Asthma Cost Calculator

Technical Documentation

Version 2

October 2024

Centers for Disease Control and Prevention (CDC)
Developed by Metas Solutions

Contents

Source Data Sets

The Asthma Cost Calculator generates estimates of asthma prevalence and related costs using data from the following key sources:

- Medical Expenditure Panel Survey (MEPS): Data from the years 2016 through 2021, covering medical expenditures, insurance coverage, and demographic characteristics.
- Behavioral Risk Factor Surveillance System (BRFSS): Data for adults (18 years of age or older) from 2018 through 2022 covering asthma status, demographic, and socio-economic characteristics.
- National Survey of Children's Health (NSCH): Data for children (17 years of age or younger) from 2021 and 2022 covering asthma status, demographic, and socio-economic characteristics.
- U.S. Census Bureau: The 2022 state-level population estimates by sex, age, race, and ethnic origin and national-level population projections for 2022 to 2100 by sex, age, race, and ethnic origin are used to develop state-level projections for asthma prevalence and costs.
- Centers for Medicare and Medicaid Services Wage Index by State: Data from the 2022 proposed wage index table by CBSA (Table 3).
- Bureau of Economic Analysis' Price Indexes for Personal Consumption Expenditures by Function: Data for Health Expenditures from line 37 (Health) from 2016 to 2021.

Methodology Overview

The Asthma Cost Calculator uses a two-part machine learning (ML) predictive model, called "random forest," to estimate medical costs. It also uses a count data regression model for absenteeism attributable to asthma (International Classification of Diseases, Tenth Revision (ICD-10), code J45). Specifically, a negative binomial regression method was chosen over the conventional poisson regression approach based on the Likelihood Ratio test for goodness of fit.[4] The two-part ML model is analogous to the parametric two-part Generalized Linear Model (GLM) method: the first part models the binary outcome of having any healthcare expenditure vs. having zero healthcare expenditure, whereas the

second part models the amount of healthcare expenditure given that the person has non-zero expenditure. The flexible nature of ML methods mitigates one of the key challenges encountered by commonly used parametric models: assuming a specific functional form, a violation of which can lead to bias in parameter estimation. In addition to capturing various nonlinearities and interactions, the ML method also has desirable predictive properties, which in turn contribute to more accurate assessment of direct medical costs. The projections include estimates for all payers, Medicaid, Medicare, and private insurance categories. All monetary values in the data are adjusted for inflation using the health price index for personal consumption expenditures, with 2021 as the base year, and population counts are based on U.S. Census Bureau current estimates and future projections.

The total state-specific costs are calculated as follows. First, we calculate per-person annual incremental cost of asthma by population subgroups. Second, we calculate the prevalence of treated asthma for each population subgroup. We obtain the population counts for each population subgroup from U.S. Census data. We multiply the estimated prevalence of treated asthma by the population counts to arrive at the number of people with treated asthma in each population subgroup. By applying the per-person annual incremental cost to each population subgroup counts with treated asthma, we obtain direct medical costs associated with asthma. We use national Medical Expenditure Panel Survey (MEPS) data to estimate the incremental missed workdays per year attributable to asthma. Based on daily wage rates among the survey respondents with treated asthma, we calculate the per-person annual absenteeism cost. Next, we estimate the prevalence of treated asthma among the adult population, which we use to calculate the number of adults with treated asthma. We multiply the number of adults with treated asthma by the per-person annual absenteeism cost to arrive at the annual total cost of absenteeism. We add up the total absenteeism cost with the total direct medical costs to arrive at the combined direct and indirect cost of asthma. We generate two sets of geographic adjustment factors using BRFSS data and restricted MEPS data to adjust our national prevalence estimates and national incremental cost estimates to state-specific values.

Random Forest Specification

The two-part model is a well-established statistical approach used in healthcare expenditure analysis to account for the distinct nature of healthcare cost data.[5,6] Suppose Y is the cost and X is a matrix of relevant predictors. The two-part model can be represented as follows:

$$E[Y \mid X] = Pr(Y > 0|X) \times E[Y|Y>0, X]$$

The random forest model is used for both parts of the model. The first part employs a random forest classifier to predict the likelihood of a patient incurring any healthcare costs, and the second part uses a random forest regressor to estimate the amount of the healthcare expenditure for those with non-zero costs.

