ASTHMA BURDEN AND COST ANALYSIS

Technical Report
Date: June 5, 2025
Prepared by: Metas Solutions for NCEH, CDC

Contents

1. Introduction

1.1 Background and Purpose

Asthma imposes a significant disease burden on patients, caregivers, employers, and health systems (American Lung Association 2024, Gruffydd-Jones et al. 2019, Ekim and Ocakci 2016), citation here). Characterized by reversible airway obstruction and hyper-responsiveness, asthma affects all age groups—from young children to older adults—and contributes to elevated healthcare utilization (e.g., emergency department visits, hospitalizations) and productivity losses (e.g., missed workdays, missed workdays to care for sick children). This report aims to offer a streamlined, yet comprehensive model for estimating the **direct (medical) and indirect (absenteeism) costs** of asthma across U.S. states.

1.2 Overview of Key Objectives

1. **Incremental Medical Cost per Person:** Develop and validate a robust predictive model (two-part random forest) to estimate the additional per-person cost of asthma relative to otherwise similar individuals without asthma.
2. **Incremental Absenteeism Cost per Person:** Quantify missed workdays for adults and caregivers of children with asthma, assigning an economic value to these lost days.
3. **State-Specific Adjustments:** Incorporate restricted MEPS data, random effects, or other geostatistical refinements to improve local accuracy.

4. **Prevalence and Population Counts:** Derive up-to-date, subgroup-specific asthma prevalence estimates from BRFSS (adults) and NSCH (children) to yield the total number of people with asthma at each state or region.

5. **Aggregate Cost and Projections:** Combine incremental costs with present census population counts as well as their future projections to form total state-level estimates. Provide future cost projections through 2035.

Each section details the relevant data and analytical steps.

2. Data Sources

2.1 Overview and Rationale

We use multiple nationally representative datasets to capture medical expenditures, prevalence rates, and demographic distributions:

1. **Medical Expenditure Panel Survey (MEPS)**
2. **Behavioral Risk Factor Surveillance System (BRFSS)**
3. **National Survey of Children's Health (NSCH)**
4. **U.S. Census Bureau** Population Estimates/Projections

Each dataset was selected to fulfill a specific analytical need—MEPS for expenditures and missed workdays, BRFSS for adult prevalence, NSCH for child prevalence, and Census for population denominators.

2.2 MEPS Data for Cost and Absenteeism

- **Years:** 2016 through 2022, pooled to increase sample size.
- **Components:** Household component data; restricted-use files for state-level analyses if available.
- **Key Variables:** Total annual expenditures (sum of inpatient, outpatient, emergency, prescription, etc.), days of work missed due to illness, demographic and socioeconomic status.
- **Inflation Adjustment:** Dollar values converted to 2022 using the personal health care expenditure index or personal consumption expenditures (PCE) recommended by the Bureau of Economic Analysis.

2.3 BRFSS Data for Adult Asthma Prevalence

- **Years:** 2018–2022.
- **Population:** Non-institutionalized adults (≥18 years).
- **Asthma Definition:** Self-report of *“Have you ever been told you have asthma?”* followed by *“Do you currently have asthma?”*
- **Weighting and Reliability:** We align with the CDC methodology for age-adjusted prevalence, controlling for small cell counts and high relative standard errors.

2.4 NSCH Data for Child Asthma Prevalence

- **Years:** 2021–2022.
- **Population:** Children (<18 years) in household settings.
- **Asthma Definition:** Self or parent/guardian report of a healthcare professional’s diagnosis of “current” asthma.
- **Rationale:** Data needed to capture children’s prevalence, given that BRFSS focuses on adults.

2.5 U.S. Census Bureau Data for Population

- **Source:** 2022 population estimates, plus projection series through 2035 for scenario-based forecasting.
- **Stratification:** Age, sex, race/ethnicity.
- **Alignment with BRFSS/NSCH:** We ensure subgroups are matched to the prevalence breakdowns (e.g., 0–17 vs. 18+ for child vs. adult distinctions).

3. Methodological Framework

3.1 Incremental Cost Estimation: General Concepts

An **incremental cost** approach isolates the added cost due to asthma by comparing predicted expenditures of individuals with asthma (treated population) to an equivalent population hypothetically without asthma, controlling for other covariates (Nurmagambetov et al. 2017).

