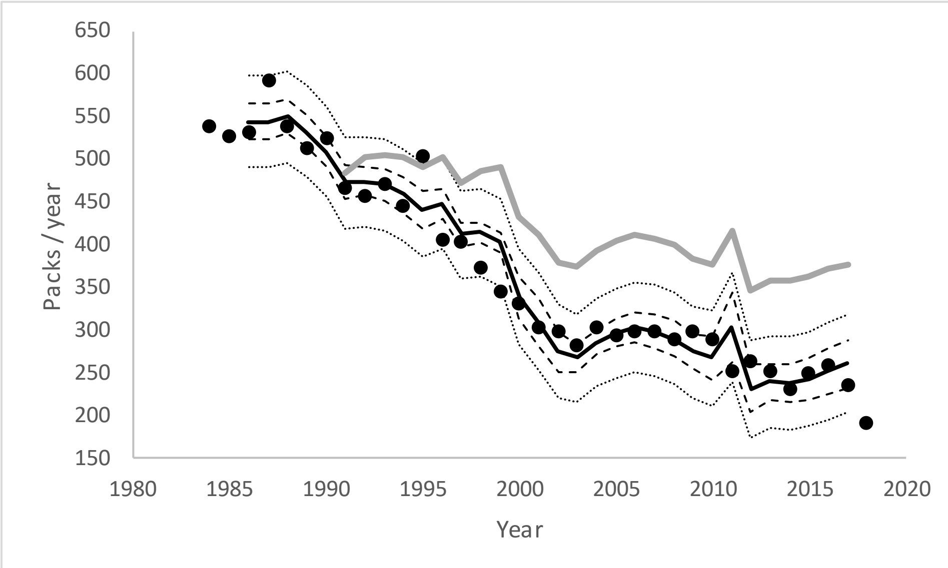


Figure 1. – Predictions for prevalence of adult current smoking (Eq. 1c). Panel a: control states – California; Panel b: California

Black circles: observed time series; Solid black line: model prediction; Dashed black lines: 95 percent confidence interval for prediction; Dotted black lines: 95 percent forecast interval for individual observations; Solid gray line: model prediction for hypothetical of no CTCP program.

Panel a.



Panel b.

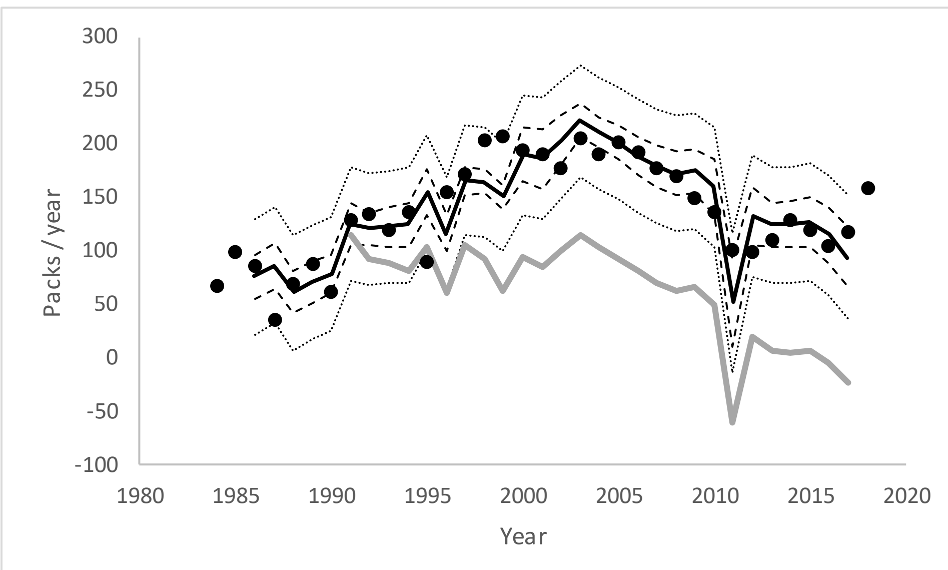
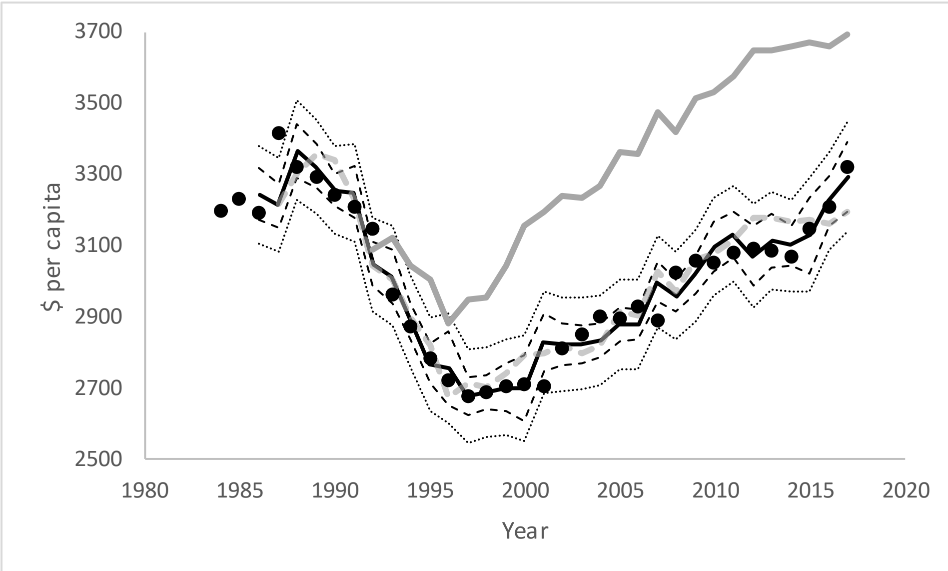


Figure 2. – Predictions for mean cigarette consumption per current smoker (Eq. 2c). Panel a: control states – California; Panel b: California

Black circles: observed time series; Solid black line: model prediction; Dashed black lines: 95 percent confidence interval for prediction; Dotted black lines: 95 percent forecast interval for individual observations; Solid gray line: model prediction for hypothetical of no CTCP program.

Panel a



Panel b

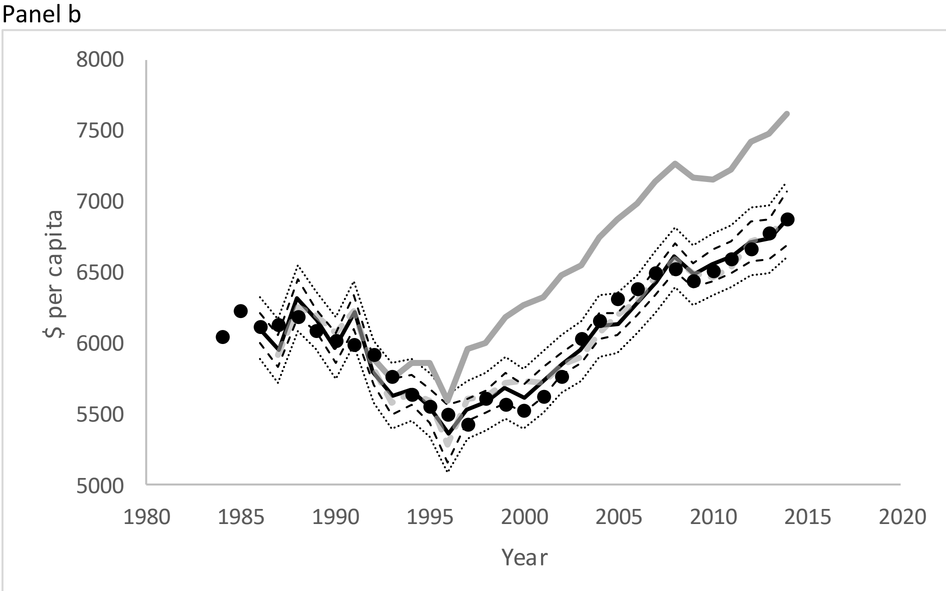
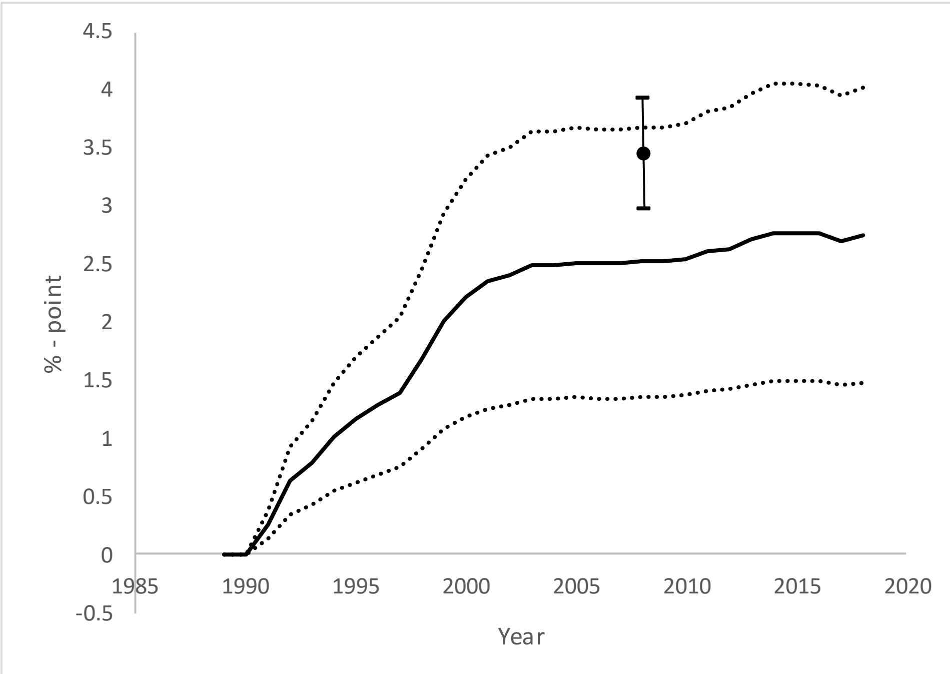


