

CHRONIC DISEASE COST CALCULATOR

TECHNICAL APPENDIX

Version 2

November 2013

Centers for Disease Control and Prevention (CDC)



Developed by RTI International^{*}



^{*} RTI International is a trade name of Research Triangle Institute.

Table of Contents

What is New in Version 2	3
Default Source Data Sets	3
<i>Enrollment</i>	3
<i>Treated Population, Per-Person Costs, and Absenteeism</i>	4
Methodology	4
<i>Enrollment</i>	5
<i>Treated Population</i>	5
<i>Per-Person Medical Costs</i>	6
<i>Per-Person Absenteeism Costs</i>	8
<i>Calculated Costs</i>	9
<i>Medical Cost Projections</i>	10
External Review	10
Glossary of Terms	11
References	13
Table 1: Chronic Disease Definitions—Clinical Classification (CC) Code* and 3-digit ICD-9	14

What is New in Version 2

The following features are new for Version 2 of the Cost Calculator:

- Inclusion of asthma, arthritis and depression
- Estimates of the medical costs of the selected chronic diseases to Medicare, private insurance, and all payers combined (including other payers such as TRICARE, out-of-pocket, the uninsured, etc.)
- Estimates of absenteeism costs for the selected chronic diseases at the state level
- Projections of estimates of the medical costs of selected chronic diseases through 2020.

Default Source Data Sets

Enrollment

U.S. Census Bureau

Total state population and breakdowns by sex and age for 2008 and state population projections by sex and age for 2010 through 2020 came from the U.S. Census Bureau.^{1,2}

Kaiser Family Foundation

Medicare beneficiary data came from the Kaiser Family Foundation 2008 Medicare Health and Prescription Drug Plan Tracker.³

Medicaid Statistical Information System (MSIS)

Medicaid enrollment data came from the Medicaid Statistical Information System (MSIS) State Summary Fiscal Year 2008.⁴ MSIS data are used by CMS to produce Medicaid program characteristics and utilization information for the states. The purpose of MSIS is to collect, manage, analyze and disseminate information on eligibles, beneficiaries, utilization and payment for services covered by State Medicaid programs.

Current Population Survey (CPS)

Private insurance enrollment data and breakdowns of enrollment by sex and age by payer (private insurance, Medicaid, and Medicare) came from the Current Population Survey (CPS).⁵ Private insurance data came from the 2008 CPS and Medicaid and Medicare data came from the 2007 through 2009 CPS. The Current Population Survey (CPS) is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The sample is scientifically selected to represent the civilian noninstitutional population. The sample provides estimates for the nation as a whole and serves as part of model-based estimates for individual states and other geographic areas.

Treated Population, Per-Person Costs, and Absenteeism

(Treated population is defined as the number of people receiving care for the disease in the previous year.)

Medical Expenditure Panel Survey (MEPS)

Data were pooled from the 2004 through 2008 Medical Expenditure Panel Survey (MEPS) Consolidated Data Files, a nationally representative survey of the civilian non-institutionalized population that provides data on annual medical expenditures, sources of payment, insurance coverage, and days missed from work due to illness or injury for each participant.⁶ Diseases were defined using ICD-9 codes based on self-reported diseases that were transcribed by professional coders and reported in the MEPS Medical Conditions files for years 2004 through 2008 (see Appendix).

The combined five-year MEPS sample included 153,012 persons of all ages living in the U.S. Estimates for both the treated population and costs have been adjusted to be nationally representative using MEPS sampling weights for years 2004 through 2008. The default data include years prior to the implementation of Medicare Part D, which took effect in 2006. All expenditure data were inflated to 2010 dollars using the gross domestic product general price index as recommended by AHRQ to reflect more current dollar values.⁷

National Nursing Home Survey (NNHS)

Estimates for the institutionalized population, which are not available in other data sources, were derived from the 2004 National Nursing Home Survey (NNHS). The NNHS is a nationally representative sample of United States nursing homes, their services, their staff, and their residents.⁷⁸ The NNHS provides information on nursing homes from two perspectives- - that of the provider of services and that of the recipient of care. For recipients, data were obtained on demographic characteristics, health status, and sources of payment. Diseases were defined using ICD-9 codes based on any diagnosis of the condition, either at admission or time of the survey and primary or secondary diagnosis (see Appendix).