3.2 The Two-Part Model and Adaptations

Under classical two-part modeling (Belotti et al. 2015):

1. Combined Parts: E(Y | X) = P(Y>0 | X) * E(Y | Y>0, X)
2. **Part 1:** Model P(Y>0 | X) via logistic (or probit) regression.
3. **Part 2:** Model E(Y | Y>0, X); Conditional on Y>0, model the continuous cost distribution using a GLM (gamma log-link).

Here, we replace these parametric regressions with **random forests**, which can capture complex interactions.

3.3 Advantages of Random Forest in a Two-Part Context

- **Non-Parametric:** No need to assume distributional forms (gamma, log-normal, etc.).
- **Enhanced Fit for High-Cost Outliers:** More robust splits in the regression trees may handle right-skewed cost data better, especially with large sample sizes.
- **Better Predictive Power for Zero Costs:** The classification forest in Part 1 can adapt to a variety of demographic or clinical indicators for zero spending.

3.4 Variable Selection and Model Inputs

We incorporate:

- **Demographics:** Age, sex, race/ethnicity, marital status.
- **Socioeconomic:** Insurance type (private, Medicare, Medicaid), income (in relation to federal poverty levels), highest education level.
- **Clinical Indicators:** Asthma diagnosis indicator, plus other comorbidities.
- **Calendar Year:** Categorical indicators for 2016–2022 to capture secular changes.

4. Estimating Medical Costs

4.1 The Random Forest Two-Part Model

4.1.1 Part 1: Random Forest Classifier

- **Outcome:** Binary: any (≥$1) vs. zero expenditures.

- **Splitting Criteria:** Gini impurity or entropy to maximize purity of the resulting classification nodes.
- **Hyperparameter Tuning:** We vary the number of trees, maximum depth, minimum samples per split, and minimum samples per leaf. We rely on 5-fold cross-validation to pick the set that yields the highest area under the curve (AUC), a commonly used metric to evaluate the performance of binary classification models.

4.1.2 Part 2: Random Forest Regressor

- **Outcome:** Positive spending among those with >$0 expenditures.
- **Splitting Criteria:** Mean Squared Error (MSE), a commonly used metric to evaluate predictive performance of continuous dependent variable models at each node.
- **Hyperparameter Tuning:** Similar to Part 1, but the performance metric is MSE on out-of-bag samples.

4.2 Data Cleaning and Variable Construction

- **Asthma Indicator (Treated Asthma):** Individuals must report a healthcare event (e.g., doctor visit, prescription) specifically coded for asthma (ICD-10 J45) within the year.
- **Zero-Expenditure Cases:** We treat individuals with no healthcare spending as part of Part 1 modeling.
- **Inflation Adjustment:** Convert all cost variables to 2022 dollars using the recommended price index.

4.3 Handling Comorbidities and Zero Expenditures

We include additional flags for conditions such as diabetes, coronary heart disease, or COPD, which can influence total medical costs. This approach follows previous studies, which recognizes that multiple chronic conditions can interact, altering cost estimates. Since we are focused on the *incremental* cost due to asthma, these comorbidities remain in the model as control covariates, so the final difference in predicted cost isolates the component associated with treated asthma.

5. Out of Pocket Incremental Costs and Event Specific Regional and Uncontrolled Asthma Subgroup Costs

5.1 Out of Pocket Incremental Costs

We estimate the incremental out-of-pocket (OOP) direct medical costs associated with asthma across population subgroups defined by sex, race/ethnicity, age, and insurance coverage type. These estimates reflect the additional financial burden borne by individuals with asthma compared to a comparable population without asthma.

Modeling Strategy

We implement a random forest-based two-part model to account for both the probability of incurring OOP expenditures and the level of spending conditional on having any expenditures:

- Part 1 (Classifier): A random forest classifier models the probability of any OOP asthma-related spending (≥$1).

- Part 2 (Regressor): A random forest regressor models the magnitude of OOP spending among those with >$0 expenditures.

Model Inputs and Covariates

Covariates mirror those used in the general model (see Section 3.4), including:

- Demographic: Age (categorical), sex, race/ethnicity, marital status

- Socioeconomic: Insurance type (private, Medicare, Medicaid, uninsured), income relative to federal poverty level, education

- Clinical: Asthma status, comorbidities (diabetes, COPD, etc.)