Figure 3. – Predictions for real per capita health care expenditure, California. Panel a: NIPA measure, (Eq. 3c), Panel b: CMS measure (Eq. 4c)
Black circles: observed time series; Solid black line: model prediction using observed smoking prevalence and mean consumption; Dashed gray line: model prediction using predicted smoking prevalence and mean consumption; Dashed black lines: 95 percent confidence interval for prediction; Dotted black lines: 95 percent forecast interval for individual observations; Solid gray line: model prediction for hypothetical of no CTCP program.
Note: vertical axes are real 2019 dollars

Panel a



Panel b

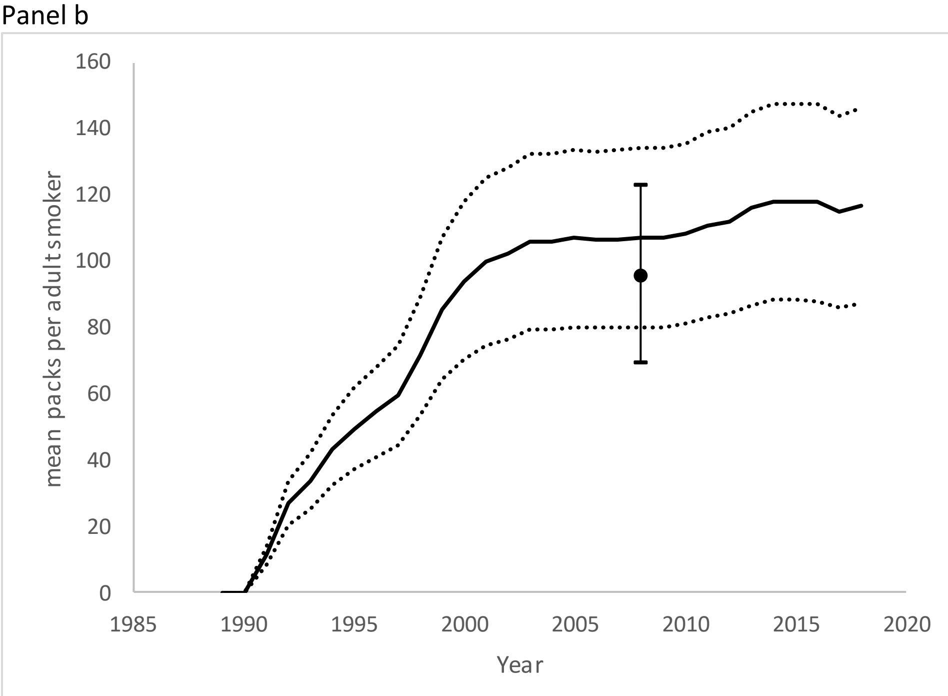
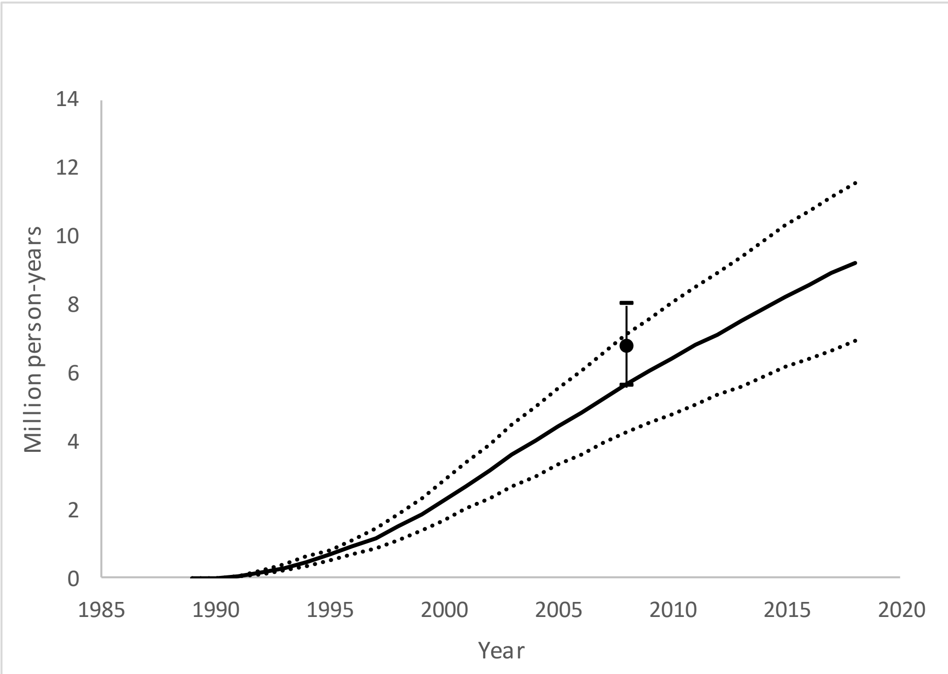


Figure 4.—Annual savings attributable to the CTCP program, 1990 to 2020: Panel a, adult smoking prevalence, Panel b, mean packs per current smoker.
Black line: mean; Dotted lines: lower and upper 95% confidence intervals for prediction; Black dots and error bars: point and 95% confidence interval reported in 2013 publication

Panel a



Panel b

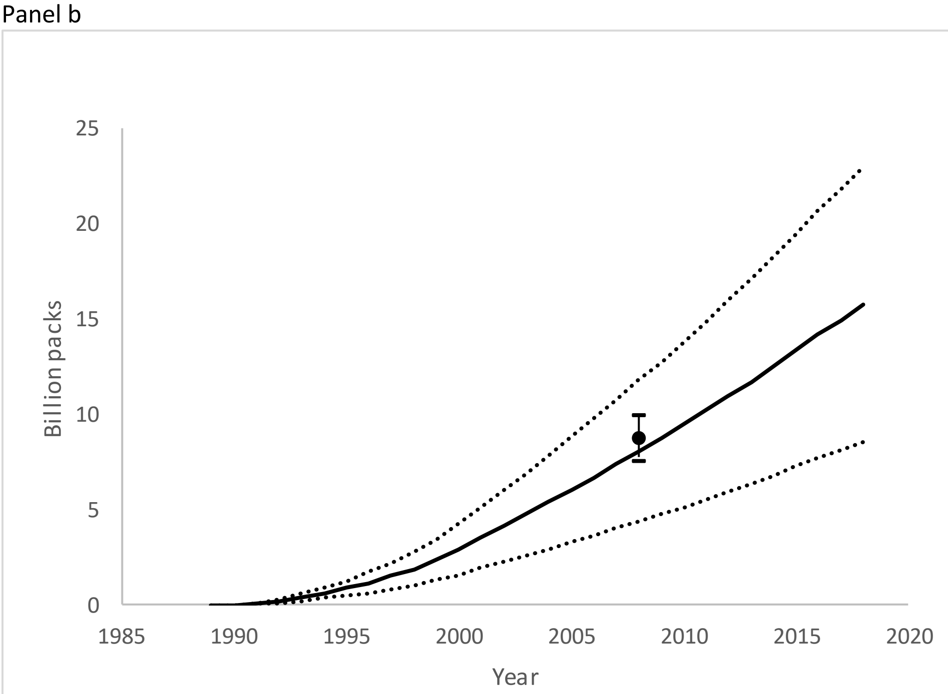
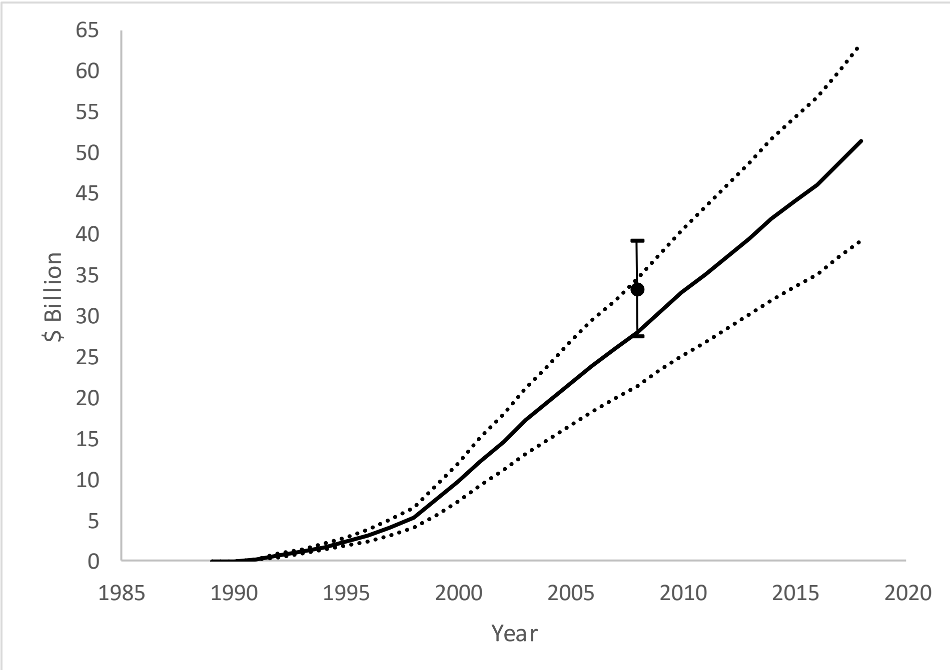