Methodology

This section describes the methods used to generate the default values in the Cost Calculator. We estimated the treated population and per-person medical and absenteeism costs by sex by age (e.g., males age 65+) using complex survey weights in MEPS and NNHS. However, small sample sizes in MEPS preclude reporting of treated population and costs by sex by age subgroups. Where sample sizes permit, we report treated population and costs separately by sex and by age group. We do this by calculating weighted averages of the estimates by sex by age, where the weights are derived from the payer population distribution in each state. Thus, we apply nationally-representative treated population and per-person medical and absenteeism costs by sex by age to states using state-level demographics. This assures that the subgroup and total population percent treated and costs will always be consistent with one another for any payer population the user enters (i.e., the sum of payer disease costs by sex and the sum of payer disease costs by age both equal the total payer disease costs).

Enrollment

Payer enrollment was taken from the following sources: all payers combined from U.S. Census Bureau estimates for 2008, Medicaid from MSIS State Summary Fiscal Year 2008, Medicare from the Kaiser Family Foundation 2008 Medicare Health and Prescription Drug Plan Tracker, and private insurance from the 2008 CPS. Total state population for 2008 was used for all payers combined. Beneficiaries who were only enrolled in capitated Medicaid managed care during the year were included in the Medicaid population. Medicare population included individuals with Medicare Parts A and/or B.

CMS does not report Medicaid or Medicare enrollment by sex by age. For each payer, we estimated the distribution of the population by sex by age using the CPS (2008 for private insurance and 2007 through 2009 for Medicaid and Medicare). We then applied these distributions to each of the state enrollment populations. The resulting payer population distributions by sex by age were used as weights to calculate weighted average percent treated and per-person medical and absenteeism costs by sex, by age, and for the overall payer population in each state.

Treated Population

Treated population is defined as the number of people receiving care for the disease in the previous year.

- The MEPS 2004-2008 Condition files were used to define treated population. We restricted our condition indicators to those for which respondents received care **within the interview year**.
- Treated population in the MEPS data was based on self-reports of receiving care for the disease within the interview year.
 - For example, a person in remission for cancer but who visited a doctor for a follow-up appointment during the interview year would be included in the cancer treated population. A person with a history of hypertension but who did not have any medical treatment (appointments or medication) during the interview year would not be included in the hypertension treated population.
 - *Persons identified as having the disease in the MEPS thus are likely more resource-intensive (higher cost) than those included in alternative prevalence definitions based on a lifetime history of the disease.*
- Treated populations were estimated by payer/sex/age subgroups at the Census region level, the most detailed geographic data available in the public use MEPS. Because sample sizes were often small at the payer/sex/age/region level, we used logistic regression analysis to smooth the percent treated estimates.
 - For each payer and disease, we used logistic regression to predict the percent treated as a function of sex, age, and region. Stepwise regression was used to test for the presence of statistically significant interactions of age, sex, and region for each payer and disease. Significant interactions were retained in the final logistic

regression specification that was used to predict the percent treated by age, sex, and region.

- Where the relative standard error of predicted percent treated was greater than 30% for sex or age breakdowns at the region level, the Cost Calculator does not report output by these subgroups.
- All default percent treated estimates were weighted using the MEPS sampling weights to be nationally representative. Nationally-representative estimates by payer/sex/age were then applied to state-level population data (see p. 4).
- Treated populations in the MEPS data did not include those in long-term care (LTC) facilities since the target population for the MEPS is the civilian non-institutionalized population. Therefore, all MEPS percent treated estimates were scaled to account for non-coverage of the LTC population. The scaling factor, calculated by combining percent treated estimates in MEPS and NNHS, was the ratio of overall percent treated including the LTC population to percent treated in the non-LTC population. Separate ratios were used for each payer/sex/age subpopulation. However, because NNHS did not include regional identifiers, scaling factors at the national level were used for each of the Census regions in MEPS.
- Children (i.e., ages 0 to 17) were assumed to have zero percent treated for all diseases except asthma and depression. The small number of children with the other diseases in MEPS prevented reliable estimates for this age group. As a check of this assumption, in separate analyses we conducted using Medicaid claims data from Illinois, Indiana, Kansas and Louisiana (unpublished), which had sufficient number of children with the diseases reported in the Cost Calculator, percent treated estimates for children were very small: heart disease = 0.25%, congestive heart failure (CHF) = 0.03%, hypertension = 1.5%, stroke = 0.04%, diabetes = 0.20%, and cancer = 0.12%.