- Temporal: Calendar year (2016–2022)

Estimation and Interpretation

We use counterfactual prediction to compute the incremental OOP cost of asthma for each subgroup. Specifically, we estimate mean predicted OOP costs under the observed asthma status and subtract predicted costs under the counterfactual no-asthma status. Estimates are averaged within subgroups and inflated to 2022 USD using the Personal Health Care (PHC) Price Index.

5.3 Event-Specific Regional Costs by Asthma Control Status

Objective and Motivation

We estimate region-specific, event-level incremental medical costs of asthma to understand how costs differ across healthcare settings and geographic areas. This analysis includes estimates for:

- Inpatient hospitalizations
- Outpatient visits
- Physician office visits
- Emergency department visits
- Prescribed medications
- Home health visits

Estimates are stratified by Census region (Northeast, Midwest, South, West) and asthma control status (controlled vs. uncontrolled).

Defining Asthma Control

Asthma control is based on clinical utilization indicators in the past 3 months. An individual is categorized as having uncontrolled asthma if any of the following apply:

- 3 SABA canisters used
- ≥1 asthma attack
- Any ED visit or inpatient stay for asthma

All other individuals with asthma are classified as having controlled asthma.

Two-Part Modeling Framework by Region and Event Type

For each Census region and event type, we apply a random forest-based two-part model to estimate incremental costs associated with asthma:

- Part 1: Random forest classifier for non-zero spending
- Part 2: Random forest regressor for continuous positive costs

Inputs and Control Variables

All models control for a consistent set of covariates:

- Demographics (age, sex, race/ethnicity, marital status)

- Socioeconomic status (insurance type, income, education)
- Comorbidities
- Calendar year indicators

Separate models are estimated by region, and marginal effects are calculated within each asthma control subgroup.

Output Interpretation

We report incremental cost estimates (asthma vs. counterfactual no-asthma scenario) for each event type within each region and asthma control stratum. All estimates are presented in 2022 dollars.

6. Estimating Absenteeism Costs

6.1 Defining Absenteeism in Asthma

Absenteeism is critical to understanding the indirect burden. Asthma exacerbations often require days away from work or school, leading to lost productivity and wages. Among adults, missed workdays frequently rank among the largest indirect costs in chronic respiratory conditions.

6.2 Estimating Costs for Missed Workdays Among Adults

- **Data Source:** MEPS includes questions about workdays missed due to the respondent's health condition(s).
- **Incremental Days Missed:** Compare mean days missed for the treated asthma sample vs. non-asthma sample using Negative Binomial Regression Model.
- **Wage Assignment:** Multiply the incremental missed days by average daily wage from the self-reported hourly wage from MEPS.

6.3 Estimating Costs for Caregiver Absences Among Children

For child asthma, parents/guardians often miss work to care for a child during flare-ups or medical visits. Following previous studies, we:

1. Identify children with documented missed school days.
2. Estimate the fraction of those days requiring a parent or caregiver to miss work.

3. Assign a daily wage for the caregiver. We use state-specific minimum wage rate and multiply it by 8 to obtain a conservative daily wage estimate.

6.4 Aggregating Absenteeism Costs at the State Level

Similar to the medical cost approach, we multiply **per-person absenteeism cost** by the estimated number of adults or children with asthma.

7. State-Specific Adjustments

7.1 Restricted MEPS for Geographically Enhanced Analysis

Public-use MEPS data typically identifies broad census regions. For more granular state-level estimates, we rely on restricted MEPS available through secure data environments. This step parallels the approach used in the previous study, where state-level random effects are estimated and used as the adjustment factors to convert national estimates to state-level estimates.

7.2 Random Intercepts and State-Level Modeling

We fit a mixed-effects random forest or linear mixed model (depending on software constraints) to estimate a **random intercept** for each state:

log(COST_RATIO) = a + b_state(i) + c*asthma_treated_status + ... + e,

where b_state(i) captures average cost differentials at the state level.

7.3 States with Insufficient Sample Sizes

Restricted MEPS dataset identifies 29 states, and remaining states are not identified. The identified states are AL, AZ, CA, CO, CT, FL, GA, IL, IN, KY, LA, MA, MD, MI, MN, MO, NC, NJ, NY, OH, OK, OR, PA, SC, TN, TX, VA, WA, WI. For the states that are not identified in the restricted dataset, we group them by region (Northeast, Midwest, South, West) and apply the region-specific predictions. This ensures that each state receives some form of local calibration but avoids extremely noisy estimates.