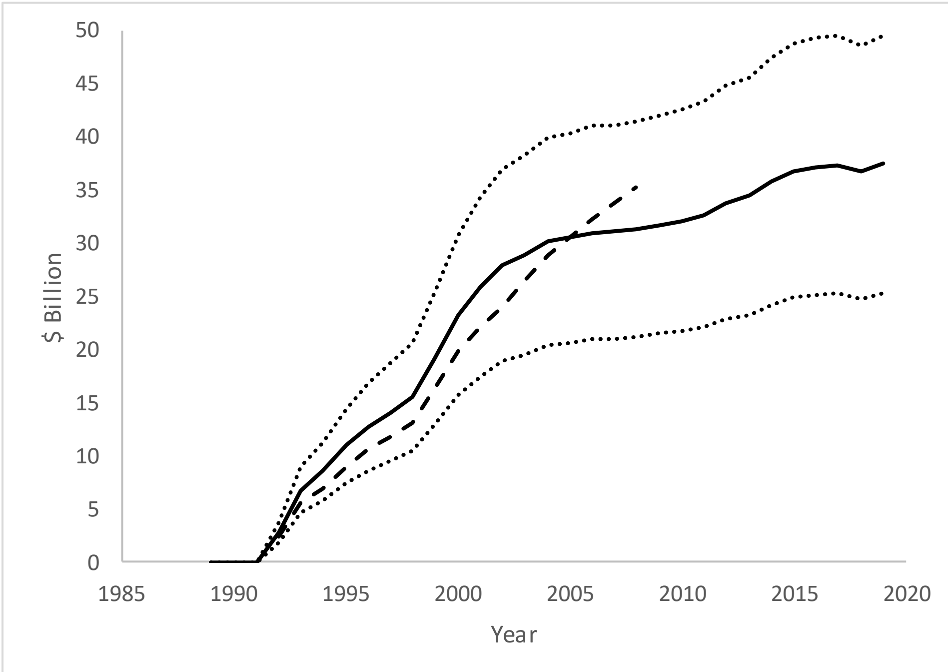
Figure 5.—Cumulative savings attributable to the CTCP program, 1990 to 2020: Panel a, adult smoking prevalence, Panel b, mean packs per current smoker.
Black line: mean; Dotted lines: lower and upper 95% confidence intervals for prediction; Black dots and error bars: point and 95% confidence interval reported in 2013 publication



Reduction in cigarette sales

Figure 6.—Cumulative savings in value of cigarette sales attributable to the CTCP program, 1990 to 2020.
Black line: mean; Dotted lines: lower and upper 95% confidence intervals for prediction; Black dots and error bars: point and 95% confidence interval reported in 2013 publication
Note: vertical axes are real 2019 dollars

Panel a



Panel b

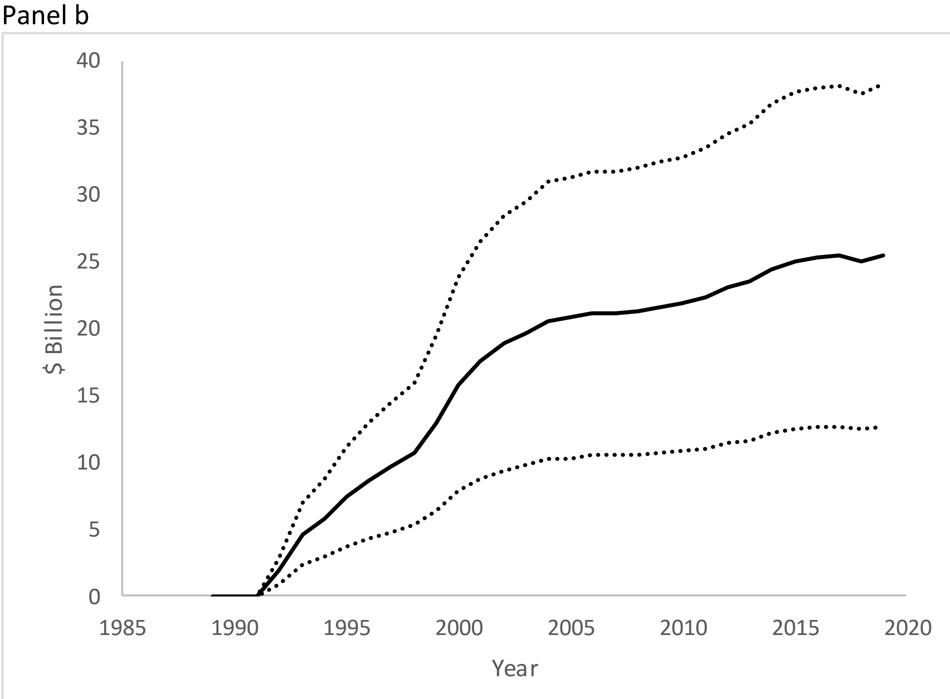
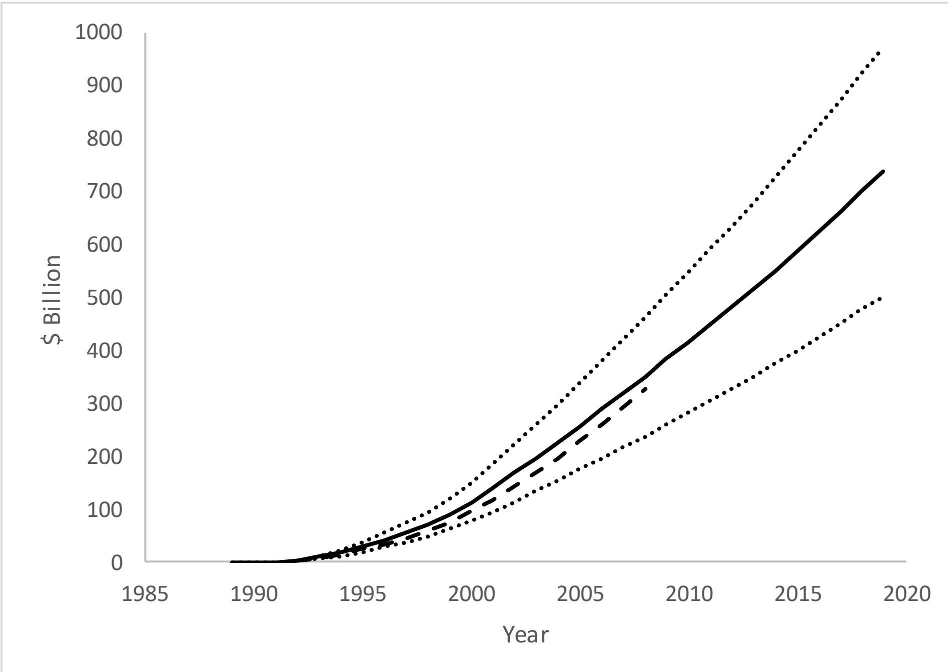


Figure 7.—Annual savings in real per capita health care expenditure attributable to the CTCP program, 1990 to 2020: Panel a, NIPA measure, Panel b, CMS measure.
Black line: mean; Dotted lines: lower and upper 95% confidence intervals for prediction.
Dashed black line in Panel a: estimated savings from ECTCP, 2016.
Note: vertical axes are real 2019 dollars