Per-Person Medical Costs

Regression modeling was used to estimate the costs associated with selected chronic diseases. The regression approach isolated the impact of diseases on health care costs while controlling for the presence of other important drivers of medical expenditures. Models were run for each payer population. Sample sizes were too small to estimate separate models by sex by age. Results reported by sex and by age category were generated from weighted averages of simulations by sex by age using the regression model for the entire payer population (see below for description of simulations). Nonlinearities in the overall model will generate different average costs per person for the different sex/age groups.

- A two-part regression model was used to predict annual per-person costs. The first part of the two-part model used logistic regression to predict the probability of any expenditure. The second part of the two-part model analyzed expenditures conditional on having any expenditures (i.e., any utilization that generated expenditures >0). To select the appropriate cost estimation model for the second part of the two-part model, we used the algorithm for choosing among alternative non-linear estimators recommended by Manning and Mullahy.⁹ The results of this algorithm indicated that we should use a generalized linear model (GLM) with a gamma distribution and a log link in the second stage to analyze non-zero expenditures.

- The dependent variable in each regression was restricted to the annual medical expenditures paid by each payer.
- All regressions included the following variables: indicators for arthritis, asthma, cancer, congestive heart failure (CHF), coronary heart disease (CHD), hypertension, stroke, other heart diseases,[†] depression, and diabetes; age; age squared; sex; race/ethnicity; education; family income; other sources of health insurance; year indicators; and indicators for other diseases including injuries, dyslipidemia, HIV/AIDS, pneumonia, chronic obstructive pulmonary disease, other mental health/substance abuse, back problems, skin disorders, renal failure, and pregnancies.
- Hypertension and diabetes are risk factors for other diseases included in the cost estimation model; therefore, the full model likely underestimated the true costs of hypertension and diabetes. To capture the costs of diseases on the causal pathway, the estimates for hypertension and diabetes are from alternate models we created that omitted controlling for diseases linked to hypertension and diabetes.
 - The hypertension model dropped CHF, CHD, other heart diseases, renal failure, and stroke.
 - The diabetes model dropped CHF, CHD, other heart diseases, hypertension, renal failure, stroke and dyslipidemia.
- Relative standard errors for the national MEPS Medicaid disease expenditures, the smallest sample among payers in the analysis, range from 21% to 35% across diseases.
- Confidential MEPS data that identified the largest 30 states and 9 census divisions were utilized to generate state-specific per-person cost estimates.
 - Payer sample sizes were not large enough by state to replicate the full analysis for each state.
 - We regressed log (positive) medical expenditures on the variables in the model plus state/census division dummies.
 - The coefficients on the dummies provided measures of the differences in average medical costs across states that were used to scale the national estimates to make them state-specific.
- The MEPS does not include expenditures associated with persons residing in institutions. The following steps were taken to include LTC costs in the per-person cost estimates:
 - We used a “top-down” approach to generate disease-specific LTC costs for all payers combined, Medicaid, and Medicare. Private insurance is not a major payer for LTC so these costs were not estimated for private insurers. All of the steps that follow were performed separately for each payer.
 - There are no nationally-representative data sets that include both diagnoses and medical expenditures for the LTC population. Therefore, we assumed that LTC costs are proportional to length of stay, which is available in the NNHS. The assumption was that the LTC costs are based on per diem payments and are constant across diseases except for length of stay. Acute treatment for specific conditions is often charged to physicians or other settings, which will be captured in the non-LTC cost category.