8. Prevalence of Asthma

8.1 BRFSS 2018–2022 for Adult Asthma

As an alternative input, in addition to estimating treated prevalence of asthma based on MEPS survey data, we also pool 5 years of BRFSS data (2018–2022) to obtain more stable estimates of current asthma prevalence by state. Key steps:

1. Identify respondents aged ≥18 who self-report current asthma.
2. Calculate weighted prevalence by state, adjusting for the complex survey design.
3. Assess reliability. We suppress or combine categories if the stratum sample is <50 or if the relative standard error is >30%, as recommended by NCHS data presentation standards.

8.2 NSCH 2021–2022 for Child Asthma

For children (<18 years):

- Use NSCH standardized measures for current asthma.
- Apply sampling weights for national representation.

8.3 Subgroup Stratifications and Suppression Criteria

We strive to present stratifications by age group, sex, race/ethnicity, and income. However, if any cell size fails reliability thresholds, we combine categories to ensure stable estimates.

9. Population Estimates and Disease Counts

9.1 Census-Based Population Data

We download or extract state-level population totals from the **U.S. Census Bureau** for 2022, also obtaining projected counts for future years (e.g., 2023–2035). We focus on major demographic categories (age, sex, race/ethnicity) that align with BRFSS and NSCH grouping.

9.2 Combining Census Counts with Survey Prevalences

Number_with_Asthma = $\hat{p}$ × Population, where $\hat{p}$ is the estimated prevalence in for a particular population subgroup in a specific state. Summing across subgroups yields the total number of individuals with treated (or current) asthma in that state.

9.3 Treated vs. Current Asthma Populations

- **Current Asthma (BRFSS/NSCH):** Typical self-reported measure capturing individuals who identify as having active asthma.
- **Treated Asthma (MEPS):** Individuals with actual expenditures for asthma care. For cost calculations, we usually multiply the *treated* cost per person by the *treated* population to avoid overestimation. However, some analyses approximate the treated population as a fraction of those with current asthma.

10. Aggregate Direct and Indirect Cost Calculations

10.1 Per-Person Costs to Aggregate Costs

$\text{Total\_Direct\_Cost}_s\text{=(Incremental\_Cost\_per\_Person}_s\text{)×Population\_with\_Treated\_Asthma}_s\text{)},$

where **Incremental_Cost_per_Person$_s$** is the estimated additional medical cost per person with treated asthma in state *s*, and **Population_with_Treated_Asthma$_s$** is the number of individuals receiving treatment for asthma in that state.

Subgrouping: If we have subgroup (age, sex) incremental costs, we multiply each subgroup's cost per person by the relevant subgroup count, then sum over subgroups.

10.2 Incorporating Payer-Specific Breakdowns (Medicare, Medicaid, Private)

To understand how total costs break down by payer, we use MEPS individual-level indicators that specify both **insurance coverage** and **spending by payer type** (e.g., Medicare, Medicaid, private insurance). This allows for more accurate attribution of costs to each payer. One approach is to fit **separate models for each payer**, using outcomes that reflect spending by that specific payer. Alternatively, users can include **payer type as an interaction term** within a unified model to capture differential effects.

11. Future Projections

11.1 Base-Year (2022) Projections to 2035

We treat 2022 as our base year for cost and prevalence inputs. For each subsequent year, we project:

1. **Population Growth**: Use U.S. Census Bureau projections disaggregated by age, sex, and state.
2. **Medical Cost Growth**: For each year beyond 2022, apply an annual growth factor (e.g., 1.92%) following a Congressional Budget Office or historical real growth estimate.

12. Access to Care

Asthma requires consistent management-including medications, specialist visits, and environmental controls-to prevent exacerbations. However, gaps in access to care, such as lacking a usual source of care, inability to obtain medical care, delaying treatment due to cost, often lead to preventable ER visits, hospitalizations. Regional disparities further compound these challenges. Therefore, understanding these patterns is critical for targeting interventions to improve outcomes.

To quantify the access related to differences between asthma patients and those without asthma, we estimate access-to-care variables available from 2018 to 2022, “HAVEUS42” and DLAYCA42” and “AFRDCA42”, while accounting for complex survey design. Furthermore, we stratify these estimates by four major Census regions.

13. Data Quality and Potential Biases

- **Survey-Based Diagnoses:** BRFSS and NSCH rely on self-reported data, which may underdiagnose or over-diagnose.
- **Sample Size Limitations:** Many states have low sample counts for certain demographic strata. We adopt commonly used standards to suppress or combine unstable estimates.