Panel a



Panel b

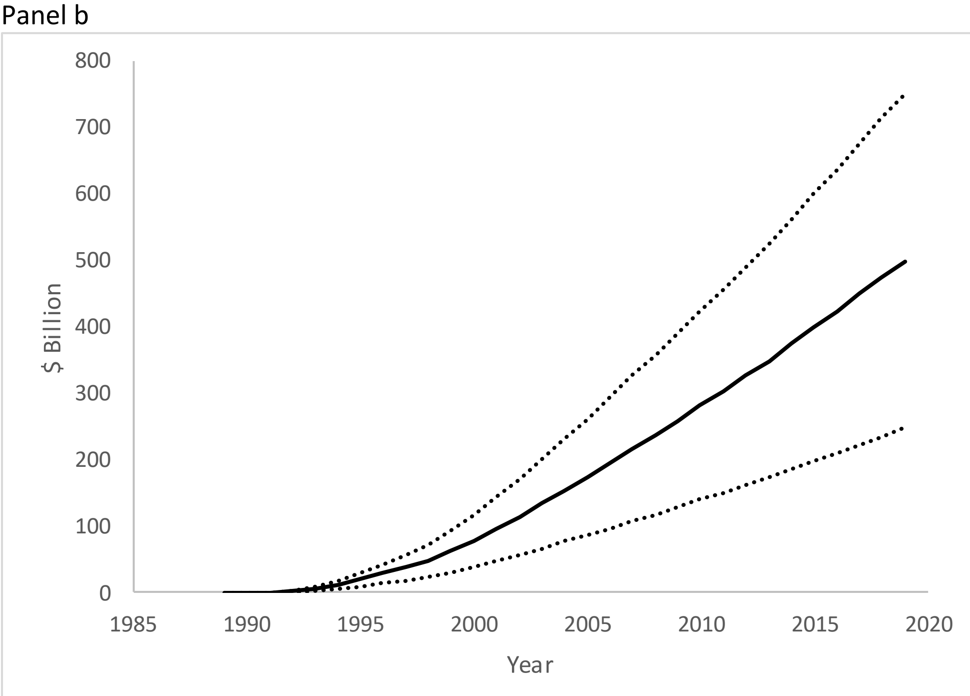


Figure 8.—Cumulative savings in real per capita health care expenditure attributable to the CTCP program, 1990 to 2020: Panel a, NIPA measure, Panel b, CMS measure.
Black line: mean; Dotted lines: lower and upper 95% confidence intervals for prediction.
Dashed black line in Panel a: estimated savings from ECTCP, 2016.
Note: vertical axes are real 2019 dollars

DATA APPENDIX

Data Processing Summary Info
This document or section lists the data to acquire and process to produce estimates of historical returns to tobacco control expenditures using existing models. It includes the data source and processing used to produce data sets needed to run each model to the last year of data available. Essentially, we acquired the latest data from various sources in the required format detailed in this data request. The list of variables (or features) is listed below, along with the source and other pertinent details. Each variable is a different data source and dataset. All datasets are processed in a way to append key additional state variables, which are state name, state abbreviate and state FIPS code. This will allow anyone who uses the data to conveniently match, merge and join to one another. In the dataset file name, the variables and the latest year the data is available is in the file name.

File Format Details*			
File Format	CSV	Delimited	Comma
Text Qualifier	“”	File Extension	*.CSV

**We downloaded each variable from its respective data sources and saved each file in csv format. The data sources are listed later in this document.*

High Level Steps

1. Download Files (download variables)
2. Transpose Data with Files (transpose rotates data from rows to columns or vice versa)
    a. Some files have year as a column but we need it as the data in panel data format; therefore, these files need to be transposed. Panel data format is longitudinal data or cross-sectional time series data.
3. Create Deflators from CPU Index file based on 2010 dollar.

	A	B	C	D	E	F
1	Behaviors – Cigarette use					
2	Cigarette Use (Adults) – BRFSS					
3						
4	Current - Overall					
5	State	1984	1985	1986	1987	1988
6	Alabama			25.4	27.5	26.4
7	Alaska					
8	Arizona	27.8	27	24.6	26.8	24.3
9	Arkansas					
10	California	26.3	26.3	25.3	22.2	22.6

	A	B	C	D	E
1		BEA_Reg	State_Nar	Year	s_s
2	1	5	Alabama	2016	21.5
3	2	5	Alabama	2015	21.4
4	3	5	Alabama	2011	24.3
5	4	5	Alabama	2012	23.8
6	5	5	Alabama	2014	21.1
7	6	5	Alabama	2013	21.5
8	7	8	Alaska	2014	19.9
9	8	8	Alaska	2015	19.1
10	9	8	Alaska	2013	22.6

4. Gather additional files need for final file
    a. Additional files provided within this document below
5. Transform Data for final file
    a. Turn nominal dollar amounts to real dollar amounts using the respective deflators (2010 $). Note there are several different deflators. This document specifies which deflator to use.
    b. Use logic to create new variable
    c. Use some variables as is
    d. Merge some variables in from additional files provided

1. **Download Files (Key for variable names x_s_n: x is variable name, s is state index, n is time index)**

Variable to Download					
Feature Name	**Description**	**Source or Download Link**	**Source or Download Details**	**Additional Source or Download Details**	**Transpose**
NIPA nhce	total nominal NIPA health care expenditure by state	https://www.bea.gov/regional/downloadzip.cfm	https://www.bea.gov/regional/index.htm	Bureau of Economic Analysis NIPA Data on health expenditures, and personal income NIPA Health Care expenditure, personal income, and resident population, annual estimates by individual state and DC Frequency: annual, calendar year There will be two files here. One with years after 1996 using NCAIS and before 1997 using SIC. There are codes to identify the specifics health care expenditure which will need to be broken for some of the calculation for the final file.	Transpose
ec_s_n	nomina1 Annual tobacco educational funding, total	https://www.cdc.gov/statesystem/index.html	https://www.cdc.gov/tobacco/about/osh/index.htm	Custom Reports>Funding>Best Practices and Funding>Expenditures	
c_s_n	cigarette consumption per smoker	https://www.cdc.gov/statesystem/index.html	https://www.cdc.gov/tobacco/about/osh/index.htm	Download file using the following path: Custom Reports>Policy>The Tax Burden on Tobacco>Cigarette Sales OW	Transpose

p_s_n	nominal cigarette price	https://www.cdc.gov/statesystem/index.html	https://www.cdc.gov/tobacco/about/osh/index.htm	Custom Reports>Policy>The Tax Burden on Tobacco>Cigarette Sales OW	Transpose
s_s	state current smoking prevalence	https://www.cdc.gov/statesystem/index.html	https://www.cdc.gov/tobacco/about/osh/index.htm	Custom Reports>Tobacco Use – Survey Data>Cigarette Use (Adults)>Current Smoking - BRFSS	Transpose
y_s_n	nominal per capita state personal income	https://www.bea.gov/regional/downloadzip.cfm	https://www.bea.gov/regional/index.htm	Download the Annual Personal Income By State File	Transpose
of_s_n om_s_n w_s_n a_s_n	age and sex demographic variables	1980-1990 https://www2.census.gov/programs-surveys/popest/datasets/1980-1990/state/asrh/st_int_asrh.txt 1990-1999 https://www2.census.gov/programs-surveys/popest/tables/1990-2000/intercensal/st-co/stch-icen1990.txt to https://www2.census.gov/programs-surveys/popest/tables/1990-2000/intercensal/st-co/stch-icen1999.txt 2000-2010 https://www2.census.gov/programs-surveys/popest/datasets/2000-2010/intercensal/state/st-est00int-agesex.csv 2010-2018 https://www2.census.gov/programs-surveys/popest/tables/2010-2018/state/asrh/sc-est2018-alldata5.csv	1980-1990 https://www.census.gov/content/census/en/data/datasets/time-series/demo/popest/1980s-state.html 1990-1999 https://www.census.gov/content/census/en/data/datasets/time-series/demo/popest/intercensal-1990-2000-state-and-county-characteristics.html 2000-2010 https://www.census.gov/content/census/en/data/datasets/time-series/demo/popest/intercensal-2000-2010-state.html 2010-2018 https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-detail.html	Download State level inter- and post-censal annual state population estimates by age and sex	
hs_s_n	Hispanic popualtion	Annual BRFSS data files 1985-2018	https://www.cdc.gov/brfss/annual_data/annual_data.htm		

aicpu deflators	All item CPI-U index	https://download.bls.gov/pub/time.series/cu/cu.data.1.AllItems	Price index files, Consumer Price Index for All Urban Consumers (CPI-U) Dimensions, national and Census Region cross-section, Annual averages, Base year = 100, 1982-1984 3.1.a. all-items, time span 1967-2018		
mccpu deflator	Medical care CPI-U index	https://download.bls.gov/pub/time.series/cu/cu.data.15.USMedical	3.1.b.i. national average: CUUR0000SA0L5 3.1.b ii. Northeast: CUUR0100SA0L5 3.1.b.iii. Midwest: CUUR0200SA0L5 3.1.b.iv. South: CUUR0300SA0L5 3.1.b.v. West: CUUR0400SA0L5 3.1.c. medical care, time span 1935-2018 national, 1978-2018, Census Region	We looked at code and decipher	

ailmcpu deflators	All item less medicare CPI-U index	https://download.bls.gov/pub/time.series/cu/cu.data.1.AllItems	3.1.a.i. national average: CUUR0000SA0 3.1.a ii. Northeast: CUUR0100SA0 3.1.a.iii. Midwest: CUUR0200SA0 3.1.a.iv. South: CUUR0300SA0 3.1.a.v. West: CUUR0400SA0 3.1.b. all-items less medical care, time span 1979-2018 URL for download for all-items less medical care:		
ailmcspu deflators	Medicare Care Series CPI-U index	https://download.bls.gov/pub/time.series/cu/cu.data.1.AllItems	3.1.c.i. national average: CUUR0000SAM 3.1.c ii. Northeast: CUUR0100SAM 3.1.c.iii. Midwest: CUUR0200SAM 3.1.c.iv. South: CUUR0300SAM 3.1.c.v. West: CUUR0400SAM 3.1.d. Medical care services time span 1984-2018		