[†] Other heart diseases includes rheumatic fever/rheumatic heart disease, diseases of mitral and aortic valves and other endocardial structures, acute and chronic pulmonary heart disease, acute and other pericardial and endocardial disease, cardiomyopathy, conduction disorders, cardiac dysrhythmias and ill-defined heart disease. See Appendix for specific diagnosis codes.

- From the 2004 NNHS, we calculated the total number of days for patients with a primary diagnosis of each disease at admission using the date of admission.
- Next, we calculated the percentage of all nursing home days accounted for by each primary diagnosis. The use of primary diagnosis assumed that length of stay (and LTC expenditure) was not compounded by the presence of other diseases.
- Then we multiplied the percentage of all nursing home days for each primary diagnosis (disease) by national health accounts total nursing home expenditures to get disease-specific total nursing home expenditures.¹⁰
- Next, we divided disease-specific nursing home expenditures by the total number of people with the disease (LTC from NNHS and the non-LTC, non-institutionalized population from MEPS).
- Finally, we added the above to the MEPS per-person cost for each disease. The national LTC per-person cost estimate was added to the MEPS estimates before converting to state-level estimates using the state/census division multipliers.
- Costs per person attributable to each disease were calculated using the following method that minimizes double-counting of expenditures across diseases.¹¹
 - First, every unique combination of the chronic diseases observed in the data was identified (e.g., stroke, stroke with hypertension, and stroke with CHF are three of the unique combinations).
 - Second, expenditures were predicted for each individual.
 - Then, for each unique combination of diseases, we subtracted from step two the predicted expenditures for an otherwise identical person without the combination of diseases. This provides an estimate of the costs attributable to every unique combination of diseases.
 - Next, the coefficients of the diseases from the second part of the two part regression model were used as importance weights to redistribute costs associated with jointly-occurring diseases to constituent diseases (e.g., to redistribute the costs of stroke with hypertension back to stroke and hypertension separately).
 - Finally, the application averages the redistributed costs over the population with each disease.
 - There should not be any double-counting of costs between the LTC and non-LTC analyses because the former only includes the LTC component of costs, which are excluded entirely from the non-LTC analyses.

Per-Person Absenteeism Costs

Regression modeling was used to estimate absenteeism costs associated with selected chronic diseases. Please note that other indirect costs of chronic disease, including productivity losses through presenteeism and premature mortality and reductions in the quality of life, are not included in the estimates. The model for work days missed was estimated for the adult working population from MEPS. Sample sizes were too small to estimate separate models by sex by age. Results reported by sex and by age category were generated from weighted averages of simulations by sex by age using the regression model for the entire working population (see below for description of simulations). The nonlinear regression model will generate different average work days missed per person for the different sex/age groups.

- A negative binomial model was used to predict annual work days missed per working person. Negative binomial models, an extension of Poisson models, are used when the dependent variable is a count (i.e., a non-negative integer).
- The dependent variable was annual work days missed due to illness or injury.
- All regressions included the following variables: indicators for arthritis, asthma, cancer, CHF, CHD, hypertension, stroke, other heart diseases, depression, and diabetes; age; age squared; sex; race/ethnicity; education; family income; sources of health insurance; family size; occupation; year indicators; and indicators for other diseases including injuries, dyslipidemia, HIV/AIDS, pneumonia, chronic obstructive pulmonary disease, other mental health/substance abuse, back problems, skin disorders, renal failure, and pregnancies.
- Hypertension and diabetes are risk factors for other diseases included in the absenteeism estimation model; therefore, the full model likely underestimated the true absenteeism costs of hypertension and diabetes. To capture the absenteeism costs of diseases on the causal pathway, the estimates for hypertension and diabetes are from alternate models we created that omitted controlling for diseases linked to hypertension and diabetes.
 - The hypertension model dropped CHF, CHD, other heart diseases, renal failure, and stroke.
 - The diabetes model dropped CHF, CHD, other heart diseases, hypertension, renal failure, stroke and dyslipidemia.
- Because asthma and depression are prevalent among children, we also estimated negative binomial models for the annual number of school days missed due to illness or injury using the full specification above. We assumed that working parents must also miss work on the days that their children miss school.
- Work days missed per person attributable to each disease was calculated using the same method as above for medical cost, which minimizes double-counting across diseases.¹¹