The same set of predictors are used in both parts: asthma indicator with two categories (meets the criteria and does not meet the criteria), sex with two categories (Male and Female), age group with six categories (0-5 years, 6-11 years, 12-17 years, 18-44 years, 45-64 years, 65 years and older), income with five categories (Poor- less than 100% of the Federal Poverty Level (FPL), Near poor- 100% to 124% of the FPL, Low income- 125% to 199% of the FPL, Middle income- 200% to 399% of the FPL, High income- 400% or more of the FPL), education with four categories (Less than high school, High school, College, Advanced college), race and ethnic origin with five categories (Hispanic, White non-Hispanic, Black non-Hispanic, Asian non-Hispanic, Other), insurance with four categories (Private, Medicare, Medicaid, Uninsured), married with two categories (Married and Not married), region with four categories (Northeast, Midwest, South, West), and year with six categories (2016, 2017, 2018, 2019, 2020, and 2021).

The output variable in the first state is a binary indicator of whether the patient has any healthcare costs (0 = no costs, 1 = any costs). The amount of healthcare expenditure for those patients who have non-zero costs is used as the output variable in the second part of the model.

Hyperparameters for both parts, such as the number of trees, maximum depth of the trees, and minimum samples per leaf, are optimized using five-fold cross-validation. The accuracy and mean squared error are used as the primary performance metrics.

Treated Population, Per-Person Costs, and Absenteeism

Treated Population

The treated population is defined as individuals who received medical care for asthma in a given year. The MEPS dataset from 2016 through 2021 is used to estimate the treated population, while BRFSS data were used to estimate state-specific variations in prevalences of treated asthma. The estimates are weighted to ensure they are nationally representative.

Per-person Medical Costs

Asthma-related costs are calculated using two-part random forest machine learning models, which predict both the probability of incurring any expenditure and the amount spent. Separate estimates are provided for different payers (e.g., Medicaid, Medicare, private insurance). These costs are adjusted for inflation and reported in 2021 dollars.

Absenteeism Costs

Absenteeism costs are estimated by calculating the number of workdays and school days missed due to asthma. For adult populations, the estimates are derived from MEPS data. Similarly, for children, the estimates are derived from NSCH data. These are then projected to state-level populations using data from the U.S. Census Bureau. We calculated missed workdays and missed school days to estimate absenteeism and indirect costs. We used 2016–2021 data from MEPS to calculate the number of patients with asthma aged ≥18 years and we also used 2021-2022 data from NSCH to calculate the number of patients with asthma aged <18.Indirect medical costs per person by absenteeism were calculated using the number of missed workdays and the number of missed school days and the mean daily wage from the JOBS files. Indirect costs were calculated by multiplying the missed workdays and missed school days by the mean daily wage of patients with asthma. The

mean daily wage of patients with asthma was estimated using an hourly wage and work hours per week of patients with asthma as reported in the MEPS data.

Medical Cost Projections

Projections of asthma-related medical costs through 2035 are based on population projections from the U.S. Census Bureau. The projection assumes that per-person costs grow at the 1.92% rate per year as per Congressional Budget Office projections of real spending growth and that asthma prevalence for each population subgroup remain steady. State-level projections are developed using national-level estimates in three steps. First, using the national projections for each population subgroup, we calculate year-to-year growth rate. We apply these growth rates to the actual 2022 population subgroup estimates in each state to generate future population counts from 2023 to 2035. Then, we apply the prevalence rates for each subgroup to generate the projected count of asthma-treated population. We multiply the projected count of asthma-treated population by the adjusted incremental cost of asthma to arrive at the projected direct medical costs of asthma by state. All cost estimates are presented in constant 2021 dollars.

Glossary of Terms

Absenteeism: Absence from work or school due to asthma.

Treated Population: Individuals who received any medical care for asthma in the previous year.

Per-Person Incremental Cost: The average difference in the amount of health expenditures that a person with treated asthma spends and the amount this very person would have spent if she or he did not asthma.

Negative binomial model: A parametric statistical method used to model the relationship between a count dependent variable that has a non-negative integer value, such as the number of days missed due to illness, and a set of predictors such as demographic, socio-economic, and health characteristics.