CAaicspu deflators	California all-item CPI-U price index	http://www.dof.ca.gov/Forecasting/Economics/Eco_Forecasts_Us_Ca/	Adjustment factors for California price index due to discontinuation of previous BLS sub-state price indices		
CMS nhce (hr_s_n)	total nominal CMS health care expenditure by state	https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsStateHealthAccountsResidence.html	Health expenditures by state of residence, 1991-2014 [ZIP, 347KB]	Within zip file, it is the file name US_PER_CAPITA14.CVS	Transpose
CMS nhcep (hp_s_n)	CMS total agg pesonal healthcare expenditure by provider	https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsStateHealthAccountsResidence.html	-		Transpose
BEA Population	BEA Population	https://www.bea.gov/regional/downloadzip.cfm	https://www.bea.gov/regional/index.htm		

2. Transpose Data

Feature Name	Description	Transpose
NIPA nhce	total nominal NIPA health care expenditure by state	transpose
c_s_n	cigarette consumption per smoker	transpose
s_s	state current smoking prevalence	transpose
y_s_n	nominal per capita state personal income	transpose
CMS nhce (hr_s_n)	total nominal CMS health care expenditure by state	transpose
CMS nhcep (hp_s_n)	CMS total agg personal healthcare expenditure by provider	transpose

3. Create Deflators (2010 $)

Feature Name	Description	Calculation
mccpu deflator	Medical care CPI-U index	Divide 2010 CPI_U Index by the respective year for each year to calculator ratio which is the deflator
aicpu deflators	All item CPI-U index	Divide 2010 CPI_U Index by the respective year for each year to calculator ratio which is the deflator
ailmcpu deflators	All item less medicare CPI-U index	Divide 2010 CPI_U Index by the respective year for each year to calculator ratio which is the deflator
ailmcspu deflators	Medicare Care Serices CPI-U index	Divide 2010 CPI_U Index by the respective year for each year to calculator ratio which is the deflator

4. Gather additional files need for final file

Cross Walk & Control File

state_region_cross walk.csv

Instrumental Variable File

irrelevant_IVs_panel data_85_18.csv

% of Population 18 years File

census_demograph y_variables_a18_adj.

5. Transform Data for Final File

Final Variable (Final File)				
Year	**Year**	Data Type	**File or Calculation**	**Individual Downloaded Files**
State_Name	State Name	Text	File	Cross Walk File
sfips	State FIPS, two characters	Text	File	Cross Walk File
Cen_Reg_Name	Census Region Name	Text	File	Cross Walk File
BEA_Reg	BEA Region Name	Text	File	Cross Walk File
hr_s	Total real CMS health care expenditure by state	Numeric	Calculation	personal_med_exp * medicare care deflator (2010 $) * (10^-3)
s_s	State current smoking prevalence	Numeric	As is	s_s
y_s	Real per capita state personal income	Numeric	Calculation	y_s = y_s_n * all items less medicare care deflator (2010 $) * (10^-4)
cps_s	Consumption per smoker by state	Numeric	Calculation	cps_s = c_s / s_s
cpsa_s	Consumption per adult smoker by state	Numeric	Calculation / File	cps_s = c_s / (s_s * % of 18 years olds) % of Population 18 years File
c_s	cigarette consumption per smoker	Numeric	As is	c_s
p_s	Real cigarette price	Numeric	Calculation	p_s = p_s_n * all item 2010 deflator (2010 $)
ec_s	Real annual tobacco educational funding, total	Numeric	Calculation	ec_s = ec_s_n * all item 2010 deflator (2010 $)
hp_s	Real CMS total agg personal healthcare expenditure by provider	Numeric	Calculation	hp_s = ((hp_s_n * 1000000) / bea population) * medicare care deflator (2010 $) * (10^-3)

n11_s	Convert NIPA state total per capita health care and social assistance production, both SIC and NAIC data, from nominal dollars to real per capita health care production base year dollars	Numeric		n11_s = (10^(6 - 3)) * ((hlth_care_n_social_assist_70 - social_assist_73)/sorted_f$bea_pop) * medical care services deflators (2010 $)
n10_s	Convert NIPA state total per capita health care and social assistance production, both SIC and NAIC data, from nominal dollars to real per capita health care production base year dollars	Numeric		n10_s = (csic_health_srvc_65/ bea population) * medical care services deflators (2010 $) * (10^(6-3))
c13	12 State Control States	Numeric	File	Control File
c20	20 State Control States	Numeric	File	Control File
c32	32 State Control States	Numeric	File	Control File
c38	38 State Control States	Numeric	File	Control File
c50	50 State Control States	Numeric	File	Control File
k1	Instrumental Variable 1	Numeric	File	Instrumental File
k2	Instrumental Variable 2	Numeric	File	Instrumental File
k3	Instrumental Variable 3	Numeric	File	Instrumental File
k4	Instrumental Variable 4	Numeric	File	Instrumental File
k5	Instrumental Variable 5	Numeric	File	Instrumental File
k6	Instrumental Variable 6	Numeric	File	Instrumental File
k7	Instrumental Variable 7	Numeric	File	Instrumental File
k8	Instrumental Variable 8	Numeric	File	Instrumental File
k9	Instrumental Variable 9	Numeric	File	Instrumental File
k10	Instrumental Variable 10	Numeric	File	Instrumental File
k11	Instrumental Variable 11	Numeric	File	Instrumental File
k12	Instrumental Variable 12	Numeric	File	Instrumental File
k13	Instrumental Variable 13	Numeric	File	Instrumental File
k14	Instrumental Variable 14	Numeric	File	Instrumental File
k15	Instrumental Variable 15	Numeric	File	Instrumental File
k16	Instrumental Variable 16	Numeric	File	Instrumental File
k17	Instrumental Variable 17	Numeric	File	Instrumental File
k18	Instrumental Variable 18	Numeric	File	Instrumental File
k19	Instrumental Variable 19	Numeric	File	Instrumental File
k20	Instrumental Variable 20	Numeric	File	Instrumental File
k21	Instrumental Variable 21	Numeric	File	Instrumental File

k22	Instrumental Variable 22	Numeric	File	Instrumental File
k23	Instrumental Variable 23	Numeric	File	Instrumental File
k24	Instrumental Variable 24	Numeric	File	Instrumental File
k25	Instrumental Variable 25	Numeric	File	Instrumental File
k26	Instrumental Variable 26	Numeric	File	Instrumental File
k27	Instrumental Variable 27	Numeric	File	Instrumental File
k28	Instrumental Variable 28	Numeric	File	Instrumental File
k29	Instrumental Variable 29	Numeric	File	Instrumental File
k30	Instrumental Variable 30	Numeric	File	Instrumental File
k31	Instrumental Variable 31	Numeric	File	Instrumental File
k32	Instrumental Variable 32	Numeric	File	Instrumental File
k33	Instrumental Variable 33	Numeric	File	Instrumental File

Additional Final File Info
The final dataset for the data processing detailed within this document also includes the following requirements:

- Excel Format
- Stata 'long' panel Data Format
- Blank or Null if no data is not present
- Keep zero value zero, and make them blank or NULL
- Include year 1984 – 2018 where possible

TECHNICAL APPENDIX

Data

All data used in ECTCP were still available, except for sub-state price indices used to calculate the California state price index; alternative consistent price index data were obtained from the California Department of Finance was used instead.

The adjusted Census Region and California price indices were estimated as follows. The price indices used were CPI-U deflators for EC: all-item CPIU, Regional; y: all-item CPIU less Medical Care; NIPA health care expenditure Medical Care Services; CMS health care expenditure: Medical Care.

The regional price indices used the East, South and Midwestern Regions. The California price index was taken from the California Department of Finance (CDOF). The CDOF price indices did not have categories of all-items less Medical Care, so we used the all-item price index; it also did not have Medical Care Services, so we used Medical Care. An adjusted West Region index which was derived from the population weighted West Region index after subtracting out the population weighted California Index.

Models with measurement adjustment for 2011 and following years.