14. Updates and Maintenance

The analytic component of the ACC exists as a two-part pipeline, where part 1 handles the data ingestion and part 2 handles the analytics and estimation.

Cost estimates can be updated yearly, corresponding to the releases of the various data resources. Updates can be completed through the Databricks User Interface. The part 1 update requires only parameterization of the year range of data ingestion. The part 2 update requires additional parameters to the years.

The only maintenance required is to ensure all data source links are appropriate for data ingestion and ensure that the ingested data schema is consistent.

15. Additional Data Sources

In addition to data sources listed in Section 2.1, ACC program also has access to additional data sources listed below:

- MarketScan Data
- American Community Survey (ACS)
- Current Population Survey (CPS)
- US Bureau of Labor Stats (BLS)
- Healthcare Cost and Utilization Project (HCUP)
- Centers for Medicare and Medicaid Services Data (CMS)

These data can further compliment the asthma cost calculator by providing a focused prevalence and cost estimates to particular subsets of population or more focused look at a particular health care resource utilization type, e.g. Medicare beneficiaries from CMS data or hospitalization charges attributable to asthma from HCUP data. The results of analyses based on these data will be added as supporting documents below the ACC dashboard at an ongoing basis depending on data availability.

16. Conclusion

This technical report outlines a comprehensive framework for estimating asthma's direct and indirect costs at both national and state levels. Leveraging robust data sources (MEPS, BRFSS, NSCH) and advanced modeling techniques (two-part random forest, random effects for state-level refinement), we generate per-person incremental cost estimates that, when combined with updated prevalence and population data, yield state-specific burden metrics.

We recommend that practitioners, policymakers, and researchers use these results in tandem with local clinical data and environmental surveillance. Future expansions could refine the absenteeism model, include intangible quality-of-life measures, or incorporate evolving trends in asthma prevalence due to climate change or emerging therapies.

Overall, this approach offers a robust and adaptable platform for ongoing monitoring of the economic burden of asthma.

References

1. (2024). *Diabetes State Burden Toolkit Technical Report.* RTI International under contract from CDC.
2. (2013). *Chronic Disease Cost Calculator Version 2: Technical Appendix.* Centers for Disease Control and Prevention, developed by RTI International.
3. 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.
4. 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.
5. Medical Expenditure Panel Survey (MEPS). (2016–2022). Agency for Healthcare Research and Quality, Rockville, MD. www.meps.ahrq.gov
6. Behavioral Risk Factor Surveillance System (BRFSS). (2018–2022). Centers for Disease Control and Prevention, Atlanta, GA. www.cdc.gov/brfss
7. National Survey of Children's Health (NSCH). (2021–2022). U.S. Census Bureau and HHS Health Resources and Services Administration.
8. U.S. Census Bureau. (2022). *Population Estimates and Projections.* www.census.gov
9. U.S. Bureau of Economic Analysis. (2022). *Price Indexes for Personal Consumption Expenditures.* www.bea.gov
10. Congressional Budget Office. (2023). *Long-Term Budget Outlook.*
11. Tursynbek Nurmagambetov, Olga Khavjou, Louise Murphy & Diane Orenstein (2017). "State-level medical and absenteeism cost of asthma in the United States." *Journal of Asthma*, 54:4, 357-370, DOI: 0.1080/02770903.2016.1218013
12. Belotti, F., Deb, P., Manning, W.G., Norton, E.C. (2015). "Twopm: Two-Part Models." *Stata J*. 2015, 15, 3–20.

13. American Lung Association. (2024). Asthma Trends and Burden. https://www.lung.org/research/trends-in-lung-disease/asthma-trends-brief/trends-and-burden.
14. Gruffydd-Jones, K., Thomas, M., Roman-Rodríguez, M., Infantino, A., FitzGerald, J. M., Pavord, I., Haddon, J. M., Elsasser, U., & Vogelberg, C. (2019). Asthma impacts on workplace productivity in employed patients who are symptomatic despite background therapy: a multinational survey. Journal of asthma and allergy, 12, 183–194. https://doi.org/10.2147/JAA.S204278
15. Ekim, A., & Ocakci, A. F. (2016). Caregiver burden in pediatric asthma: a systematic review. Health Science Journal, 10(6), 1.