References

1. U.S. Census Bureau. Population Estimates, 2022. Available at www.census.gov.

2. Medical Expenditure Panel Survey (MEPS). Data for 2016–2021. Available at www.meps.ahrq.gov.

3. Behavioral Risk Factor Surveillance System (BRFSS). Asthma prevalence data for 2018–2022. Available at www.cdc.gov/brfss.

4. A. Colin Cameron and Pravin K. Trivedi (2013), Regression Analysis of Count Data, 2nd edition, Econometric Society Monograph No.53, Cambridge University Press, 1998

5. Manning, W. G., Basu, A., & Mullahy, J. (2005). Generalized modeling approaches to risk adjustment of skewed outcomes data. Journal of Health Economics, 24(3), 465–488. https://doi.org/10.1016/j.jhealeco.2004.09.011

6. Buntin, M. B., & Zaslavsky, A. M. (2004). Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. *Journal of Health Economics, 23*(3), 525–542. https://doi.org/10.1016/j.jhealeco.2003.10.005

7. U.S. Census Bureau. (2022). National Survey of Children's Health, 2021-2022 datasets. U.S. Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Bureau. Retrieved from https://www.census.gov/programs-surveys/nsch/data/datasets.html

8. Centers for Medicare & Medicaid Services. (2022). FY 2022 proposed wage index by CBSA: Table 3. U.S. Department of Health and Human Services. Retrieved from https://www.cms.gov/medicaremedicare-fee-service-paymentacuteinpatientppswage-index-files/fy-2022-wage-index-home-page

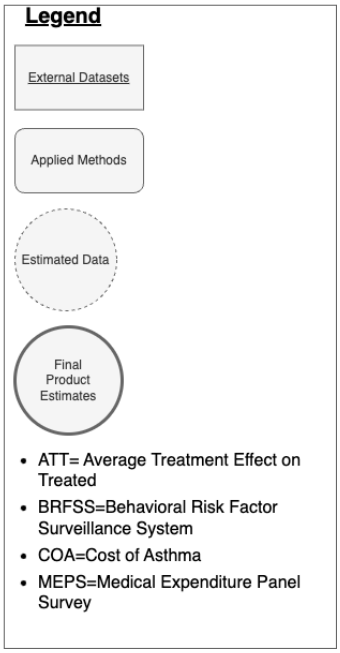
9. U.S. Bureau of Economic Analysis. (2022). Price indexes for personal consumption expenditures by function: Health expenditures (Line 37), 2016-2021. U.S. Department of Commerce. Retrieved from

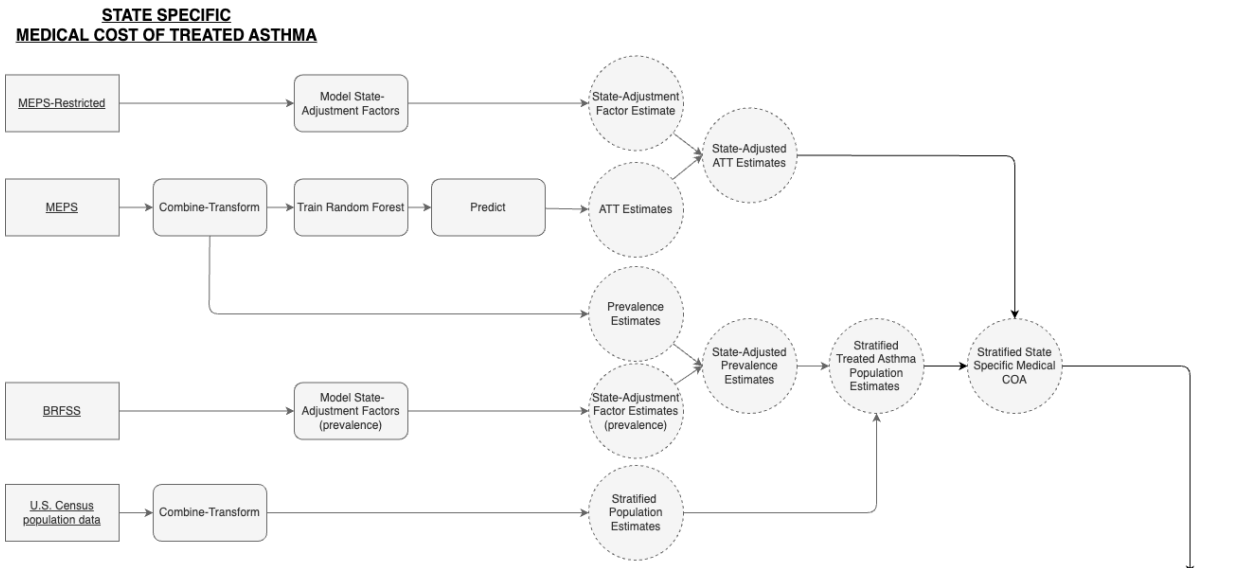
https://apps.bea.gov/iTable/?reqid=19&step=2%23reqid%3D19&step=2&isuri=1&1921=survey#eyJhcHBpZCI6MTksInN0ZXBzIjpbMSwyLDNdLCJkYXRhIjpbWyJDYXRlZ29yaWVzIiwiU3VydmV5Il0sWyJOSVBBX1RhYmxlX0xpc3QiLCI3NCJdXX0=