Published Model

Current adult smoking prevalence in California:

$$(prev_{c,t} - prev_{CA,t}) = \alpha_1(EC_{CA,t-1} - EC_{c,t-1}) + \alpha_2(p_{c,t-1} - p_{CA,t-1}) + \alpha_3(y_{c,t-1} - y_{CA,t-1}) + \alpha_4 D2011 + \alpha_5[D2011] * (prev_{c,t-1} - prev_{CA,t-1}) + \varepsilon_{1a,t} \quad \text{(S1a)}$$

Cigarette Consumption per Smoker:

$$(cpsa_{c,t} - cpsa_{CA,t}) = \beta_0 + \beta_1(EC_{CA,t-1} - EC_{c,t-1}) + \beta_2(p_{c,t-1} - p_{CA,t-1}) + \beta_3(y_{c,t-1} - y_{CA,t-1}) + \beta_4 D2011 + \beta_5[D2011] * (cpsa_{c,t} - cpsa_{CA,t}) + \varepsilon_{2a,t} \quad \text{(S2a)}$$

NIPA Healthcare Expenditures:

$$n_{CA,t} = \gamma_0 + \gamma_1 n_{c,t} + \gamma_2(prev_{c,t} - prev_{CA,t}) + \gamma_3(cpsa_{c,t} - cpsa_{CA,t}) + \gamma_4(y_{c,t-1} - y_{CA,t-1}) + \gamma_5[D2011] * (prev_{c,t-1} - prev_{CA,t-1}) + \gamma_6[D2011] * (cpsa_{c,t} - cpsa_{CA,t}) + \varepsilon_{3a,t} \quad \text{(S3a)}$$

CMS Healthcare Expenditures:

$$h_{CA,t} = \delta_0 + \delta_1 h_{c,t} + \delta_2(prev_{c,t} - prev_{CA,t}) + \delta_3(cpsa_{c,t} - cpsa_{CA,t}) + \delta_4(y_{c,t-1} - y_{CA,t-1}) + \delta_5[D2011] * (prev_{c,t-1} - prev_{CA,t-1}) + \delta_6[D2011] * (cpsa_{c,t} - cpsa_{CA,t}) + \varepsilon_{4a,t} \quad \text{(S4a)}$$

Forecast Model

Current adult smoking prevalence in California:

$$(prev_{c,t} - prev_{CA,t-1}) = \alpha_1(EC_{CA,t-1} - EC_{c,t-1}) + \alpha_2(p_{c,t-1} - p_{CA,t-1}) + \alpha_3(y_{c,t-1} - y_{CA,t-1}) + \alpha_4 D2011 + \alpha_5[D2011] * (prev_{c,t-1} - prev_{CA,t-1}) + \varepsilon_{1b,t} \quad \text{(S1b)}$$

Cigarette Consumption per Smoker:

$$\left(cpsa_{c,t} - cpsa_{CA,t-1}\right) = \beta_0 + \beta_1\left(EC_{CA,t-1} - EC_{c,t-1}\right) + \beta_2\left(p_{c,t-1} - p_{CA,t-1}\right) + \beta_3\left(y_{c,t-1} - y_{CA,t-1}\right) + \beta_4 D2011 + \beta_5[D2011] * \left(cpsa_{c,t} - cpsa_{CA,t}\right) + \varepsilon_{2b,t} \quad \text{(S2b)}$$

NIPA Healthcare Expenditures:

$$n_{CA,t} = \gamma_0 + \gamma_1 n_{c,t-1} + \gamma_2\left(prev_{c,t} - prev_{CA,t}\right) + \gamma_3\left(cpsa_{c,t} - cpsa_{CA,t}\right) + \gamma_4\left(y_{c,t-1} - y_{CA,t-1}\right) + \gamma_5[D2011] * \left(prev_{c,t-1} - prev_{CA,t-1}\right) + \gamma_6[D2011] * \left(cpsa_{c,t} - cpsa_{CA,t}\right) + \varepsilon_{3b,t} \quad \text{(S3b)}$$

CMS Healthcare Expenditures:

$$h_{CA,t} = \delta_0 + \delta_1 h_{c,t-1} + \delta_2\left(prev_{c,t} - prev_{CA,t}\right) + \delta_3\left(cpsa_{c,t} - cpsa_{CA,t}\right) + \delta_4\left(y_{c,t-1} - y_{CA,t-1}\right) + \delta_5[D2011] * \left(prev_{c,t-1} - prev_{CA,t-1}\right) + \delta_6[D2011] * \left(cpsa_{c,t} - cpsa_{CA,t}\right) + \varepsilon_{4b,t} \quad \text{(S4b)}$$

Final Forecast Model:

Current adult smoking prevalence in California:

$$\left(prev_{c,t} - prev_{CA,t-1}\right) = \alpha_1\left(EC_{CA,t-1} - EC_{c,t-1}\right) + \alpha_2\left(p_{c,t-1} - p_{CA,t-1}\right) + \alpha_3\left(y_{c,t-1} - y_{CA,t-1}\right) + \alpha_4 D2011 + \alpha_5[D2011] * \left(prev_{c,t-1} - prev_{CA,t-1}\right) + \epsilon_{1c,t} \quad \text{(S1c)}$$

Cigarette Consumption per Smoker:

$$\left(cps_{c,t} - cps_{CA,t-1}\right) = \beta_0 + \beta_1\left(EC_{CA,t-1} - EC_{c,t-1}\right) + \beta_2\left(p_{c,t-1} - p_{CA,t-1}\right) + \beta_3\left(y_{c,t-1} - y_{CA,t-1}\right) + \beta_4 D2011 + \beta_5[D2011] * \left(cpsa_{c,t} - cpsa_{CA,t}\right) + \epsilon_{2c,t} \quad \text{(S2c)}$$

Healthcare Expenditures, NIPA measure of health care expenditure:

$$n_{CA,t} = \gamma_0 + \gamma_1 n_{c,t-1} + \gamma_2\left(prev_{c,t-1} - prev_{CA,t-1}\right) + \gamma_3\left(cpsa_{c,t-1} - cpsa_{CA,t-1}\right) + \gamma_4\left(y_{c,t-1} - y_{CA,t-1}\right) + \gamma_5\left(of_{c,t-1} - of_{CA,t-1}\right) + \gamma_5\left(om_{c,t-1} - om_{CA,t-1}\right) + \gamma_6\left(w_{c,t-1} - w_{CA,t-1}\right) + \gamma_7[D2011] * \left(prev_{c,t-1} - prev_{CA,t-1}\right) + \gamma_8[D2011] * \left(cpsa_{c,t} - cpsa_{CA,t}\right) + \epsilon_{3c,t} \quad \text{(S3c)}$$

Healthcare Expenditures, CMS measure of health care expenditure:

$$h_{CA,t} = \delta_0 + \delta_1 n_{c,t-1} + \delta_2\left(prev_{c,t-1} - prev_{CA,t-1}\right) + \delta_3\left(cpsa_{c,t} - cpsa_{CA,t}\right) + \delta_4\left(y_{c,t-1} - y_{CA,t-1}\right) + \delta_5\left(a_{c,t-1} - a_{CA,t-1}\right) + \delta_5\left(hs_{c,t-1} - hs_{CA,t-1}\right) + \delta_6[D2011] * \left(prev_{c,t-1} - prev_{CA,t-1}\right) + \delta_7[D2011] * \left(cpsa_{c,t} - cpsa_{CA,t}\right) + \varepsilon_{4c,t} \quad \text{(S4c)}$$

Where $D2011$ is a dummy variable for the BRFSS change in survey methodology, = 0 until year 2010, = 1 for 2011 and following years.

Table S1. Estimated California smoking prevalence, cigarettes per capita, and per capita healthcare expenditures.

Equation	Dependent Variable	Statistic	ECTPC, 2013	ECTCP, 'published', updated data to 2008	ECTCP, 'published', updated data	ECTCP, 'forecast', updated data	ECTCP, final 'forecast', updated data	Dimension
Model Equations								
1	$(prev_{c,t} - prev_{CA,t})$ $(prev_{c,t} - prev_{CA,t-1})$**	$\alpha 0$	6.30 (0.610)	6.05 (1.09)	5.95 (1.09)	6.29 (1.17)	6.29 (1.17)	
		$\alpha 1$	0.0497 (0.00347)	0.0489 (0.0101)	0.0494 (0.00889)	0.0503 (0.00970)	0.0503 (0.00970)	/$ per capita
		$\alpha 2$	-1.00 (0.477)	-0.861 (0.634)	-0.940 (0.517)	-0.419 (0.573)	-0.419 (0.573)	/$ per pack
		$\alpha 3$	0.416 (0.0730)	0.432 (0.160)	0.352 (0.140)	0.312 (0.147)	0.312 (0.147)	/$1000 per capita
		R2 (%)	77	81	76	95	95	
		r1	0.154	0.121	0.14	-0.0230	-0.0230	
2	$(cps_{c,t} - cps_{CA,t})$ $(cps_{c,t} - cps_{CA,t-1})$**	$\beta 0$	67.9 (10.2)	92.8 (33.6)	71.3 (30.2)	53.6 (32.0)	53.6 (32.0)	
		$\beta 1$	1.39 (0.132)	2.11 (0.271)	1.93 (0.250)	2.22 (0.267)	2.22 (0.267)	/$ per capita
		$\beta 2$	-26.6 (6.80)	-24.2 (17.3)	-40.5 (14.4)	-42.3 (15.9)	-42.3 (15.9)	/$ per pack
		$\beta 3$	2.97 (1.21)	2.17 (4.30)	-0.328 (3.89)	2.82 (4.05)	2.82 (4.05)	/$1000 per capita
		R2 (%)	81	95	96	95	95	
		r1	0.148	0.132	0.19	-0.0788	-0.0788	
3	$n_{CA,\ t}$	$\gamma 0$	-550 (433)	-530 (407)	-433 (297)	-350 (397)	1680 (464)	$
		$\gamma 1$	1.15 (0.180)	0.992 (0.165)	1.06 (0.114)	1.02 (0.142)	0.668 (0.153)	
		$\gamma 2$	-35.4 (9.85)	-63.7 (16.5)	-69.5 (13.8)	-54.5 (16.7)	-53.2 (23.5)	$/%point
		$\gamma 3$	-3.14 (0.786)	-3.16 (0.401)	-3.22 (0.335)	-3.40 (0.433)	-2.95 (0.602)	$ pack per smoker
		$\gamma 4$	-108 (6.79)	-70.1 (15.8)	-62.3 (12.1)	-88.6 (15.9)	-40.9 (17.3)	$/$1000 per capita
		$\gamma 5$	--	--	--	--	1361 (1351)	$/%point
		$\gamma 6$	--	--	--	--	-2010 (4925)	$/%point
		$\gamma 7$	--	--	--	--	3772 (3075)	$/%point
		R2 (%)	80	90	92	86	92	
		r1	0.262	0.380	0.410*	0.433*	-0.27	