Calculated Costs

- The total payer population with a disease is calculated by multiplying the payer population by the percent treated.
- The total annual payer costs attributable to a disease are calculated by multiplying the average annual costs per person with the disease by the payer population with the disease.
- The total payer costs for all selected diseases are calculated by summing the total annual payer costs attributable to each disease. However, as described above, the cost estimates for hypertension and diabetes include the costs of complications such as CHD, CHF, stroke, and other heart diseases. *The sum of costs over selected diseases that include hypertension and diabetes could overestimate the costs associated with all the selected diseases.* The Cost Calculator provides estimates for “Diseases of the Heart” and “Total CVD” that avoid double-counting of costs across diseases. The costs for “Diseases of the Heart” include CHD, CHF, and other heart disease. The costs for “Total CVD” include “Diseases of the Heart,” stroke, and an estimate of hypertension costs that avoids double-counting of costs with other diseases. Excluding the costs of complications lowers the estimates for hypertension and diabetes by approximately 34% and 39%, respectively.
- The total absenteeism costs attributable to a disease are calculated by multiplying the following: the number of people treated for the disease, the percentage of people treated

for the disease that are employed, the average annual work days missed per person due to the disease, and average daily earnings.

- The percentage of people treated for the disease that are employed was taken from MEPS. For children with asthma and/or depression, we used the percentage of children whose parents work full time.
- Average daily earnings were taken from the 2009 CPS. The earnings were not disease specific.¹² We used average daily earnings for females ages 18 to 44 to proxy for parents of young children with asthma and/or depression.

Medical Cost Projections

We generated projections of the medical cost of the chronic diseases to all payers combined through 2020 using default input data from MEPS and the U.S. Census Bureau. The medical cost projections include nursing home costs but exclude absenteeism costs. The projections are reported in real 2010 dollars and do not project inflation. **The projections assume no changes in policy or technology and exclude changes due to the Affordable Care Act (PL 111-148).** All the changes in the real medical cost of disease are driven by growth in real medical costs¹³⁻¹⁵ and change in the treated population, which is in turn driven by changes in population size and age and gender distributions.

- We obtained projections of population counts by sex by age for each state through 2020 from the Census.
- We then applied the percent treated for each disease by sex by age from MEPS to the Census population counts by sex by age and by state through 2020 to obtain projections of the number of people with each disease by sex by age by state.
- Next, we multiplied the number of people with each disease by sex/age/state by the state-specific per-person medical cost estimate by sex by age (see above).
- We used Congressional Budget Office assumptions for future health care cost growth above and beyond growth due to population growth and aging.¹⁴⁻¹⁵ We assumed that per-person costs will increase at the same rate as overall medical expenditures between 2010 and 2020: an average annual rate of 3.6%.
- Finally, we aggregated cost estimates across sex and age cells for each state to obtain total medical cost projections for each disease by state through 2020.

External Review

The methods used in generating estimates for the Cost Calculator were reviewed by an external panel of subject matter experts in the autumn of 2010. Comments from Drs. Steven Teutsch and Allison Rosen were discussed and responses agreed upon via conference call.

Glossary of Terms

Absenteeism: Absence from work. The Cost Calculator's primary measure of lost productivity is absence from work due to chronic disease.

Attributable: The portion of expenditures directly associated with a disease.

BRFSS: Behavioral Risk Factor Surveillance System

CHD: Coronary heart disease

CHF: Congestive heart failure

CMS: Centers for Medicare & Medicaid Services

CPS: Current Population Survey

CVD: Cardiovascular disease

Generalized linear model: A type of non-linear estimation used to compute the relationship between a set of independent variables and a dependent variable that has a non-normal distribution. In this case, the independent variables are the demographics and the disease indicators and the dependent variable is strictly positive payer expenditures, which is positively skewed.

ICD-9: International Classification of Diseases, Ninth Revision

Logistic regression: A type of non-linear estimation used to compute the relationship between a set of independent variables and a discrete dependent variable. In this case, the independent variables are the demographics and the disease indicators and the dependent variable is the probability of having positive payer expenditures.

LTC: Long-term care

MEPS: Medical Expenditure Panel Survey

MSIS: Medicaid Statistical Information System

Negative binomial model: A type of non-linear estimation used to compute the relationship between a set of independent variables and a count dependent variable (i.e., a dependent variable that takes on non-negative integer values). In this case, the independent variables are the demographics and the disease indicators and the dependent variable is the annual number of work days missed due to illness or injury.

NNHS: National Nursing Home Survey

Per-person costs: Average payer costs per person with the disease attributable to the disease. These numbers represent the extent to which payer expenditures per person with the disease would be lower in the absence of the disease, all else constant.

Treated population: The number of people receiving care for the disease in the previous year.

References

1. U.S. Census Bureau. <http://www.census.gov/popest/data/intercensal/state/state2010.html>. Last accessed March 10, 2011.
2. U.S. Census Bureau. <http://www.census.gov/population/projections/files/stateproj/DownldFile4.xls>. Last accessed June 6, 2011.
3. Kaiser Family Foundation. <http://kff.org/medicare/state-indicator/total-medicare-beneficiaries/>. Last accessed May 5, 2011.
4. Centers for Medicare & Medicaid Services. <http://www.cms.hhs.gov/MedicaidDataSourcesGenInfo/MSIS/list.asp>. Last accessed May 5, 2011.
5. U.S. Census Bureau. http://www.census.gov/hhes/www/cpstc/cps_table_creator.html. Last accessed May 5, 2011.
6. Full documentation and data is available from Agency for Healthcare Research and Quality, <http://www.meps.ahrq.gov/>. Accessed November 16, 2006.
7. Agency for Healthcare Research and Quality. http://www.meps.ahrq.gov/mepsweb/about_meps/Price_Index.shtml. Last accessed June 1, 2011.
8. National Center for Health Statistics. <http://www.cdc.gov/nchs/nnhs.htm>. Last accessed March 16, 2009.
9. Manning WG, Mullahy J. Estimating log models: To transform or not to transform? *Journal of Health Economics* 2001;20:461-494.
10. Centers for Medicare & Medicaid Services. http://www.cms.hhs.gov/NationalHealthExpendData/02_NationalHealthAccountsHistorical.asp#TopOfPage. Last accessed March 17, 2009.
11. Trogon JG, Finkelstein EA, Hoerger TJ. Use of econometric models to estimate expenditure shares. *Health Services Research* 2008;43(4):1442-1451.
12. Haddix AC, Teutsch SM, Corso PS. Prevention effectiveness: a guide to decision analysis and economic evaluation. 2nd ed. New York, NY: Oxford University Press; 2003.
13. Heidenreich, P.A., Trogon, J.G., Khavjou, O.A., Butler, J., Dracup, K., Ezekowitz, M.D., & et al. (2011). Forecasting the future of cardiovascular disease in the United States: A policy statement from the American Heart Association. *Circulation*, 123 (8):933-944.
14. Congressional Budget Office. Updated Long-Term Projections for Social Security. August 2008. The Congress of the United States: Washington DC. Available from: <http://www.cbo.gov/ftpdocs/96xx/doc9649/08-20-SocialSecurityUpdate.pdf>.
15. Congressional Budget Office. The Long-Term Budget Outlook, June 2009. The Congress of the United States: Washington DC. Available from: <http://www.cbo.gov/ftpdocs/102xx/doc10297/06-25-LTBO.pdf>.

Table 1: Chronic Disease Definitions—Clinical Classification (CC) Code* and 3-digit ICD-9

Arthritis	ICD-9: 274, 354, 390, 391, 443, 446, 710-716, 719-721, 725-729
Asthma	CC: 128 ICD-9: 493
Cancer	CC: 11-43, 45
Cardiovascular Disease (CVD)	
Congestive heart failure	ICD-9: 428
Coronary heart disease	ICD-9: 410-414
Hypertension	ICD-9: 401-405
Stroke	ICD-9: 430-434, 436-438
Other heart diseases	ICD-9: 390-392, 393-398, 415-416, 420-427, 429
Depression	ICD-9: 296, 311
Diabetes	ICD-9: 250

*<http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>.