Appendix

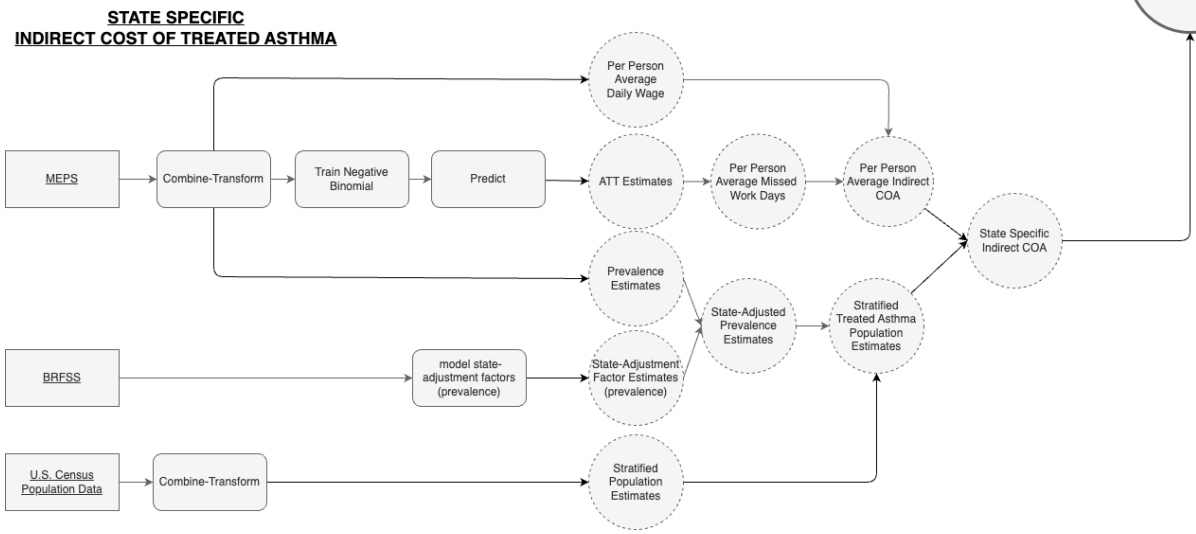
Note: This is supplementary material for the Asthma Cost Calculator Technical Documentation Version 2.

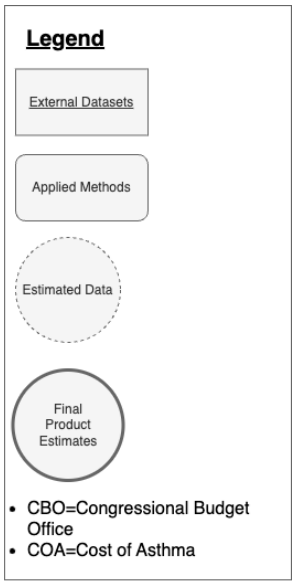
Table of Contents


Legend
External Datasets
Applied Methods
Estimated Data
Final Product Estimates
• ATT= Average Treatment Effect on Treated
• BRFSS=Behavioral Risk Factor Surveillance System
• COA=Cost of Asthma
• MEPS=Medical Expenditure Panel Survey