Equation	Dependent Variable	Statistic	ECTPC, 2013	ECTCP, 'published', updated data to 2008	ECTCP, 'published', updated data	ECTCP, 'forecast', updated data	ECTCP, final 'forecast', updated data	Dimension
4	$h_{CA,\ t}$	$\delta 0$	1056 (112)	1476 (372)	1627 (303)	1477 (279)	1391 (447)	$
		$\delta 1$	0.847 (0.0542)	0.977 (0985)	0.908 (0.0590)	0.845 (0.0492)	1.11 (0.107)	
		$\delta 2$	-67.8 (7.31)	-144 (31.2)	-130 (25.3)	-86.3 (21.5)	-87.6 (21.4)	$/%point
		$\delta 3$	-5.48 (0.928)	-5.69 (0.848)	-5.24 (0.641)	-5.00 (0.573)	-4.26 (0.650)	$ pack per smoker
		$\delta 4$	-107 (22.3)	-42.7 (30.6)	-59.7 (23.0)	-81.6 (74.7)	-61.33 (25.0)	$/$1000 per capita
		$\delta 5$	--	--	--	--	2657 (1093)	$/%point
		$\delta 6$	--	--	--	--	-410 (141)	$/%point
		$R2$ (%)	89	87	93	94	94	
		$r1$	0.486*	0.514*	0.499*	0.460*	-0.0367	

*significant at the 5% level.

** specification of dependent variable for the 'forecast' model.

r_1: first order autocorrelation coefficient.

$prev_{j,\ t}$: Prevalence of current smoking in population j, for California and control states in year t,(percentage points).

$cps_{j,\ t}$: Cigarettes consumption per current smoker in population j, for California and control states in year t, (packs/year per smoker).

$EC_{j,\ t}$: Cumulative per capita funding in population j, for California and control states in year t, (dollars).

$p_{j,\ t}$: Price per pack of cigarettes in population j, for California and control states in year t, (dollars).

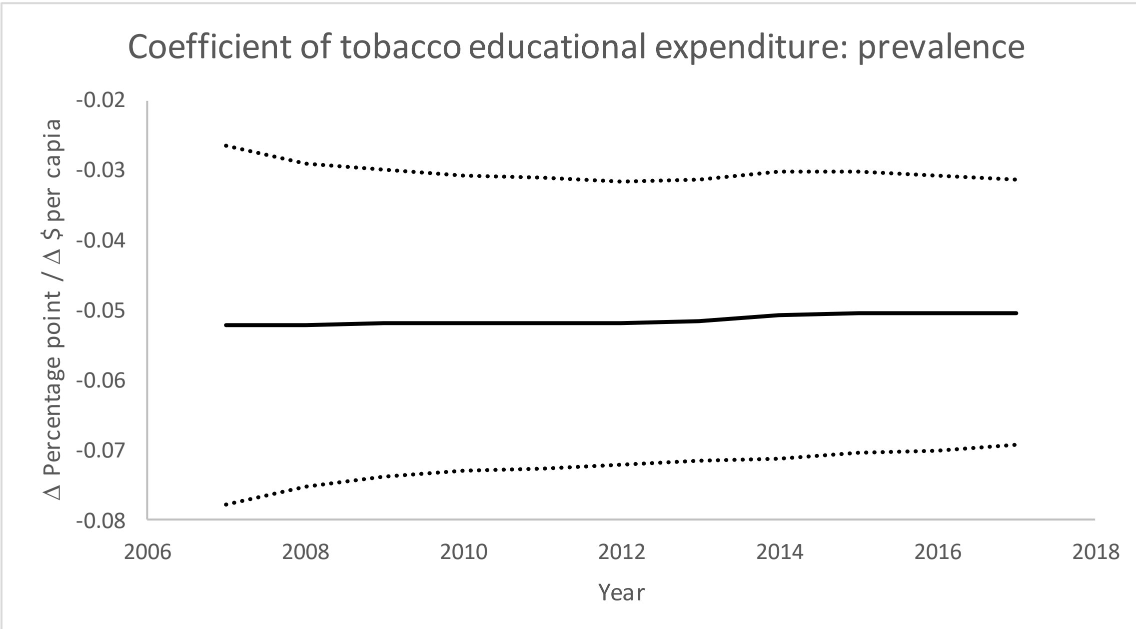
$y_{j,\ t}$: Per capita personal income in population j, for California and control states in year t, (thousands of dollars).

$n_{j,\ t}$: Per capita healthcare expenditures in population j, for California and control states in year t, (thousands of dollars).

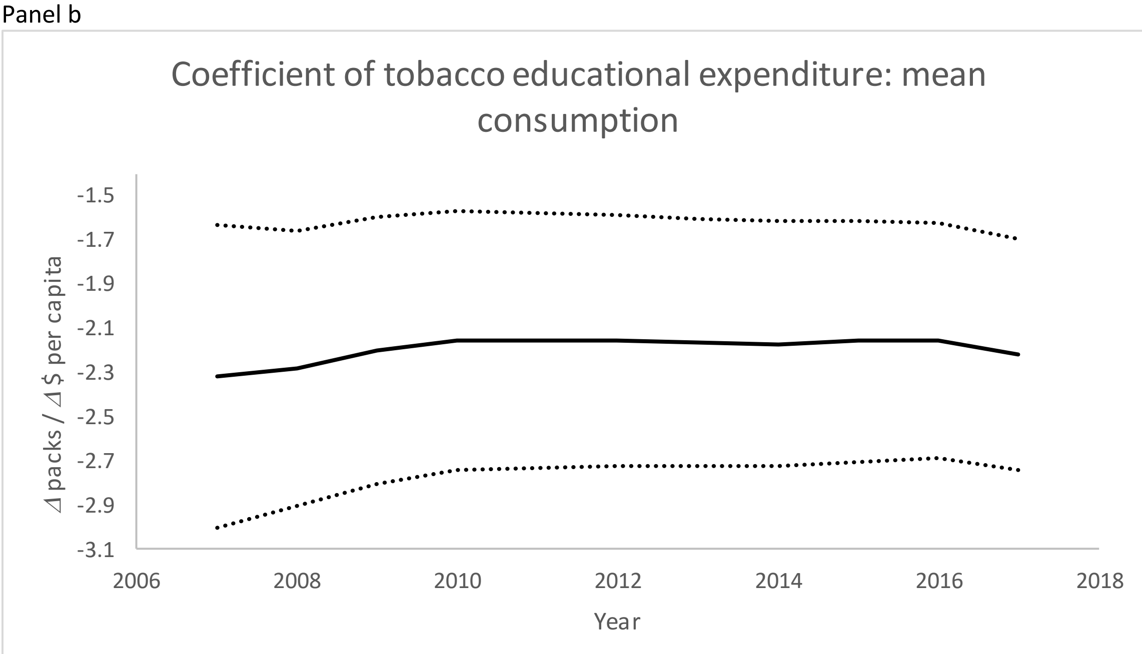
$h_{j,\ t}$: Per capita healthcare expenditures in population j, for California and control states in year t, (thousands of dollars).

Note: dollar amounts are in 2010 dollars.