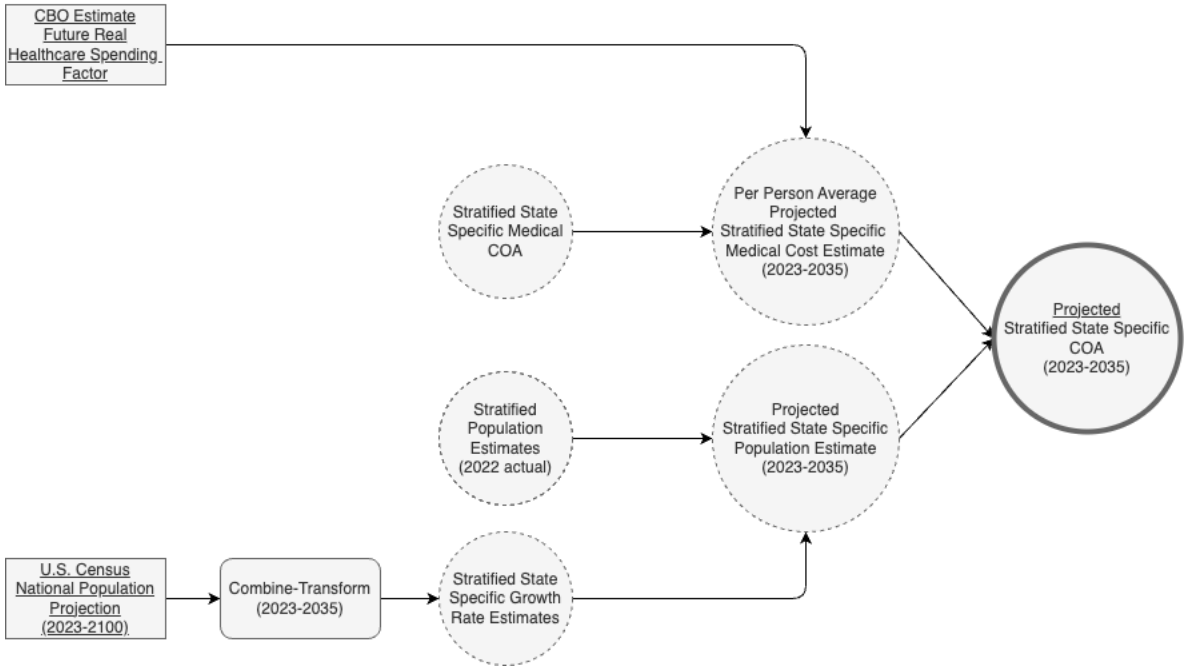

STATE SPECIFIC
MEDICAL COST OF TREATED ASTHMA
MEPS-Restricted
Model State-Adjustment Factors
State-Adjustment Factor Estimate
State-Adjusted ATT Estimates
MEPS
Combine-Transform
Train Random Forest
Predict
ATT Estimates
Prevalence Estimates
State-Adjusted Prevalence Estimates
Stratified Treated Asthma Population Estimates
Stratified State Specific Medical COA
BRFSS
Model State-Adjustment Factors (prevalence)
State-Adjustment Factor Estimates (prevalence)
U.S. Census population data
Combine-Transform
Stratified Population Estimates