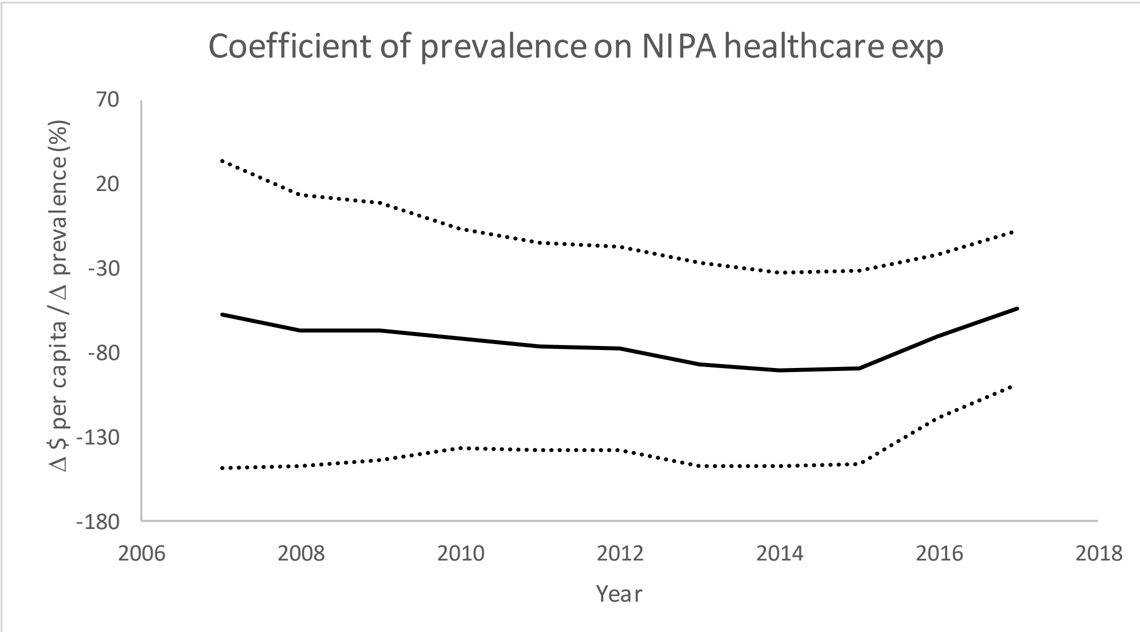
Panel a



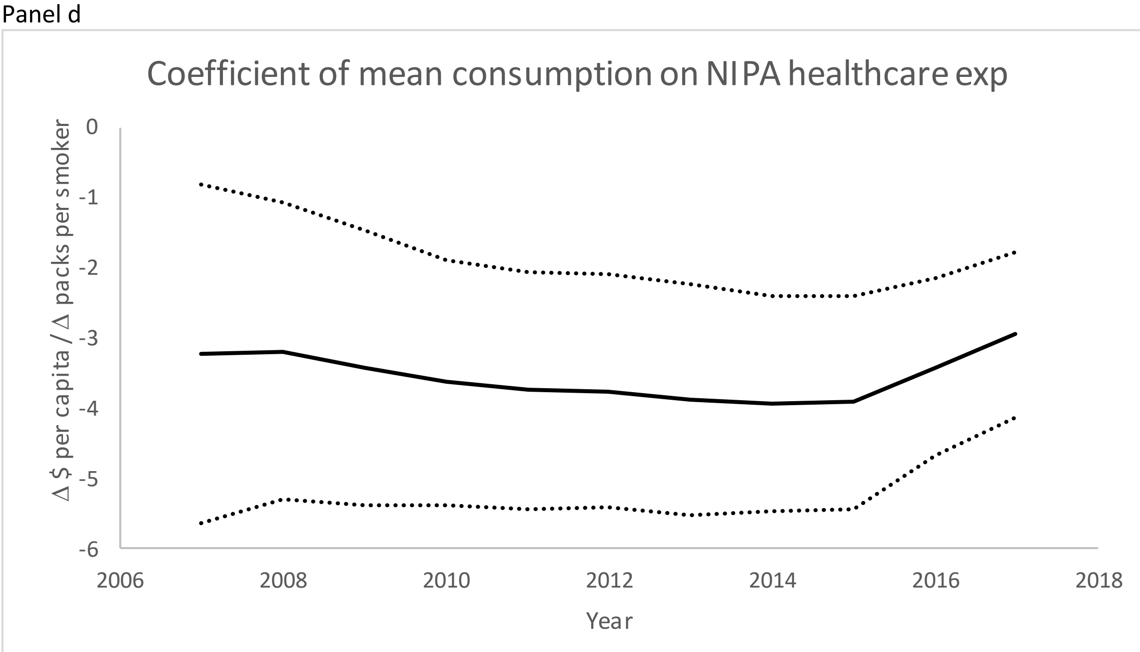
Panel b



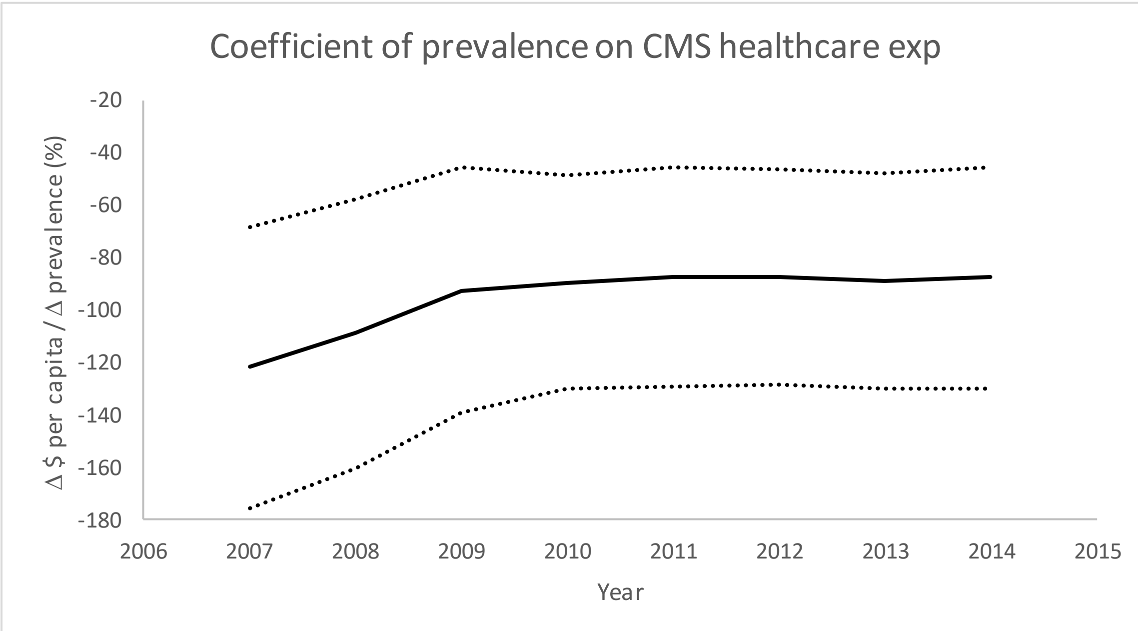
Panel c



Panel d



Panel e



Panel f

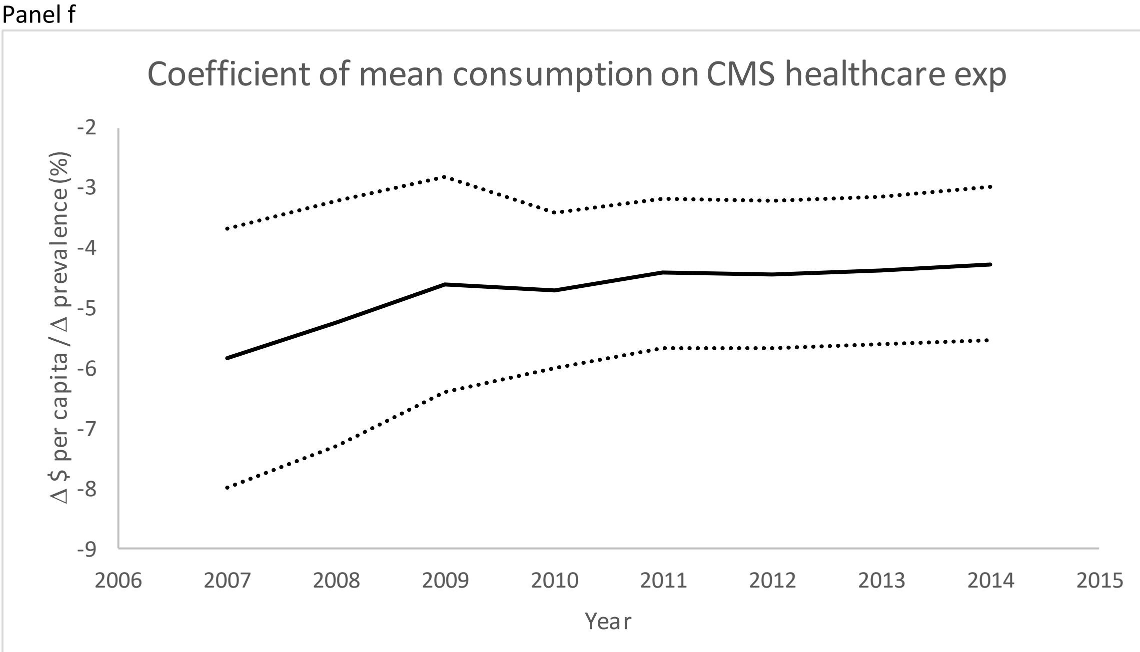


Figure S1.- Recursive coefficient estimates for variables used to estimate CTCP program effect.

REFERENCES

Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program. J Am Stat Assoc 105: 493–505

Lightwood, J., et al. (2020). "Predictive validation and forecasts of short-term changes in healthcare expenditure associated with changes in smoking behavior in the United States." PLoS One **15**(1): e0227493.

Lightwood, J. and S. Glantz (2013). "The Effect of the California Tobacco Control Program on Smoking Prevalence, Cigarette Consumption, and Healthcare Costs: 1985-2008." PLoS One **8**(2): e47145.

Lightwood, J. and S. A. Glantz (2016). "Smoking Behavior and Healthcare Expenditure in the United States, 1992-2009: Panel Data Estimates." PLoS Med **13**(5): e1002020.