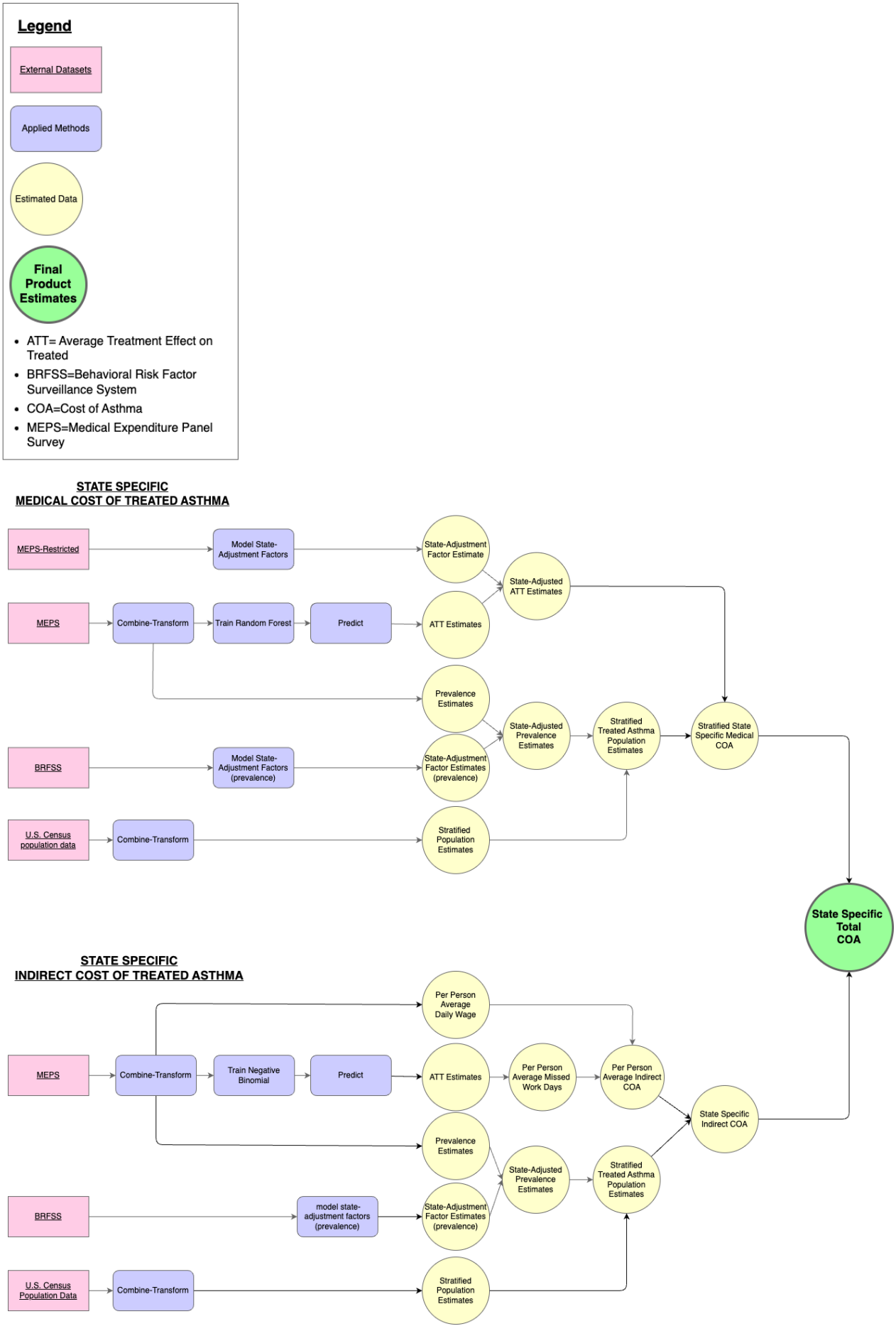

State Specific Total COA

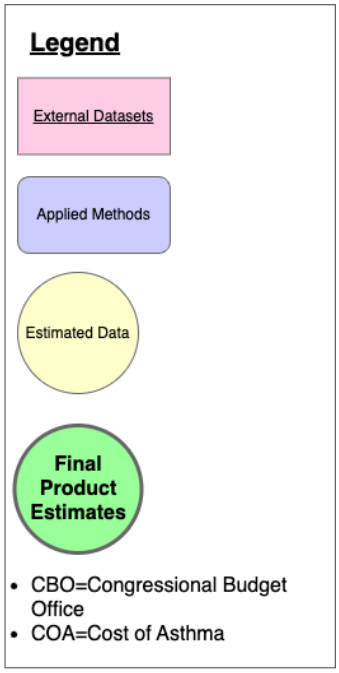

STATE SPECIFIC
INDIRECT COST OF TREATED ASTHMA
Per Person Average Daily Wage
MEPS
Combine-Transform
Train Negative Binomial
Predict
ATT Estimates
Per Person Average Missed Work Days
Per Person Average Indirect COA
State Specific Indirect COA
Prevalence Estimates
State-Adjusted Prevalence Estimates
Stratified Treated Asthma Population Estimates
BRFSS
model state-adjustment factors (prevalence)
State-Adjustment Factor Estimates (prevalence)
U.S. Census Population Data
Combine-Transform
Stratified Population Estimates



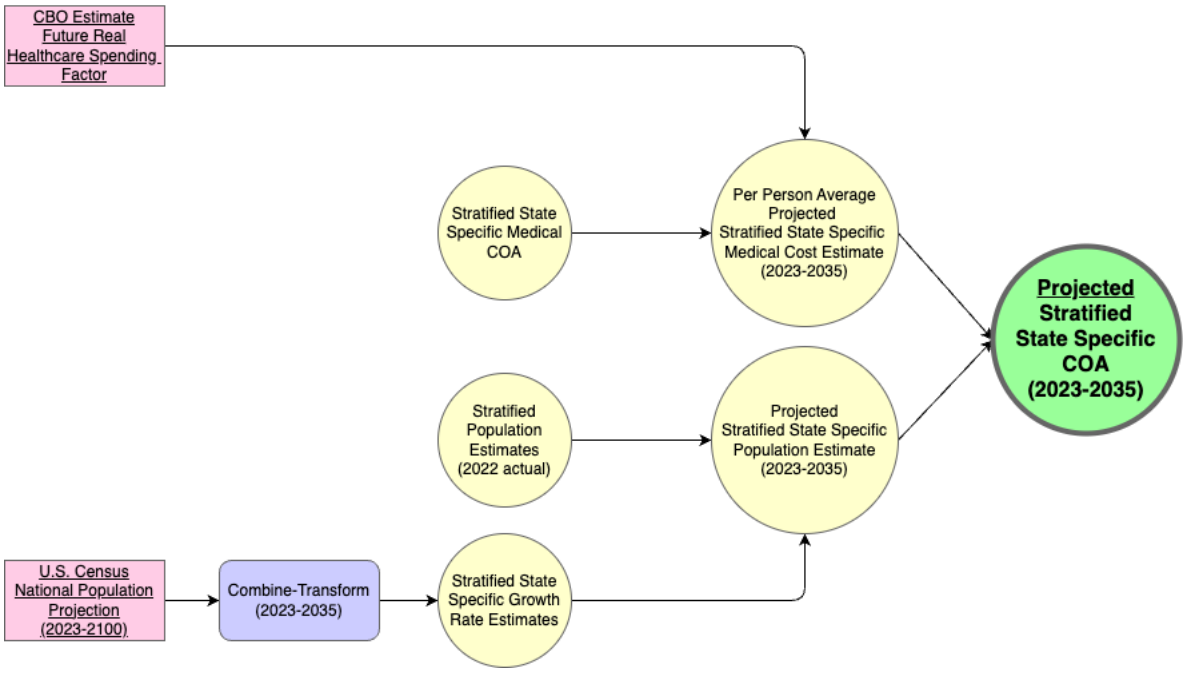
PROJECTED
STATE SPECIFIC COST OF TREATED ASTHMA




Legend
External Datasets
Applied Methods
Estimated Data
Final Product Estimates
ATT= Average Treatment Effect on Treated
BRFSS=Behavioral Risk Factor Surveillance System
COA=Cost of Asthma
MEPS=Medical Expenditure Panel Survey
STATE SPECIFIC MEDICAL COST OF TREATED ASTHMA
MEPS-Restricted
Model State-Adjustment Factors
State-Adjustment Factor Estimate
State-Adjusted ATT Estimates
MEPS
Combine-Transform
Train Random Forest
Predict
ATT Estimates
Prevalence Estimates
State-Adjusted Prevalence Estimates
Stratified Treated Asthma Population Estimates
Stratified State Specific Medical COA
BRFSS
Model State-Adjustment Factors (prevalence)
State-Adjustment Factor Estimates (prevalence)
U.S. Census population data
Combine-Transform
Stratified Population Estimates
State Specific Total COA
STATE SPECIFIC INDIRECT COST OF TREATED ASTHMA
Per Person Average Daily Wage
MEPS
Combine-Transform
Train Negative Binomial
Predict
ATT Estimates
Per Person Average Missed Work Days
Per Person Average Indirect COA
State Specific Indirect COA
Prevalence Estimates
State-Adjusted Prevalence Estimates
Stratified Treated Asthma Population Estimates
BRFSS
model state-adjustment factors (prevalence)
State-Adjustment Factor Estimates (prevalence)
U.S. Census Population Data
Combine-Transform
Stratified Population Estimates



PROJECTED
STATE SPECIFIC COST OF TREATED ASTHMA



The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention.