June 2024

Diabetes State Burden Toolkit

Technical Report

CONTENTS

TABLES

Number **Page**

This technical report was developed by RTI International under contract numbers 75D30122P14019 and 75D30122P14908 from the Centers for Disease Control and Prevention. Report contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention, the Department of Health and Human Services, or the U.S. government.

We would like to acknowledge the Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP) for the staffs' valuable collaboration to produce reports for states whose data were not publicly available. This project would not be possible without the contributions of HCUP Partners from across the United States (list of state organizations is available at https://www.hcup-us.ahrq.gov/db/hcupdatapartners.jsp); and state partners that contributed to the Behavioral Risk Factor Surveillance System. The 2021 Florida Behavioral Risk Factor Surveillance System data used in this report were collected by the Florida Department of Health (FDOH). The views expressed herein are solely those of the author(s) and do not necessarily reflect those of the FDOH.

1. INTRODUCTION

Diabetes is a serious health condition and is a major contributor to heart disease, kidney disease, stroke, vascular disease, and vision loss. The Centers for Disease Control and Prevention (CDC), Division of Diabetes Translation, has been working with state health departments to reduce the health and economic burdens associated with diabetes-related complications and premature death due to diabetes by improving diabetes care and management. State health departments and other groups interested in diabetes prevention and control have an urgent need for information on the health and economic burdens of diabetes in their respective states. In 2016, CDC released the Diabetes State Burden Toolkit, which was developed under a contract with RTI International. The toolkit reported state-specific data on the health, economic, and mortality burdens of diabetes. Estimates reported in that toolkit were based on data from 2013 and are now outdated.

In 2022, CDC contracted with RTI again to develop updated estimates of the health, economic, and mortality burdens of diabetes among U.S. adults in each state and the District of Columbia (DC) (hereafter referred to as states) and at the national level. These new estimates are included in the updated version of the Diabetes State Burden Toolkit. The first set of updates in the toolkit is the use of the most recently available data. States and other organizations may also be interested in understanding disparities in the burden of diabetes within their states. Thus, a second set of updates to the toolkit includes additional stratifications of diabetes outcomes by race/ethnicity, income, education level, and rural/urban status where these data are available.

This report describes the data and methods used to generate the updated estimates for each section of the toolkit. Each section of the report includes subsections that refer to a specific toolkit section.

2. DATA AND METHODS

The diabetes burden toolkit consists of three sections: (1) diabetes health burden, (2) diabetes economic burden, and (3) diabetes mortality and health-related quality of life. In the following subsections, we describe the data and methods used to generate updated estimates for each section of the toolkit. The estimates reported in the toolkit are for U.S. adults only, excluding children younger than age 18. The reported estimates are for both type 1 and type 2 diabetes combined because of data limitations.

We have updated the toolkit estimates using most recently available data at the time of the analysis. In 2020, COVID-19 increased mortality and caused severe disruptions to the healthcare system and society, including shutdowns, stay-at-home orders, and delays in care. Furthermore, the impact of COVID-19 may have been even more significant for people with diabetes as, when contracted with COVID-19, they have an increased risk of complications and mortality compared with people without diabetes (ADA, 2023). To some extent, the issues caused by COVID-19 persisted into 2021 as mortality levels remained high, but the healthcare system and society adjusted; thus, the disruptions became less severe. Because 2020 did not accurately represent typical healthcare utilization due to disruptions in healthcare access and delivery, data from 2020 were omitted for estimates of diabetes-related healthcare utilization and costs.

2.1 Diabetes Health Burden

This section of the burden toolkit reports the health burden of diabetes in each state and at the national level; it describes statistics on diabetes prevalence, diabetes incidence, and diabetes-associated conditions. The information on diabetes-associated conditions described in the toolkit is based on self-reported survey data, hospitalization data, and Medicare data.

The following annual estimates are reported in the health burden section of the toolkit at the national and state levels:

1. Diabetes prevalence
    a. Age-adjusted prevalence of diabetes, overall and by sex
    b. Estimated numbers of people with diabetes, overall and by age group/sex
    c. Prevalence of diabetes, overall and by the following categories (with 95% confidence intervals [CI])
        i. Age group/sex
        ii. Race/ethnicity
        iii. Education level
        iv. Income level
        v. Rural and urban areas

2. Diabetes incidence
   a. Crude rate of newly diagnosed diabetes cases per 1,000 population (with 95% CI)
   b. Age-adjusted rate of newly diagnosed diabetes cases per 1,000 population (with 95% CI)
   c. Number of newly diagnosed diabetes cases (with 95% CI)
3. Associated health conditions
   a. Self-reported data
      i. Age-adjusted prevalence of selected conditions among adults with diabetes, overall and by sex
      ii. Prevalence of selected conditions among adults with diabetes, overall and by age group (with 95% CI)
      iii. Number of adults with diabetes and selected conditions, overall and by age group
      iv. Number of cases with selected conditions attributable to diabetes, overall and by age group
   b. Hospitalization data
      i. Age-adjusted rate of hospitalizations with selected conditions among adults with diabetes, overall and by sex
      ii. Number of hospitalizations with selected conditions among adults with diabetes by age group/sex, race/ethnicity, and urban/rural status
      iii. Rate of hospitalizations with selected conditions among adults with diabetes by age group/sex, race/ethnicity, and urban/rural status (with 95% CI)
      iv. Number of diabetes-attributable hospitalizations with selected conditions by age group/sex
   c. Medicare data
      i. Age-adjusted prevalence of selected conditions among Medicare beneficiaries with diabetes by age group/sex
      ii. Prevalence of selected conditions among Medicare beneficiaries with diabetes by age group/sex and urban/rural status (with 95% CI)
      iii. Number of people with selected conditions among Medicare beneficiaries with diabetes by age group/sex and urban/rural status
      iv. Number of selected conditions attributable to diabetes

Each component of the health burden section is described in detail in the following subsections.

2.1.1 Diabetes Prevalence

We used the 2021 Behavioral Risk Factor Surveillance System (BRFSS) data and followed approaches used by the CDC United States Diabetes Surveillance System (USDSS) (available at https://gis.cdc.gov/grasp/diabetes/diabetesatlas-surveillance.html) to estimate prevalence of diabetes. BRFSS is a state-based, cross-sectional telephone interview survey sponsored by CDC and conducted by state health departments annually. The survey covers the civilian noninstitutionalized adult population in each of the 50 states and DC. BRFSS collects prevalence data regarding health-related risk behaviors, chronic health conditions, and preventive healthcare practices among U.S. adults. The 2021 BRFSS data file that included all states except Florida was downloaded directly from the CDC BRFSS website. The 2021 Florida BRFSS dataset was obtained directly from the Florida Department of Health.

We excluded survey responses with missing age or missing diabetes status from analysis, and applying BRFSS sample weights, we calculated the percentage of adults who answered "yes" to the survey question, "Has a doctor, nurse, or other health professional ever told you that you had diabetes?" We calculated diabetes prevalence for each state by respondent's age group and sex, race/ethnicity, and education attainment; household income level; and rural and urban status of county of residence. We used data presentation standards developed by CDC's National Center for Health Statistics (NCHS) to evaluate whether each stratification provided reliable estimates and could be reported without suppressing the data (Parker et al., 2017; Ward, 2019). We multiplied the percentage of people with diabetes in each category by the weighted number of total respondents in that category to estimate the total number of adults with diabetes in each category. We also age-adjusted estimates of prevalence of diabetes, overall and by sex, to the 2000 U.S. Standard Population following the methodology described by Klein et al. (2001).

Age group and Sex

We used the following age groups for the age/sex estimates of diabetes prevalence: 18–44, 45–64, 65–74, and 75 years or older. The data were not reliable by sex in the 18 to 44 age category in DC and nine states (Alaska, Delaware, Florida, North Dakota, New Hampshire, Nevada, South Dakota, Vermont, and Wyoming) because they did not meet at least one of the NCHS data presentation standards (Parker et al., 2017, Ward, 2019). For these locations, we combined the data for men and women and reported only one estimate for the 18 to 44 age group (without the sex stratification).

Race/Ethnicity

We used the following race/ethnicity categories: White non-Hispanic, Black non-Hispanic, Hispanic, and Other Races non-Hispanic. The data for the Black Non-Hispanic group were insufficient in the following 11 states: Hawaii, Idaho, Maine, Montana, New Hampshire, New Mexico, North Dakota, Oregon, South Dakota, Vermont, and Wyoming. The data for the

Hispanic group were not reliable in six states, namely Mississippi, New Hampshire, North Dakota, South Dakota, Tennessee, and Vermont. We did not report prevalence of diabetes for these racial/ethnic groups in these states.

Education

We used the following education attainment categories based on the highest grade or years of school completed: less than high school, high school graduate, and more than high school. These categories align with those used in the CDC USDSS.

Income

We used the following income categories based on the annual household income: low income (<$35,000), middle income ($35,000 - <$75,000), and high income ($75,000 or more).

Rural and Urban Areas

BRFSS uses the 2013 NCHS Urban-Rural Classification Scheme to categorize U.S. counties; the Scheme states that urban counties include large central metropolitan, large fringe metropolitan, medium metropolitan, small metropolitan, and micropolitan counties (Ingram et al, 2013). Rural counties include noncore counties (i.e., nonmetropolitan counties that do not qualify as micropolitan). DC and seven states (Connecticut, Delaware, Hawaii, Massachusetts, New Hampshire, New Jersey, and Rhode Island) did not have any respondents from rural counties in the 2021 BRFSS. In two other states (California and Nevada), the 2021 data for diabetes prevalence were not reliable in rural counties because they did not meet at least one of the NCHS data presentation standards (Parker et al., 2017, Ward, 2019). We did not report prevalence of diabetes by urban and rural status in these locations.

2.1.2 Diabetes Incidence

For diabetes incidence, we obtained data from the CDC USDSS (https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html). These estimates reflect crude and age-adjusted rates of newly diagnosed diabetes cases for 2021 (for all states, except Florida) and for 2020 for Florida. The estimates were derived from the BRFSS data and represented as a two-year average based on 2020 and 2021 survey data.

Self-reported diagnosed diabetes was defined by using valid responses to the two survey questions in series, (1) the self-reported diabetes status, “Has a doctor, nurse, or other health professional ever told you that you had diabetes?” and (2) the age when the respondent was diagnosed with diabetes for the first time, “How old were you when you were told you had diabetes?” The number of years each person was living with diagnosed diabetes was calculated by subtracting the age when they were diagnosed from their current age at the time of survey. Adults who had a value of zero were identified as being diagnosed with diabetes within the past year. In addition, half of the adults who had a value of one were classified as being diagnosed with diabetes within the past year. In calculating the rate of newly diagnosed diabetes cases, the numerator was the weighted number of adults who were diagnosed with diabetes within the past year, and the denominator was the weighted adult population estimate, excluding those who had been diagnosed with diabetes for more than 1 year or who answered "refused," "don't know," or had missing values on the diabetes status question.

For national estimates of newly diagnosed diabetes, we report the median crude and age-adjusted rates across all states, and the total number of newly diagnosed cases of diabetes calculated as a sum of the number of new cases across the states.

2.1.3 Diabetes-Associated Conditions

The toolkit reports statistics on selected diabetes-associated conditions from three types of data sources: (1) self-reported chronic health conditions from BRFSS, (2) hospitalization events from State Inpatient Databases (SID), and (3) claims from the Medicare beneficiary data. BRFSS-based self-reported health conditions are hypertension, high cholesterol, blindness, mobility limitations, limitations in instrumental activities of daily living, and coronary heart disease (CHD). Estimates generated from self-reported data are mostly based on positive responses to disease status questions ("Have you ever been told that you had…"). Hospitalization events obtained from SID are congestive heart failure (CHF), stroke, myocardial infarction (MI), lower extremity amputations (LEAs), hyperosmolar hyperglycemic nonketotic syndrome (HHNS), diabetic ketoacidosis (DKA), and hypoglycemia. Results from SID represent medical episodes for which individuals were hospitalized within a given year. Claims from the Medicare data are used to report CHD, CHF, chronic kidney disease (CKD), and peripheral vascular disease (PVD). Estimates from Medicare data represent clinical encounters for which eligible Medicare beneficiaries received medical treatment within a given year. Because of differences in definitions of conditions from different data sources, estimates of the same diseases (e.g., CHD) across the data sources are not comparable. For each condition, the toolkit reports the rate of the condition among people with diabetes and the number of cases attributable to diabetes.

2.1.2.1 Self-reported Data

We used the 2021 BRFSS data to estimate the prevalence of self-reported diabetes-associated conditions. A list of these selected conditions and survey questions used to assess each of them are presented in Table 2-1.

Table 2-1. Definitions of Self-Reported Diabetes-Associated Conditions, Behavioral Risk Factor Surveillance System, 2021

Condition	Definition
Hypertension (high blood pressure)	Have you ever been told by a doctor, nurse, or other health professional that you had high blood pressure? (If "Yes" and respondent is female, ask "Was this only when you were pregnant?")
High cholesterol	Have you ever been told by a doctor, nurse or other health professional that your blood cholesterol is high?
Blindness	Are you blind or do you have serious difficulty seeing, even when wearing glasses?
Mobility limitations	Do you have serious difficulty walking or climbing stairs?
Limitations in instrumental activities of daily living	Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone such as visiting a doctor's office or shopping?
Coronary heart disease	Has a doctor, nurse, or other health professional ever told you that you had angina or coronary heart disease? OR Has a doctor, nurse, or other health professional ever told you that you had a heart attack, also called a myocardial infarction?

We estimated the prevalence of each selected self-reported condition among people with diabetes by calculating the percentage of people with diabetes who responded "yes" to the corresponding health status question based on survey sample weights. We excluded respondents with missing information on age and diabetes status from analyses of all selected conditions. For each condition, we also excluded respondents with missing information for that specific condition. Additionally, women reporting diabetes or hypertension during pregnancy only (gestational (pregnancy-induced) diabetes and hypertension) were considered to not have the condition (diabetes or hypertension). We report prevalence estimates by four age categories (18–44, 45–64, 65–74, and 75 years or older). As with diabetes prevalence, we followed data reporting standards developed by NCHS, and we aggregated age categories that did not meet these standards (Parker et al, 2017, Ward, 2019). The level of aggregation varied across conditions and across states. For example, for hypertension, we report its prevalence for all four age categories in 28 states; however, in the remaining 22 states and DC, we combined the two youngest age categories and report the prevalence of hypertension for only three age groups: 18–64, 65–74, and 75 years or older. We calculated the estimated number of people with each condition as well as diabetes by age by multiplying the prevalence of the selected condition among people with diabetes in each age category by the weighted number of people with diabetes from BRFSS in each age category.

We estimated the number of condition cases attributable to diabetes using an attributable fraction (AF) approach. In the epidemiologic literature, AFs are used to estimate the proportion of disease risk in a population that can be attributed to a single risk factor or a set of multiple risk factors (Flegal, Graubard, & Williamson, 2004; Rockhill, Newman, & Weinberg, 1998). Because the prevalence of diabetes and its attributable conditions increase with age, the AFs were estimated separately by age group. According to Rockhill, Newman, and Weinberg (1998) and Flegal, Graubard, and Williamson (2004), when confounding factors and/or effect modifications are present, the correct formula for calculating the diabetes AF for disease i is shown below in Equation 1:

$$AF_i = pd_i\left[\frac{RR_i - 1}{RR_i}\right] \tag{1}$$

where pd_i is the adjusted prevalence of diabetes in the subsample with the condition *i*, and RR_i is the adjusted relative risk (RR) of condition *i* in the diabetes subsample relative to the non-diabetes subsample. For each age group, we predicted the probability of having diabetes among individuals with the condition (pd_i) using a logit command in Stata and controlling for age (in years), sex, and race/ethnicity (Black non-Hispanic, Hispanic, Other Races non-Hispanic [including missing race], and White non-Hispanic [variable omitted from the regression model]). The model was weighted using BRFSS sample weights to account for the complex survey design.

For each age group, we also estimated the RR of each condition, which is the ratio of the condition prevalence among people with diabetes to the condition prevalence among people without diabetes (see Equation 2). We estimated the RR using a generalized linear model (GLM) with a Poisson family and a log link, controlling for age, sex, and race/ethnicity. The GLM regressions were also weighted using BRFSS sample weights.

$$RR = \frac{\text{Complication Prevalence Among People with Diabetes}}{\text{Complication Prevalence Among People without Diabetes}} \tag{2}$$

Our standard specification included four race/ethnicity groups: Black non-Hispanic, Hispanic, Other Race non-Hispanic (including missing race), and White non-Hispanic (omitted category). However, we identified quasi-complete separation (QCS) in several age/condition/state stratifications, which occurs when all individuals in one race/ethnicity group have a single-value outcome variable such as all zeroes or all ones (e.g., every Hispanic aged 18 to 44 who has CHD also has diabetes given their age [in years]). When QCS occurred, we aggregated race/ethnicity into three groups: Black non-Hispanic, Other Races (including missing race and Hispanics), and White non-Hispanic (omitted category). If QCS was still present in the more aggregated model, we used results from a model without race/ethnicity variables.

For each age/condition/state stratification, we used the same model specification for predicting probability of diabetes as for estimating the RR. For example, if we used results from a logistic regression model without race/ethnicity controls to predict the probability of diabetes among adults aged 18 to 44 with CHD in one state, then we used the same specification (i.e., without race controls) in the GLM regression predicting the RR in that state.

We estimated the number of cases of each selected condition attributable to diabetes by multiplying the number of cases of each condition by the diabetes AF (see Equation 3):

$$\text{Number of condition}_i \text{ cases attributable to diabetes} = \text{Number of people with condition}_i * \text{AF}_i \qquad (3)$$

Note that in some states where the prevalence of the diabetes-associated condition was estimated for three age groups and overall (instead of four age groups and overall), the number of attributable cases is also reported for three age groups and overall. At the national level, estimates are reported for four age groups and overall; thus, the total number of diabetes-attributable cases summed across four age groups does not add up to the reported grand total (because in some states the data are not available by four age groups). For example, in Alaska, the number of diabetes-attributable cases of high blood pressure is reported for age groups 18 to 64, 65 to 74, 75 years or older, and overall (18 years or older). At the national level, the number of cases reported for the 18 to 44 or 45 to 64 age groups does not include the estimated number from Alaska, but they are included in the overall total count for ages 18 years or older.

We do not report the number of selected chronic health condition cases attributable to diabetes for any age category where the p-value for the RR was >0.10.

2.1.2.2 Hospitalization Data

We used the SID data from the Healthcare Cost and Utilization Project (HCUP) sponsored by the Agency for Healthcare Research and Quality (AHRQ) to estimate hospitalizations with selected diabetes-associated conditions. The SIDs capture hospital inpatient stays in a given state and contain clinical and resource-use information that is included in a typical hospital discharge summary. We used publicly available data for 34 states and DC that we obtained from the HCUP Central Distributor. Estimates for nine other states that participate in HCUP but do not make their discharge data available through the HCUP Central Distributor were generated through an intramural collaboration with AHRQ. Data for four remaining states (California, Connecticut, Louisiana, and Ohio) that participate in HCUP were not available.

The most recent year for which the SID files were available varied across the states, ranging from 2016 to 2021. For states where the most recent data were from 2020, we used the data from 2019 because inpatient utilization in 2020 was greatly impacted by COVID-19

(Table 2-2). New Hampshire's latest year of available data is 2009, and the state was therefore excluded from our analyses. Two states (Alabama and Idaho) do not participate in the HCUP SID; thus, we cannot report hospitalization data for them.

Table 2-2. The Year of State Inpatient Databases Used

State	Year of Data
Alaska	2019
Arizona	2021
Arkansas	2019
Colorado	2019
Delaware	2019
District of Columbia	2019
Florida	2019
Georgia	2019
Hawaii	2016
Illinois	2019
Indiana	2019
Iowa	2021
Kansas	2019
Kentucky	2021
Maine	2018
Maryland	2019
Massachusetts	2019
Michigan	2019
Minnesota	2019
Mississippi	2021
Missouri	2019
Montana	2019
Nebraska	2019
Nevada	2019
New Jersey	2019
New Mexico	2019
New York	2019
North Carolina	2019
North Dakota	2019
Oklahoma	2019
Oregon	2021
Pennsylvania	2019

State	Year of Data
Rhode Island	2019
South Carolina	2019
South Dakota	2019
Tennessee	2019
Texas	2019
Utah	2019
Vermont	2019
Virginia	2019
Washington	2019
West Virginia	2021
Wisconsin	2021
Wyoming	2019

Note: Data from California, Connecticut, Louisiana, and Ohio were not available. Alabama and Idaho do not participate in the HCUP SID; thus, we cannot report hospitalization data for them. New Hampshire's latest year of available data was 2009, and it was excluded from our analyses.

In SID, we identified persons with diabetes based on the International Classification of Diseases 10th Revision Clinical Modification (ICD-10-CM) codes, E10-E13, listed anywhere under the diagnosis code on their discharge record. Diabetes-associated conditions were identified using their primary (i.e., first-listed) diagnosis code (with an exception of LEAs). Selected conditions, for which we report hospitalization data, along with their respective ICD-10-CM codes are shown in Table 2-3. The counts of hospitalizations with diabetes and diabetes-associated conditions reported in the toolkit may be different from other documents published by AHRQ due to differences in the use of diagnosis codes applied to identify these conditions.

Table 2-3. Diagnosis Codes for Selected Diabetes-Associated Conditions, State Inpatient Databases

Condition	ICD-10-CM Code[a]
Congestive heart failure	I50, I11.0, I13.0, I13.2, I09.81 (AHRQ, 2015a)
Stroke	I60-I69 (Tsao et al, 2020)
Myocardial infarction	I21, I22 (AHRQ, 2015b)
Lower extremity amputation	0Y6[b] (but exclude S78, S88, S98) (AHRQ, 2015c)
Hyperosmolar hyperglycemic nonketotic syndrome	E110, E130 (Dugan et al, 2017)
Diabetic ketoacidosis	E101, E111, E131 (Dugan et al, 2017)
Hypoglycemia	E08.641, E08.649, E09.641, E09.649, E10.641, E10.649, E11.641, E11.649, E13.641, E13.649, E15, E16.0, E16.1, E16.2, T383 (Karter et al, 2019)

[a] With a reference to a particular method.

[b] Based on procedure codes.

We calculated hospitalization rate with each selected condition per 1,000 U.S. adults with diabetes by age group/sex, race/ethnicity, and rural and urban status using Equation 4:

$$\begin{array}{c}\text{Condition Hospitalization Rate}\\ \text{Per1,000 Adults with Diabetes}\end{array} = \frac{\begin{array}{c}\text{Number of People Hospitalized}\\ \text{with Condition and Diabetes}\end{array}}{\text{Number of People with Diabetes}} \times 1{,}000 \tag{4}$$

The number of people hospitalized with a selected condition and diabetes in a given state was obtained from the SID, and the estimated total number of people with diabetes in that particular state was obtained from the 2021 BRFSS data.

The two requirements for reporting hospitalization rates for a selected condition for each stratification are: (1) a reliable estimate of diabetes prevalence from the BRFSS and (2) count of hospitalizations greater than 10. For rural/urban stratifications, we are able to report hospitalization rates in 28 states for CHF, MI, stroke, LEA, and DKA; 21 states for HHNS; and 25 states for hypoglycemia. Reporting of hospitalization rates by race/ethnicity is possible whenever reliable and sufficient data are available for at least two – races/ethnicities. For race/ethnicity stratification, we are able to report hospitalization rates in 26 states for CHF, MI, stroke, LEA, and DKA; 25 states for hypoglycemia; and 23 states for HHNS.

For CHF, MI, and LEAs, we also calculated the number of hospitalizations with each condition attributable to diabetes by age/sex. We used the AF approach presented in Formula 1 where pd_i was the adjusted prevalence of diabetes among those hospitalized with condition I, and RR_i is the adjusted RR of hospitalization with condition *i* among those hospitalized with and without diabetes. For each age/sex group, we used a logistic regression model to predict the probability of having diabetes among people with the selected condition while controlling for age (in years) and race/ethnicity. In three states, age was coded in 5-year intervals. In those cases, we recoded the age variable as continuous setting it to the middle point of the 5-year interval. We followed the same approach in dealing with QCS as described for those selected conditions. We then used GLM with a Poisson family and a log link to estimate RR of each condition.

When reporting the number of estimated cases of diabetes-attributable hospitalizations in the toolkit, we rounded the estimated value to the nearest 10. We used the following rules to replace unreliable or insufficient results based on the HCUP data reporting rules:

1. Replace the number of hospitalizations with the condition and diabetes with 11 if the original number is <11 (these replacements occurred in 46 out of 2,444 state/condition/age/sex categories).
2. Replace the number of diabetes-attributable hospitalizations with zero if the number of hospitalizations with the condition and diabetes is <11 or the *p*-value for the RR is >0.10 (these replacements occurred in 61 out of 1,128 state/condition/age/sex categories).

We did not report the number of stroke hospitalizations attributable to diabetes in any state where there was high frequency of unreliable data. We assumed that all hospitalizations with HHNS, DKA, or hypoglycemia were attributed to diabetes.

2.1.2.3 Medicare Data

We used data from the Centers for Medicare & Medicaid Services (CMS) 2021 Master Beneficiary Summary File (MBSF) to estimate diabetes-associated conditions among Medicare beneficiaries with diabetes. We merged data from the MBSF Base Segment, the 30 Chronic Conditions Warehouse (CCW) Segment, the Other Chronic or Potentially Disabling Conditions Segment, and the Cost and Utilization Segment (ResDAC, 2023). We used the 30 CCW segment, as CCW recommends switching from the 27 CCW segment to the 30 CCW segment as soon as possible (CCW, 2022). We restricted our analysis to beneficiaries aged 65 or older; we also omitted individuals aged 100 or older with no healthcare use in the past 12 months to eliminate possible deceased cases. Furthermore, we restricted the analysis sample to beneficiaries with full fee-for-service (FFS) coverage during a 2-year reference period (with Part A and Part B coverage and without health management organization coverage).

The following diabetes-associated conditions were estimated from the Medicare data: CHD, CHF, CKD, and peripheral vascular disease (PVD). We used existing variables from the 30 CCW Chronic Conditions segment and the Other Chronic or Potentially Disabling Conditions segment of the MSBF to identify beneficiaries with diabetes and diabetes-associated conditions. In these data sources, the variables indicate medical treatment for a condition and are defined using algorithms based on frequencies of salient inpatient and outpatient claims within a reference time period (Chronic Conditions Data Warehouse, 2023). For all of the conditions used in our analysis, CMS uses a reference period of 2 years to identify the presence of a condition. We used variables called the end-of-year indicators to identify beneficiaries with the selected conditions, which means that the algorithm criteria were applied using December 31, 2021, as the end of the reference period. To restrict the sample to fully covered FFS beneficiaries, we excluded beneficiaries with the diabetes end-of-year flag equal to 0 (claims and coverage criteria not met) or 1 (claims met, coverage not met). Beneficiaries with the end-of-year flag indicators equal to 3 (claims and coverage met) were defined as having the condition. These end-of-year flag indicators are existing variables in the CMS data files that we used.

For each state, we calculated the prevalence of selected conditions among beneficiaries with diabetes as the percentage of all beneficiaries with diabetes who also have the selected conditions. Estimates were generated by age group/sex with two age groups (65 to 74 and 75 or older). We also reported the prevalence for the number of individuals with CAD, HF, CKD, or PVD and diabetes by rural and urban status. We used the 2013 NCHS Urban-Rural Classification Scheme to categorize U.S. counties. For rural/urban stratifications, we are able to report the prevalence for CAD, HF, and CKD in 45 states and PVD in 44 states as other remaining states and DC do not have rural counties. We excluded PVD prevalence in Massachusetts due to a small sample size.

For each state, we also calculated the number of selected condition cases attributable to diabetes using the AF approach presented earlier in Equation 1 where pd_i is the adjusted prevalence of diabetes among those with condition *i*, and RR_i is the adjusted RR of condition *i* among those with and without diabetes. For each age/sex group, we used a logistic regression model to predict the probability of having diabetes among people with the condition controlling for age (in years) and race/ethnicity. We used the same approach to address the QCS as with the BRFSS data. We then used a GLM with a Poisson family and a log link to estimate the RR of each condition. P-values for all RRs that we estimated were <0.05.

In the MSBF, CCW indicators are not available for beneficiaries enrolled in managed care, thus our estimates were based on a sample restricted to the fully covered FFS beneficiaries.

We extrapolated the number of cases with selected condition and diabetes and the number of diabetes-attributable cases to the entire Medicare beneficiary population in a given state using a state/age group/sex-specific multiplier. For each state/age group/sex stratification, this multiplier was calculated as the number of total Medicare beneficiaries divided by the number of fully covered FFS beneficiaries.

When reporting the number of estimated cases of diabetes-attributable conditions in the Burden Toolkit, we rounded the estimate to the nearest 10.

2.2 Diabetes Economic Burden

This section of the burden toolkit reports the economic burden of diabetes in each state, which consists of medical (direct) and indirect costs of diabetes. All costs are reported in 2021 dollars.

The following annual estimates are reported in the economic burden section of the toolkit at the state and national levels:

1. Total costs attributable to diabetes, overall and by age group/sex:
    a. Direct costs
    b. Indirect costs
    c. Total costs, overall and per person with diabetes
2. Medical costs attributable to diabetes:
    a. All Payers, overall and by age group/sex:
        i. Per person medical costs
        ii. Total medical costs
    b. By Payer:
        i. Per person and total medical costs paid by Medicare
        ii. Per person and total medical costs paid by Medicaid
        iii. Per person and total medical costs paid by other payers
        iv. Per person and total medical costs paid by all payer types
    c. By Payer, by age group/sex:
        i. Total medical costs paid by Medicare
        ii. Total medical costs paid by Medicaid
        iii. Total medical costs paid by other payers
        iv. Total medical costs paid by all payer types
3. Indirect costs attributable to diabetes:
    a. Total:
        i. Morbidity costs: total and per person with diabetes
        ii. Work absenteeism costs: total and per person with diabetes
        iii. Presenteeism costs: total and per person with diabetes
        iv. Household productivity losses: total and per person with diabetes
        v. Inability to work costs: total and per person with diabetes
        vi. Mortality costs: total and per person with diabetes
        vii. Total indirect costs: total and per person with diabetes

- b. Work absenteeism, overall and by age group/sex:
  - i. Number of workdays lost per employed person with diabetes
  - ii. Cost per employed person with diabetes
  - iii. Cost per person with diabetes
  - iv. Total cost
- c. Presenteeism, in total and by age group/sex:
  - i. Number of workdays lost per employed person with diabetes
  - ii. Cost per employed person with diabetes
  - iii. Cost per person with diabetes
  - iv. Total cost
- d. Household productivity losses, overall and by age group/sex:
  - i. Number of days lost per person with diabetes
  - ii. Cost per person with diabetes
  - iii. Total cost
- e. Inability to work, overall and by age group/sex:
  - i. Number of persons unable to work because of diabetes
  - ii. Cost per person with diabetes unable to work
  - iii. Cost per person with diabetes
  - iv. Total cost
- f. Mortality, in total and by age group/sex:
  - i. Number of deaths attributable to diabetes
  - ii. Labor costs
  - iii. Household productivity costs
  - iv. Total costs

4. Costs by perspective, overall and by age group/sex:
   - a. State Medicaid Program:
     - i. Estimated per person costs incurred by the state Medicaid program
     - ii. Estimated total costs incurred by state Medicaid program
   - b. Private Insurers:
     - i. Estimated per person costs incurred by private insurers
     - ii. Estimated total costs incurred by private insurers
   - c. Employers:
     - i. Estimated per person costs incurred by employers

ii. Estimated total costs incurred by employers

2.2.1 Total Costs of Diabetes

This section of the burden toolkit reports the total costs attributable to diabetes in each state, which includes both diabetes-attributable medical costs and indirect costs. Medical costs are estimated as the portion of state health expenditures from National Health Expenditure Accounts (NHEA) attributable to diabetes (including nursing home costs for institutionalized residents), as described in detail in Section 2.2.2. Indirect costs reflect the labor and household productivity losses that arise when diabetes causes missed workdays (i.e., absenteeism costs), on-the-job productivity losses (i.e., presenteeism costs), household productivity losses, diabetes-related disability that prevents people from working, or early mortality. Methods for estimating indirect costs are described in Section 2.2.3. Total costs are shown in total and by age and sex groups.

2.2.2 Medical Cost of Diabetes

This section of the burden toolkit reports diabetes-attributable direct medical costs, which are presented as costs for all payers, and costs by payer, by age group, and by sex. We used an AF approach to estimate state health expenditures attributable to diabetes. National and state health expenditures are regularly compiled by CMS (https://www.cms.gov/Research-Statistics-Data-and-systems/Statistics-Trends-and-reports/NationalHealthExpendData/index.html). To implement this approach, we first estimated the fraction of medical spending for various services attributable to diabetes. Consistently with other studies (ADA, 2018, Shultz et al, 1991, Shrestha et al., 2018), we used the AF formula presented in Equation 5 to estimate the fraction of medical costs attributable to diabetes.

$$AF = \frac{pd \times (RR-1)}{1+pd \times (RR-1)} \qquad (5)$$

where *pd* represents the prevalence of diabetes, and RR represents the ratio of medical costs for people with diabetes to medical costs for those without diabetes. The medical costs attributable to diabetes are then calculated as AF * total medical costs or expenditures.

A modified version of this AF formula is recommended when the RR is adjusted for confounding (Rockhill et al, 1998), and we used the modified version to estimate the number of diabetes-attributable complications (as shown in Equation 1). However, we could not apply the modified formula to estimate diabetes-attributable medical expenditures. In Equation 1 (the modified formula), *pd* refers to the prevalence of diabetes in a subsample of people with a specified condition. In Equation 5, *pd* refers to the prevalence of diabetes in the entire population, not a subset with a specified condition. Using Equation 1 to estimate diabetes-attributable costs would require us to estimate condition-specific costs. This approach is not feasible because other data elements used in the calculations, such as state health expenditures, are not available at the condition level. To minimize the problem of confounding encountered in Equation 5, we stratified the RR calculations by age group, by sex, by type of service, and by payer.

We applied this general approach to estimate the costs attributable to diabetes for medical services used by the noninstitutionalized population and for nursing home residents. To estimate state health expenditures attributable to diabetes, we used the 2014 NHEA from CMS. These data provided total medical expenditures by state of residence, including administrative costs and medical spending. We used the 2014 data file because it was the most recent year for which we had access to both NHEA and State Health Expenditure Account (SHEA) data. National expenditures are available by age, sex, payer (Medicaid, Medicare, or Other [which includes private insurance paid, out-of-pocket payment, and other payer paid]), and types of service (ambulatory care, hospital care, prescription drugs, nursing home care, durable medical equipment, and other care [including home health, nonprescription drugs, and nondurable medical products]). The expenditures by state in the SHEA are available by payer and type of service but not by age and sex.

To obtain state expenditure estimates by age, sex, payer, and type of medical service, we allocated state aggregate expenditures across age, sex, payer, and service categories. Specifically, we adjusted the NHEA spending by age, sex, payer, and type of service for each state using a state-specific ratio of SHEA spending by payer and by type of service to the NHEA spending. We then estimated diabetes-attributable costs by age, sex, payer, and service type and summed them to state and national levels for reporting in the toolkit. We describe our approach in more detail in the section below. We organized our approach around the following four major tasks:

1. Estimate state expenditures by age, sex, payer, and service type
2. Estimate state prevalence of diabetes (pd)
3. Estimate diabetes cost ratios (RR)
4. Estimate diabetes-attributable cost

2.2.2.1 Estimate State Expenditures by Age Group, Sex, Payer, and Service Type

Although healthcare spending likely varies by age group, sex, payer, and service type, the SHEA does not provide data broken down for all of these categories. Table 2-4 shows the availability of national- and state-level expenditure data for each of these categories. At the national level, both total and per capita spending are available for each category. Total and service-level spending are available by payer at the state level from the SHEA; however, an algorithm is required to estimate the rest of the state-level components in Table 2-4. We estimated state health expenditures by payer, age group, sex, and service type, as described briefly in this section.

Table 2-4. Availability of National Health Expenditure Accounts and State Health Expenditure Accounts Data by Payer Category

Level/Payer	Total Spending	Spending by Age, Sex	Spending by Service Type	Spending by Age, Sex, and Service Type
National	Y	Y	Y	Y
Medicare	Y	Y	Y	Y
Medicaid	Y	Y	Y	Y
Private Health Insurance	Y	Y	Y	Y
Other + OOP	Y[a]	Y[a]	Y[a]	Y[a]
State	Y		Y	
Medicare	Y		Y	
Medicaid	Y		Y	
Private Health Insurance	Y			
Other + OOP	Y[a]			
Private Health Insurance + Other + OOP	Y[a]		Y[a]	

[a] Residual of national or state minus available payers.

Other=other payers; OOP=out-of-pocket payments for insured, under-insured, and uninsured.

First, we combined NHEA and SHEA data on aggregated Personal Healthcare expenditures for 2014 to estimate expenditures for the following strata:

1. Age, in years:
   a. 0–18
   b. 19–44
   c. 45–64
   d. 65–84
   e. 85+
2. Sex:
   a. Male
   b. Female
3. Payer[1]:
   a. Medicaid
   b. Medicare (fee-for-service and managed care)

[1] SHEA includes only state-aggregated health expenditures for the privately insured and does not break down private health expenditures by age, sex, or service type. Hence, we limited our "cost by payer" analysis to include only the three original SHEA payer categories (Medicare, Medicaid, other).

   c. Other payers and programs
   d. Out-of-pocket (OOP)
   e. Private health insurance
4. Service Type:
   a. Dental services
   b. Durable medical equipment
   c. Home health care
   d. Hospital care
   e. Nursing care facilities and continuing care retirement communities
   f. Other health residential and personal care
   g. Other nondurable medical products
   h. Other professional services
   i. Physician and clinical services
   j. Prescription drugs

To estimate 2014 state expenditures by age group, sex, payer, and service type, we used a multi-step process, as summarized below:

1. Calculated per-capita 2014 NHEA costs by age group, sex, payer, and service type.
2. Generated per-capita 2014 NHEA cost estimates by payer and service type.
3. Created an adjustment index equal to 2014 state per capita spending by payer and service type relative to national per capita spending by payer and service type.
4. Multiplied this adjustment index by 2014 NHEA spending by age group, sex, payer, and service type.

We collapsed categories to the following stratifications to have a sufficient sample size for estimating all components needed for the state diabetes-attributable cost calculation:

1. Age:
   a. 19–64
   b. 65+
2. Payer (to be consistent with the state-level payer type):
   a. Medicaid
   b. Medicare
   c. Other than Medicare and Medicaid (Note: This includes private health insurance + other payers + OOP payments for insured and uninsured patients.)

3. Service Type:
    a. Hospital care
    b. Ambulatory, including physician and clinical services, other professional services
    c. Prescription drugs and other nondurable medical products
    d. Other, including dental services, durable medical equipment, home health care, and other health residential and personal care

The next steps were to calibrate the estimated total 2014 state expenditures at the payer, service type, age group, and sex levels so the aggregated cost estimates matched the 2014 actual total expenditures from SHEA. We then inflated the state health spending from 2014 to 2021 using expenditure growth from NHEA and calibrated estimates to ensure that the sum across all 2021 state estimates matched 2021 national health expenditures. After reviewing the state estimates, we changed our approach for imputing Medicaid spending, because the imputation method described above was not performing well, in the sense that the age and sex imputation based on national per-capita spending could not account for the large geographic variation across Medicaid programs in different states. This variability suggests that program benefit design and eligibility criteria are more important drivers of Medicaid spending than beneficiary's age and sex. To estimate 2021 state Medicaid costs, we used publicly available data from the Kaiser Family Foundation (KFF) on state Medicaid enrollment groups (Children [0–18], Disabled [0–64], Adults [19–64], and Aged [≥65]) and spending by enrollment group. Because the KFF state Medicaid enrollment and spending data by enrollment group did not differentiate spending by types of service, we combined spending across the four types of services in the 2014 Medicaid spending data from SHEA.

2.2.2.2 State Diabetes Prevalence

We used the 2021 BRFSS data to estimate state diabetes prevalence by age group, sex, and payer, using an approach similar to the one described in Section 2.1. We assigned payer categories (Medicare, Medicaid, other payers) using valid responses to the question, "What is the current primary source of your health insurance?" We estimated diabetes prevalence by payer, by age group (19 to 64 and 65 years or older), and by sex for each state.

2.2.2.3 Diabetes Cost Ratios

We generated diabetes cost ratios to estimate the impact of having diabetes on annual healthcare spending by type of service. We computed a diabetes cost ratio, which was the ratio of predicted costs for people with diabetes over predicted costs for people with diabetes under the scenario in which they did not have diabetes (i.e., a recycled prediction approach) using multivariate regression analysis.

Ideally, cost ratios would have been calculated for each state. However, we lacked comprehensive data containing all the variables needed to calculate diabetes cost ratios, including diabetes disease indicator, confounding variables such as socioeconomic variables and other risk factors, and healthcare expenditures by types of service, as well as the state indicator that enables us to calculate state-specific cost ratios. Although the MEPS restricted file has the majority of these required variables and the state indicator, accessing these data requires approval and onsite access and will have sample size issue in the majority of the states when calculating cost ratios at the level of granularity needed for this analysis. Using claims data would have been another possibility. However, claims data do not contain socioeconomic variables or other risk-factor variables (e.g., smoking status and obesity). Additionally, obtaining approval to use Medicaid and Medicare claims data takes additional time and resources. For these reasons, we used the publicly available MEPS data file to calculate cost ratios by payer, types of service (all services combined for Medicaid), and age group.

We used the 2015 to 2019 MEPS datasets to calculate, for each individual survey respondent, annual spending by type of service and payer, as shown in Table 2-5. Table 2-5 shows the crosswalk of service types between MEPS and NHEA.

Although MEPS asked detailed questions on survey respondents' insurance coverage, we could not use these insurance indicators directly because, in NHEA and SHEA, spending was separated by payer rather than by the primary insurance. We therefore identified the denominator population for the analysis involving Medicare and Medicaid payers as those who reported having Medicare or Medicaid as their primary insurance for at least 1 month during the survey year or those who did not self-identify as having Medicaid or Medicare, but who appeared to have payments made by Medicare or Medicaid on any of their healthcare encounters. To calculate cost ratios for other payers (including private insurance, out-of-pocket payment, and all other payers), we used the entire population from the household consolidated file as the denominator for the analysis.

Table 2-5. Medical Expenditure Panel Survey (MEPS) Spending and Payer Categories for Diabetes Cost Ratio Analysis

Types of Service for This Analysis	MEPS Service Categories	SHEA Service Type Category	Payer			
			Medicare	Medicaid	Other Payer	Private Plan
Hospital inpatient	Hospital inpatient	Hospital care	X	X	X	X
Ambulatory care	Emergency room visits, outpatient visits, and office-based provider visits	Physician and clinical services, Other professional services	X	X	X	X
Pharmacy and non-durable medical equipment	Prescription medication and nondurable medical equipment from other medical expenses	Prescription drugs, Other nondurable medical products	X	X	X	X
Other	Dental, vision, home health, and durable medical equipment from other medical expenses	Dental services, durable medical equipment, home health, other health residential, and personal care	X	X	X	X

MEPS= Medical Expenditure Panel Survey; NHEA =National Health Expenditure Accounts; SHEA=State Health Expenditure Accounts.

Notes: Although NHEA and SHEA include nursing care facilities and continuing care at retirement communities, these costs are not included in the MEPS. Hence, the attributable nursing home cost was calculated using a different approach (see Section 2.2.2.5).

Although NHEA includes only spending incurred in free-standing emergency centers in the physician and clinical services category, we were unable to distinguish free-standing emergency room visits from hospital-based emergency department visits in MEPS. Hence, we included emergency room related costs in MEPS in the Ambulatory Care category.

We do not compute the cost ratios for private insurance payer for attributable cost calculation because the SHEA does not have detailed spending by private payers. However, we calculated the cost ratios for payment by private insurance anyway along with Medicare, Medicaid, and Other.

Nursing care facilities and continuing care retirement communities were excluded from this part of the analysis because MEPS did not capture data for individuals residing in nursing homes or other institutions. The nursing home cost attributable to diabetes were instead estimated using the Minimum Data Set (MDS) collected by CMS, as described in 2.2.2.5.

Using multivariate regression analysis, we estimated the cost ratios by payer, age group (19 to 64 or 65 years or older), and sex. The denominator populations for Medicare and Other had sufficient sample sizes to calculate cost ratios by service type. However, for Medicaid, we calculated a single cost ratio (i.e., not by service type).

Multivariate regressions controlled for confounding factors, such as age, age squared, sex, race/ethnicity (White non-Hispanic, Black non-Hispanic, Hispanic, Asian non-Hispanic, Other

Races non-Hispanic), poverty status (defined by family income to poverty line ratios of <=1; >1 - <=1.25; >1.25 - <=2; >2- <=4; >4), education (no degree, high school graduate, college graduate, master/doctoral graduate, other), and Census region (East, Midwest, South, West). We also included a variable on the number of months a person was continuously covered by a particular insurance in regression analyses for Medicare and Medicaid to adjust for lengths of observation.

As a sensitivity analysis, we ran additional regressions that controlled for the following comorbidities that were high-cost drivers but not necessarily related to diabetes: arthritis, asthma, cancer, depression, injury, HIV/AIDS, pneumonia, chronic obstructive pulmonary disease (COPD), mental health conditions, back pain, and skin disorders, as well as pregnancy (included only for Medicaid and Other, but not Medicare). These comorbidities were selected as covariates in the regression analysis because they had significant impact on medical spending and might not be on the causal pathway between diabetes and spending, which made them good candidates for risk adjustment. However, there is also evidence that healthcare spending for many chronic conditions that are unrelated to diabetes tend to increase diabetes costs owing to, for example, longer hospital stays, when a patient has diabetes. Following the previous methodology, we did not use the cost ratios that controlled for additional comorbidities in our final estimates.

The cost data are highly skewed and include many nonusers of the healthcare system with zero spending as well as users with high spending. We used two-part models that included a logit model in the first part and a GLM with a log-link and gamma distribution in the second part. This model was selected after examining the distribution of cost variables, looking at model goodness-of-fit statistics, as well as analyzing the results of the family link test.

$$RR_{a,s,p,t} = \frac{E_{a,s,p,t}(DM=1|DM)}{E_{a,s,p,t}(DM=0|DM)} \tag{6}$$

where RR is the cost ratio for individuals in age group a, sex s, payer p, and, if applicable, service type t; $E_{a,s,p,t}(DM=1|DM)$ is the expected expenditures for people with diabetes; and $E_{a,s,p,t}(DM=0|DM)$ is the expected expenditures for people with diabetes under the counterfactual where they do not have diabetes. This ratio compares the expenditures of people with diabetes to what the expenditures would have been if these people did not have diabetes.

2.2.2.4 Calculating Medical Costs, Except Nursing Home Costs, Attributable to Diabetes

Using diabetes prevalence at the state level by payer, sex, and age group described in Section 2.2.2.2 and the cost ratios at the national level by payer, sex, age group, and type of service (except Medicaid) described in Section 2.2.2.3, we calculated the AF for each payer, type of service (all services combined for Medicaid), sex, and age group. This analysis excluded nursing home costs.

As described in Section 2.2.2, we used the following AF formula to estimate diabetes-attributable medical expenditures (Equation 7).

$$AF = \frac{pd_j \times (RR_j - 1)}{1 + pd_j \times (RR_j - 1)} \tag{7}$$

where AF is the AF for diabetes, RR is the diabetes cost ratio, and pd is the state prevalence of diabetes. The subscript j indicates payer, sex, age group, and service type.

Equation 7 can be rewritten as follows:

$$AF = \frac{pd_j \times (RR_j - 1)}{1 + pd_j \times RR_j - pd_j} \tag{8}$$

$$= \frac{pd_j \times (RR_j - 1)}{pd_j \times RR_j - pd_j + 1}$$

$$= \frac{pd_j \times (RR_j - 1)}{pd_j \times (RR_j - 1) + pd_j + (1 - pd_j)} \tag{9}$$

Now Equation 9 can be rewritten as Equation 10 by introducing the cost concept to calculate the total attributable cost Y_j:

$$Y_j = \frac{(pd_j \times \ RR_j - 1) \times C_{0DM}}{pd_j \times (RR_j - 1) \times C_{0DM} + pd_j \times C_{0DM} + (1 - pd_j) \times C_{noDM}} \tag{10}$$

where C_{0DM} is the per-person spending for a person with diabetes under the counterfactual that they do not have diabetes, and C_{noDM} is the per-person spending for a person who does not have diabetes. The first term in the denominator is the diabetes-attributable costs for persons with diabetes, the second term is the "regular" non-diabetes-attributable costs for persons with diabetes, and the last term is the costs for persons without diabetes.

A subtle, implicit, and important assumption in the usual formula for AF (formula 7) is that $C_{0DM} = C_{noDM}$. However, in the way we calculated C_{0DM} as described in Section 2.2.1.4, it is different from C_{noDM}. Hence, the implicit assumption would be violated if we were to use the usual AF formula directly, which can lead to either overestimates or underestimates of costs. Therefore, we adjusted the usual AF formula to account for this.

The cost ratio between C_{0DM} and C_{noDM} can be defined as follows: $1+\varphi = C_{0DM} \div C_{noDM}$, meaning that $C_{0DM} = C_{noDM} \times (1+\varphi)$. Now Equation 10 can be rewritten as follows:

$$Y_j = \frac{pd_j \times (RR_j - 1) \times C_{noDM} \times (1+\varphi)}{pd_j \times (RR_j - 1) \times C_{noDM} \times (1+\varphi) + pd_j \times C_{noDM} \times (1+\varphi) + (1 - pd_j) \times C_{noDM}} \quad (11)$$

Canceling off the C_{noDM} term in both the numerator and the denominator, Equation 11 becomes

$$AF_j = \frac{pd_j \times (RR_j - 1) \times (1+\varphi)}{pd_j \times (RR_j - 1) \times (1+\varphi) + pd_j \times (1+\varphi) + (1 - pd_j)}$$

$$= \frac{pd_j \times (RR_j - 1) \times (1+\varphi)}{pd_j \times RR_j \times (1+\varphi) - pd_j \times (1+\varphi) + pd_j \times (1+\varphi) + (1 - pd_j)}$$

$$= \frac{pd_j \times (RR_j - 1) \times (1+\varphi)}{pd_j \times RR_j \times (1+\varphi) + (1 - pd_j)} \quad (12)$$

We applied this adjusted AF Equation 12 to the aggregated state expenditure estimates described in Section 2.2.2.1 to calculate diabetes-attributable cost. We calculated diabetes attributable direct medical costs by state, payer, types of service (all services combined for Medicaid), age group (19 to 64 and 65 or older), and sex for 2013.

2.2.2.5 Nursing Home Costs

We used the CMS MDS and the estimates of state nursing home expenditures by age group, sex, and payer described in Section 2.2.2.1 to estimate state-level diabetes-attributable nursing home costs by age group, sex, and payer.

Calculate the Diabetes AF for Nursing Home Costs. We first calculated the AF for nursing home costs by age and sex as the excess diabetes prevalence in nursing homes compared to the community, as shown in Equation 13:

$$\text{AF} = \left[\frac{N^D * RUG^D}{N^D * RUG^D + N^N * RUG^N} - C^D \right] \quad (13)$$

where AF represents the excess diabetes prevalence in nursing homes compared with the community, N^D is the number of nursing home residents with diabetes, N^N is the number of nursing home residents without diabetes, RUG^D is the average Resource Utilization Group (RUG) payment for nursing home residents with diabetes, RUG^N is the average RUG payment for residents without diabetes, and C^D is the prevalence of diabetes in the community. The number of nursing home residents by diabetes status was estimated from the MDS data using a data reference period of October 2018 to September 2019. In calculating the RUG-weighted AF in Equation 13, we included only long-term stay residents (residents with nursing home episodes of at least 100 days). Episodes were defined according to the MDS User Manual’s definition and can span multiple nursing home stays that may be separated by brief time intervals where the resident is discharged (CMS, 2015).

We weighted the number of nursing home residents by the mean RUG payments over the same reference period to capture the higher potential cost of people with diabetes. The estimates of diabetes prevalence in the community are from BRFSS 2021, and we use the same estimation approach described in Section 2.1.2.

2.2.3 Indirect Cost of Diabetes

This section of the burden toolkit reports diabetes-attributable indirect costs and consists of costs of absenteeism, presenteeism, household productivity losses, inability to work, and premature mortality. In this section, we describe the methods used to estimate each component of the indirect costs of diabetes.

2.2.3.1 Total Costs

We calculated total morbidity costs attributable to diabetes as a sum of diabetes-attributable costs related to work absenteeism, work presenteeism, household productivity losses, and inability to work, as described below. We then calculated total indirect costs attributable to diabetes as a sum of diabetes-attributable morbidity and mortality costs. We calculated per capita costs as the cost per person with diabetes, where the number of people with diabetes includes the noninstitutionalized general population from BRFSS and nursing home residents.

2.2.3.2 Absenteeism Costs

Absenteeism cost is the cost of workdays lost. To estimate the diabetes-attributable absenteeism costs among those who are currently employed, we first estimated the number of workdays missed that are attributable to diabetes. We estimated the diabetes-attributable workdays missed per person with diabetes by Census region, age group, and sex. We then valued these days missed using national age group- and sex-specific earnings adjusted to the state level using a state-to-national adjustment factor. We next multiplied the value of the workdays missed by the estimated number of employed workers with diabetes in each state, by age group, and by sex. The steps below provide additional details on our approach.

Step 1: Estimated work loss days attributable to diabetes. We used the National Health Interview Survey (NHIS) to estimate the number of work loss days attributable to diabetes. NHIS is a cross-sectional household interview survey administered by CDC, which is designed to monitor the health of the U.S. population through the collection and analysis of data on a broad range of health topics. The survey covers the civilian noninstitutionalized population residing in the United States at the time of the interview.

Pooling data from the 2016 through 2021 NHIS, we estimated work loss at the regional (Census region) and national levels. We used regional estimates instead of state-specific estimates because person-level state identifiers are not included on the NHIS public use data files. In NHIS, persons with diabetes are identified by the question "Have you ever been told that you had diabetes?" The work-loss analysis was restricted to individuals employed at any point during the year. Number of workdays lost was defined using the following NHIS question: "During the past 12 months, about how many days did you miss work at a job or business because of illness or injury (do not include maternity leave)?" To estimate workdays lost due to diabetes, we tested three different models for best fit: one-part negative binomial model, two-part generalized linear model with a logit, and a zero-inflated negative binomial model. Based on a comparison of the model residuals, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC), our final estimation used a two-part model with a logit model for the first part and a GLM for the second part:

$$\text{Missed\_Work}_0 = f(\text{Diab}_t, \text{Comorbid}_t, \text{Region}_t, X_t), \tag{14}$$

$$\text{Missed\_Work}_t = f(\text{Diab}_t, \text{Comorbid}_t, \text{Region}_t, X_t), \tag{15}$$

where Missed_Work_0 indicates whether a workday was missed in the past year due to illness or injury; Missed_Work_t denotes the annual number of workdays missed because of illness or injury if at least 1 workday was missed; Diab_t denotes whether the person has diabetes, Comorbid_t represents the presence of other comorbidities; Region_t represents region of residence (Census region), and X_t denotes sociodemographic characteristics. We used a gamma distribution and log link to model the number of missed workdays for those with nonzero missed workdays. We controlled for the following comorbidities: arthritis, asthma, cancer, depression, chronic bronchitis, back problems, and pregnancy. To capture work loss attributable to diabetes and its complications, we did not control for diabetes risk factors and complications, such as CHF, CHD, other heart diseases, hypertension, renal failure, stroke, and high cholesterol. We also included the following sociodemographic controls: age, age squared, race/ethnicity, education, family income, health insurance, and occupation. The occupation variable was unavailable in the 2019 NHIS data. We estimated per-person number of workdays missed, calculated by age, sex, and region, as the mean difference between the predicted number of workdays missed for a person with diabetes and the predicted number of workdays missed for that person, assuming no diabetes. Predicted values were estimated using coefficients from both the logit and GLM models. Productivity losses for employed individuals on short-term disability are captured in this portion of the analysis.

Step 2: Obtained earnings estimates by age group/sex/state. Earnings data were not available by all three stratifications (age group/sex/state), so we used a two-step approach to convert national wage estimates by age group/sex to state-level estimates by age group/sex. First, we obtained mean per-capita earnings by age group and sex from the Current Population Survey (CPS) Microdata Access Tool (MDAT) to estimate daily earnings at the national level. We used 2022 CPS annual earnings data, which reflect annual earnings from 2021 and include income from wage and salary earnings amount. Average earnings were estimated by 5-year age groups and sex and included working individuals.

Second, we used 2021 national and state-level occupational employment mean wage estimates from the Bureau of Labor Statistics (BLS) to estimate a state-to-national wage ratio. BLS estimates are collected from employers and provide occupation-level wages by state, but they are not available by age and sex. We applied the BLS state-to-national wage ratios to the CPS national wages by age and sex to obtain state-level wage estimates by age and sex. We weighted the 5-year age groups from the CPS annual earnings data to our age groups (18–44, 45–64, 65–74) using 2021 Census population estimates. We calculated mean earnings per day of work by dividing annual mean earnings by 250, which is the typical number of weekdays worked per year for full-time employees.

Step 3: Calculated per-capita diabetes-attributable absenteeism costs. We calculated state-level per-capita diabetes-attributable absenteeism costs by age and sex ($Per\_Cap\_Missed\_Work\_Cost_{sag}$) as follows:

$$Per\_Cap\_Missed\_Work\_Cost_{sag} = Work\_loss\_Diab_{rag} * Daily\_Earn_{sag}, \qquad (16)$$

where $Work\_loss\_Diab_{rag}$ represents the number of workdays lost attributed to diabetes by region, age, and sex (from the NHIS analysis); and $Daily\_Earn_{sag}$ represents state-level age group- and sex-specific average daily earnings.

Step 4: Estimated the number of people with diabetes who are employed. We used 2016 through 2021 NHIS data to calculate the percentages of people with diabetes who were employed by region, age group, and sex. We identified employed individuals using the same methodology as in Step 1. We estimated the number of employed people with diabetes in each state, by age and sex ($Diab\_work_{sag}$), as follows:

$$Diab\_work_{sag} = Perc\_employ|diab_{rag} * Num\_diab_{sag}, \qquad (17)$$

where $Perc\_employ|diab_{sag}$ denotes the region, age group-, and sex-specific percentage of people with diabetes who are employed, as estimated from NHIS. Num_diab_{sag} represents the number of people with diabetes by state, age group, and sex, which we obtained from the Health Burden section of the toolkit.

Step 5: Calculated total absenteeism costs. Our final step was to calculate total absenteeism costs by age group and sex for each state, as follows:

$$Absenteeism\_tot_{sag} = Per\_Cap\_Missed\_Work\_Cost_{sag} * Diab\_work_{sag} \qquad (18)$$

For this calculation, we multiplied the per-capita cost of missed work attributable to diabetes ($Per\_Cap\_Missed\_Work\_Cost_{sag}$) by an estimate of the total number of people in the state, by age and sex, who have diabetes and are employed ($Diab\_work_{sag}$).

In the toolkit, we report per capita annual work absenteeism costs calculated as cost per employed person with diabetes and as cost per person with diabetes, where persons with diabetes include the noninstitutionalized population from BRFSS and the nursing home residents.

2.2.3.3 Presenteeism Costs

Presenteeism cost is the cost of productivity losses while on the job. To estimate the costs of reduced productivity while at work, we used published estimates of the impact of diabetes on reducing productivity. In the American Diabetes Association (ADA) report, the authors assumed that, on average, 6.6% of annual productivity is lost as a result of diabetes while people are at work (ADA, 2018). We multiplied this reduced productivity estimate by state-level daily earnings by age group and sex ($Daily\_Earn\_per\_cap_{sag;}$ estimated in Step 2 in Section 2.2.3.2) and then applied it to the average number of days worked by employed people with diabetes minus the number of days missed by people with diabetes, as follows:

$$Present\_Cost\_per\_cap_{sag} = 0.066 * Daily\_Earn\_per\_cap_{sag} * (250 - Days\_Missed\_Work_{ag}) \quad (19)$$

Daily_Earn_per_cap denotes daily CPS earnings data (annual earnings divided by 250 days), which are the same data we used to estimate absenteeism costs, thus consistently valuing productivity losses from absenteeism and presenteeism. $Days\_Missed\_Work_{ag}$ is the average number of days of work loss among people with diabetes by age group and sex.

Because presenteeism costs apply only to employed people with diabetes, we used data on employment among people with diabetes to estimate total state costs of presenteeism. For each state, age group, and sex, we multiplied per capita presenteeism costs by the estimated number of employed individuals with diabetes from NHIS 2016 – 2021. We identified employed individuals using the same methodology as in Step 1 in Section 2.2.3.2 (those who had a job in the last week or no job in the last week but a job in the past 12 months were identified as employed). In the toolkit, we report per capita annual work presenteeism costs calculated as cost per employed person with diabetes and as cost per person with diabetes, where persons with diabetes include the noninstitutionalized general population from BRFSS and the nursing home residents.

2.2.3.4 Household Productivity Losses

Household productivity losses arise when people are unable to perform household services. Although our absenteeism costs value lost market production due to diabetes, these estimates do not value lost non-market production due to diabetes. We estimated household production losses using the number of days spent in bed attributable to diabetes to value non-market production lost due to diabetes, such as housework, food cooking and clean-up, household management, caring for children in the household, etc. To estimate the number of bed days attributable to diabetes, we used 2014 – 2018 NHIS data and the same methodology that we used to estimate the number of workdays lost for the absenteeism cost analysis. The bed days variable was not available in the NHIS data after 2018, so we pooled 5 years of data prior to 2019 for this analysis. Because both employed and unemployed individuals may experience bed days, the bed days analysis included all respondents aged 18 or older, regardless of whether they were employed. For an employed individual, a sick day spent in bed would result in losses in both labor and household productivity. Consequently, valuing both labor and household productivity losses for a missed workday spent in bed did not result in double counting of costs.

Step 1: Estimated bed days attributable to diabetes. We defined number of days spent in bed using valid responses to the following NHIS question: "During the past 12 months, about how many days did illness or injury keep you in bed more than half of the day? (include days while an overnight stay as an inpatient in a hospital)." Using the 2014 – 2018 NHIS data, we tested three different regression models for best fit: one-part negative binomial, two-part generalized linear model with a logit, and a zero-inflated negative binomial model. Based on a comparison of the model residuals, the AIC, and the BIC, our final bed days estimation used a two-part model with a logit model in the first part and GLM in the second:

$$\text{Bed\_Days}_0 = f(\text{Diab}_t, \text{Comorbid}_t, \text{Region}_t, X_t), \tag{20}$$

$$\text{Bed\_Days}_t = f\,(\text{Diab}_t, \text{Comorbid}_t, \text{Region}_t, X_t), \tag{21}$$

where Bed_Days_0 indicates whether at least 1 bed day was reported; Bed_Days_t denotes the annual number of bed days reported if at least one day was spend in bed; Diab_t represents whether the individual has diabetes; Comorbid_t stands for the presence of other comorbidities; Region_t is region of residence (Census region); and X_t represents other sociodemographic characteristics. We used a gamma distribution and log link to model the number of bed days for those with nonzero bed days. We controlled for the following comorbidities: arthritis, asthma, cancer, depression, chronic bronchitis, back problems, and pregnancy. To capture the downstream effects of diabetes, we did not control for diabetes risk factors and complications, such as CHF, CHD, other heart diseases, hypertension, renal failure, stroke, and high cholesterol. We also included the following sociodemographic characteristics: age, age squared, race/ethnicity, education, family income, health insurance, and employment status. We estimated per person number of days spent in bed, calculated by age and sex at the regional level, as the mean difference between the predicted number of bed days for each person with diabetes and the predicted number of bed days assuming no diabetes. Predicted values were estimated using coefficients from the GLM and logit models.

Step 2: Valued a lost day of household production. We obtained an estimate of the average per capita monetary value of a day of household production by age group and sex from the Expectancy Data Economic Demographers' "Dollar Value of a Day, 2020" publication (Expectancy Data, 2021) and inflated it to 2021 dollars. That report provides a market estimate of the value of a day for various activities, including household production and caring for and helping others in the household, such as inside housework, food cooking and clean-up, shopping, and household management. The estimates are based on time-diary data from BLS' American Time Use Survey, combined with data from a wage survey conducted by BLS. These value-of-time estimates are available at the national level only. We adjusted the age group and sex-specific estimates to state estimates by creating state multipliers using BLS 2021 average wages for each state, by age group and sex, as a ratio of average national wages, by age group and sex (see Step 2 in Section 2.2.3.2 for further details about the state multipliers).

Step 3: Calculated per-capita diabetes-attributable household productivity costs. We then calculated state-level per-capita diabetes-attributable household productivity losses by age group and sex ($HH\_prod\_loss\_PC_{sag}$) as follows:

$$HH\_prod\_loss\_PC_{sag} = Bed\_days\_diab_{rag} * HH\_daily\_value_{sag}, \tag{22}$$

where $Bed\_days\_diab_{rag}$ represents the estimated per capita number of bed days attributable to diabetes by region, age group, and sex; and $HH\_daily\_value_{sag}$ denotes the state-level value of a day of household production and caring for and helping others in the household, by age group and sex.

Step 4: Calculated the cost for the total household productivity loss. We calculated total household productivity losses by age group and sex for each state ($HH\_prod\_loss\_tot_{sag}$) as follows:

$$HH\_prod\_loss\_tot_{sag} = HH\_prod\_loss\_PC_{sag} * Num\_diab_{sag}. \tag{23}$$

In Equation 23, $HH\_prod\_loss\_PC_{sag}$ denotes per capita state-level diabetes-attributable household productivity losses by age group and sex. Num_diab_{sag} is the estimated number of people with diabetes by age group (a) and sex (g) among the noninstitutionalized in each state (s), which we obtained from the Health Burden section of the toolkit. In the toolkit, the per person household productivity costs are reported per person with diabetes, where persons with diabetes include the noninstitutionalized general population from BRFSS and the nursing home residents.

2.2.3.5 Inability to Work Costs

Inability to work costs arise when people are disabled and unable to work. If people are too sick to work because of diabetes, they lose the full value of their expected earnings over the course of a year. We assume that these disabled, unemployed individuals would have been employed if they did not have severe diabetes causing them being unable to work. Our approach to estimating these losses involves first estimating the probability of being unable to work because of diabetes by region, age group, and sex; then applying this probability to state estimates of the number of people with diabetes by age group and sex, and finally assessing the value work loss for those unable to work using state-, age group-, and sex-specific annual earnings data. We include estimates for the noninstitutionalized population only.

Step 1: The estimated probability of being unable to work attributable to diabetes. We estimated the probability of being unable to work because of diabetes using the 2016 – 2021 NHIS data. We defined a person as being unable to work if he or she answered "Disabled" to the NHIS survey question "What is the main reason you did not work last week?" We estimated the probability of being unable to work because of diabetes at the national level by region, age group, and sex.

We used a logistic regression model as follows:

$$\text{Unable\_to\_work}_t = f(\text{Diab}_t, X_t), \tag{24}$$

where Unable_to_work_t represents whether an individual reports being unable to work because of a health condition or not; Diab_t denotes whether the individual has diabetes; and X_t represents demographic variables such as age, sex, and comorbidities (e.g., arthritis, COPD). We used coefficients from the model to estimate the mean difference in the predicted probability of being unable to work for someone with diabetes relative to their predicted probability of being unable to work if they did not have diabetes. We estimated the probability of being unable to work due to diabetes by region, age group, and sex, denoted as $\text{Pr\_unable\_to\_work}_{rag}$.

Step 2: Estimated the number of people with diabetes who are unable to work. We multiplied the estimated probability of being unable to work because of diabetes (by age/sex) by the number of people with diabetes by state, age group, and sex. This calculation resulted in an estimate of the number of people unable to work because of diabetes ($\text{Num\_unable\_to\_work}_{sag}$) as follows:

$$\text{Num\_unable\_to\_work}_{sag} = \text{Pr\_unable\_to\_work}_{rag} * \text{Num\_diab}_{sag}. \tag{25}$$

The estimated number of people with diabetes, Num_diab_{sag}, was from Section 2.1.2.

Step 3: Calculated the total cost of inability to work costs. We multiplied the number of people unable to work because of diabetes in each state by state-level mean annual earnings by age group and sex (estimated in Step 2 under Absenteeism Costs section) as follows:

$$\text{Unable\_to\_Work\_Diab\_Cost}_{sag} = \text{Num\_unable\_to\_work}_{sag}\_ * \text{Annual\_Earn}_{sag}, \quad (26)$$

where $\text{Unable\_to\_Work\_Diab\_Cost}_{sag}$ represents total state-, age group-, and sex-specific costs that arise when people with diabetes are too sick to work; and Annual_Earn_{sag} denotes state-, age group-, and sex-specific annual earnings estimates from the CPS. These are the same earnings estimates that we used to value absenteeism and presenteeism costs for employed people with diabetes (see Step 2 in Section 2.2.3.2 for further details about the state-level earnings estimates).

In the toolkit, we report per capita annual costs of inability to work calculated as cost per person with diabetes who is unable to work and as cost per person with diabetes.

2.2.3.6 Mortality Costs

We estimated mortality costs using a human capital approach, which values premature death from a disease as future productivity losses foregone (Haddix, Teutsch, and Corso, 2003; Rice, Hodgson, and Kopstein, 1985; Rice, 1966). Our diabetes-attributable mortality cost estimates provide separate estimates for the value of labor productivity losses and the value of household productivity losses resulting from premature mortality. We used the number of deaths attributable to diabetes by age group and sex in each state estimated in Section 2.3.1 and multiplied those estimates by estimates of the present value of lifetime earnings and household productivity costs to calculate total mortality costs.

We estimated labor losses due to premature mortality for adults aged 18 to 74 and household production losses due to premature mortality for adults aged 18 to 84. We did not calculate labor costs associated with premature mortality for adults aged 75 or older to be consistent with other labor loss estimates (absenteeism and presenteeism costs). We did not calculate household productivity losses due to premature mortality for adults aged 85 or older because we assumed that participation in household activities among this group is low. For the mortality cost analysis, we used finer age categories than in other sections of the indirect cost estimation to better capture the distribution of deaths within age groups and therefore more accurately assign estimates of lost earnings or household productivity. The finer age groups were 18 to 34, 35 to 44, 45 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74, 75 to 79, and 80 to 84 years. We then aggregated the mortality cost estimates into the standard age groups used for the rest of the indirect cost estimates (18 to 44, 45 to 64, 65 to 74, and 75 to 84).

Step 1: Calculated lifetime earnings and lifetime household production costs. We estimated the present value of future earnings and household production using national estimates of annual earnings and the dollar value of household production that we used to value work loss and household production losses (described in Steps 2 of Sections 2.2.3.2 and 2.2.3.4). We then adjusted these present-value estimates to state estimates by multiplying them by the ratio of state-to-national wages that we used for the morbidity-related cost estimates. Future costs were discounted by the probability of surviving to each year of age at which the expected production occurs. We used the 2021 U.S. life tables from the National Vital Statistics Report to calculate compounded survival rates for each age group (Arias, 2022b). To ensure that losses were applied only to the populations expected to incur the losses, we multiplied the age group- and sex-specific labor costs for each state by age group- and sex-specific employment rates, and we multiplied age group- and sex-specific percentages of people living in households by household production losses by state, age, and sex (Haddix, Teutsch, & Corso, 2003). We also adjusted for an expected future growth in productivity using a 1% annual growth rate and discounted the costs using a 3% annual discount rate, as recommended in Haddix, Teutsch, and Corso (2003).

Step 2: Calculated total mortality costs. We calculated total mortality costs for each age/sex group by multiplying lifetime earnings and lifetime household production costs by the number of deaths attributable to diabetes (calculated in Section 2.3.1). We then aggregated the mortality costs into the standard age groups used in the rest of the indirect cost estimation section.

2.2.4 Costs by Perspective

This section of the burden toolkit reports diabetes costs from the perspective of the state Medicaid program, private insurers in the state, and all employers in the state. The purpose of these estimates is to provide different groups and organizations with estimates of costs or losses that they incur as a result of diabetes. The costs reported in this portion of the toolkit are estimates that may be useful for planning their likely expenditures, given diabetes prevalence among enrollees or employees and for assessing the potential value of investments in approaches to manage or prevent diabetes. Those who are interested in assessing the potential costs and impacts of investing in the National Diabetes Prevention Program for enrollees or employees should see the Diabetes Impact Toolkit (available from https://nccd.cdc.gov/toolkit/diabetesimpact).

2.2.4.1 Medicaid Costs

We obtained state health expenditures paid for by Medicaid (Section 2.2.2.1) from SHEA data and allocated Medicaid spending across age and sex groups. We used the state Medicaid expenditures for all healthcare service types (including nursing home costs). As described in detail in Sections 2.2.2.4 and 2.2.2.5, we used an AF approach to estimate the amount of each state’s Medicaid expenditures attributable to diabetes by age group and sex. We provide these estimates as the state Medicaid costs attributable to diabetes, showing both total costs and costs per adult with diabetes enrolled in Medicaid.

2.2.4.2 Private Insurance Costs

We estimated annual diabetes-attributable medical costs incurred by private insurers by starting with the medical costs paid by payers other than Medicare or Medicaid, including private insurers, military insurers, out-of-pocket expenditures, and other payers, as described in Section 2.2.2. We then multiplied Other Payer costs by the fraction of these costs paid by private insurers, which we calculated for each state from the SHEA data. Because expenditures from SHEA were not available by age group and sex, we assumed that the fraction of Other Payer costs paid by private insurers did not vary by age group or sex. On average, about 55% of Other Payer costs were paid by private insurers across all states in the SHEA. We did not include nursing home costs in the Other Payer costs because most private insurance costs are for the noninstitutionalized populations. Consequently, our private insurer cost estimates reflect costs incurred for the noninstitutionalized population only. We applied the state fractions of private payer costs to Other Payer costs by state, age group (19 to 64 and 65 or older), and sex to estimate total private insurer costs by state, age group, and sex.

To estimate private insurance costs per person with a private payer, we first estimated the number of privately insured people with diabetes in each state by age group (19 to 64 and 65 years or older) and sex. We used the 2021 BRFSS data to estimate the total number of people in each state with a private payer by age group and sex, as described in Section 2.2.2.2. We then estimated diabetes prevalence among the privately insured by age group and sex for each state, also using the 2021 BRFSS data. Combining the privately insured and diabetes prevalence among the privately insured estimates resulted in estimates of the number of privately insured people in each state with diabetes by age group and sex. We estimated private insurance costs per person by dividing total diabetes attributable costs paid by private payers for each state, age group, and sex by the estimated number of privately insured people with diabetes for each state, age group, and sex category.

2.2.4.3 Employer Costs

The estimated annual diabetes-attributable costs incurred by employers in each state consist of the medical costs paid by private insurers for employees with diabetes and the diabetes-attributable indirect costs of absenteeism and presenteeism, which reflect productivity losses borne by employers. The medical costs incurred by private insurers serve as a fair representation of costs for employers that are self-insured and are a proxy for other employers because even though they do not directly pay the private insurance expenditures, premiums for a given year are usually negotiated based on previous year's medical expenditures. Our approach for estimating private insurance costs is described in more detail in Section 2.2.4.2. We multiplied diabetes-attributable per-person private insurance cost estimates by the number of employees with diabetes to estimate the private insurance costs attributable to diabetes that are incurred by employers. This component of employer costs was estimated by state, age group, and sex, for all employees aged 18 to 74.

The absenteeism and presenteeism costs attributable to diabetes were drawn directly from our estimates of indirect costs of diabetes. Our methods for estimating absenteeism and presenteeism costs attributable to diabetes are described in detail in Sections 2.2.3.2 and 2.2.3.3. In sum, we estimated the number of missed workdays attributable to diabetes by region, age group, and sex and valued lost workdays using state average earnings. To estimate presenteeism costs, we estimated the annual number of hours lost while on the job because of reduced productivity attributable to diabetes. We valued these productivity losses using state average earnings. For employers' annual absenteeism and presenteeism costs attributable to diabetes, we estimated costs by age group and sex for all employees aged 18 to 74.

Our estimated employer costs attributable to diabetes reflect total costs incurred by all employers in a given state and average cost per employee with diabetes in that state.

2.3 Diabetes Mortality and Health-related Quality of Life

This section of the burden toolkit reports diabetes-related mortality statistics in each state and nationally; it consists of diabetes-attributable deaths, years of potential life lost (YPLLs), and quality-adjusted life years (QALYs) lost due to diabetes.

The following annual estimates are reported in the mortality section of the toolkit at the state and national levels among persons aged 15 or older:

1. Mortality
    a. Number of diabetes-attributable deaths, overall, by sex, by age group, and by sex/age group
        i. Diabetes as the underlying cause of death
        ii. Cause-specific deaths attributable to diabetes: all causes of death, CVD deaths, and end-stage renal disease (ESRD) deaths
    b. Diabetes-attributable deaths per 100,000 persons, overall, by sex, by age group, and by sex/age group
        i. Diabetes as the underlying cause of death
        ii. Cause-specific deaths attributable to diabetes: all causes of death, CVD deaths, and ESRD deaths
2. YPLLs, overall and by age group/sex
    a. Estimated average YPLLs attributable to diabetes
    b. Number of persons with diabetes (in thousands)

   c. Total YPLLs attributable to diabetes (in thousands)

3. QALYs lost, overall and by age group/sex

   a. Estimated average QALYs lost due to diabetes

   b. Number of persons with diabetes (in thousands)

   c. Total QALYs lost due to diabetes (in thousands)

Each component of the mortality section is described in detail in the following subsections.

2.3.1 Mortality Data

The mortality section of the toolkit reports the number and rate per 100,000 of diabetes-attributable deaths in persons aged 15 or older.[2] The mortality data are presented by four age groups (15–44, 45–64, 65–74, 75 years or older) by sex and by state using 2021 CDC WONDER mortality data (https://wonder.cdc.gov/mcd-icd10.html). CDC WONDER is a public-use online database for epidemiologic research that contains information about mortality (deaths) and census data. Death counts are automatically calculated in the CDC WONDER interface and are downloadable by cause, age, sex, and state. The toolkit reports the number and the rate of deaths with diabetes as the underlying cause of death and diabetes-attributable deaths for all causes of death, CVD, and ESRD for 780 (52*5*3) different combinations of states (51 plus the United States as a whole), age categories (4 plus overall), and sex (2 plus overall). Aggregating up to four age groups matches diabetes prevalence calculated in Section 2.1.2 and drastically reduces the percentage of suppressed or unreliable cohorts.

In CDC WONDER, mortality statistics are suppressed when $n < 10$ for any specified strata and are considered unreliable when $n < 20$; thus, we are unable to report the data for these strata in the toolkit. We developed a set of rules to aggregate the data in an effort to minimize the amount of suppressed and unreliable data at the state level. Whenever possible, our aim was to report state-level data using actual numbers of death.

We used the following rules to report results from the mortality data, from most desirable to least desirable:

1. Use 2021 state/age group/sex deaths (100% of observations with all-cause deaths are in this category, meaning that we have no unreliable or suppressed data for all-cause deaths; 97.8% of CVD deaths; and 89.0% of diabetes deaths).
2. When #1 is suppressed or unreliable, pool state data through 2019–2021 and divide by 3 to calculate an average annual death rate (0.5% of deaths with CVD as the underlying cause falls in this category).

[2] Herein we start with age 15 to 19 because CDC WONDER reports deaths in 5-year bins. However, for QALYs and YLLs, we start at age 18.

3. When #1 and #2 do not produce numbers above the reliable threshold, use 2021 regional death rates and apply to state cohort population (0.5% of CVD deaths; 1.0% of diabetes deaths).
4. If #1–3 all yield suppressed numbers, we report "suppressed" in the toolkit (0.7% of CVD deaths; 5.9% of diabetes deaths). All of these occurred in the 15-44 age group. If #1–3 all yield unreliable numbers, we report the 2021 state value (#1), but note it as unreliable (0.5% of CVD deaths; 4.2% of diabetes deaths). Eleven of the 17 of these cases occurred in the 15-44 age group.

Mortality increased significantly between 2019 and 2020. Based on the data from the CDC WONDER, rates of all-cause mortality increased by 18% between 2019 and 2020, rates of deaths with diabetes as the underlying cause of death increased by 16%, and rates of CVD as the underlying cause of death increased by 5%. Mortality rates then increased slightly (2%) between 2020 and 2021. According to the NCHS, deaths from COVID-19 accounted for 61% of the increase in mortality between 2019 and 2020 and for 50% of the increase between 2020 and 2021 (Arias et al, 2022a, 2022b). As a result, mortality data used in our analysis reflect the increased deaths due to the COVID-19 pandemic and other causes.

2.2.1.2 Attributable Fraction of Diabetes

Because diabetes is not always listed as a cause of death on death certificates, diabetes-attributable mortality from all-cause and CVD was calculated using the AF approach. To do so, we used Miettinen's formula (1974) presented in Equation 27:

$$AF = \frac{pd \times (RR_i - 1)}{1 + pd \times (RR_i - 1)} \tag{27}$$

We used diabetes prevalence, pd, from BRFSS 2021, stratified by age group, sex, and state. RR_i is the adjusted RR of disease *i* in the diabetes subsample relative to the non-diabetes subsample. We applied this AF formula for all-cause and CVD mortality attributable to diabetes because CVD is a key diabetes-related causes of death (i). The diabetes-attributable mortality from all specific causes (including ESRD) would approximately sum to the value calculated using the AF from all-cause mortality, assuming that the RRs of specific causes outside our analysis (e.g., accidental deaths, cancer deaths) are always equal to 1. However, the RR=1 condition may not hold. The inclusion of diabetes-attributable deaths for causes other than diabetes potentially overcomes the concern that diabetes may not be listed as a principal (underlying) cause of death on death certificates.

Similarly to the medical costs attributable to diabetes, we could not apply the modified AF formula presented in Equation 1 to estimate diabetes-attributable mortality even though it is recommended when the RR was adjusted for confounding (Rockhill et al., 1998). Because the presence of diabetes is underreported on death certificates, we do not have accurate measures of diabetes prevalence among people with all-cause or CVD deaths. No source of diabetes prevalence data conditional on death at the state level is available. Therefore, we cannot use the modified AF formula to estimate AF for diabetes mortality. In the absence of this information, the AF formula presented in Equation 27 (also shown in Equations 5 and 7) is more appropriate (Steenland & Armstrong, 2006).

We partly avoid the problem of confounding by stratifying the RR calculation by age group and sex. This is potentially important because the prevalence of diabetes increases with advancing age, RRs decrease with age, and overall deaths increase with age. Still, some concerns about confounding may remain because RR estimates are controlled for race/ethnicity. However, further stratification of mortality data is problematic as the count of reliable numbers of deaths, per strata, especially for the younger cohorts, markedly smaller as the number of strata increases.

To calculate RR, we used NHIS data and approach described by Gregg et al. (2012) but included more recent NHIS base years (2013-2017) and follow-up (using mortality data up to 2019). We estimated the RRs stratified by age group (18-44, 45–64, 65–74, and 75 years or older) and by sex. Due to the small number of deaths with CVD among people with diabetes in the 18-44 age group, we combined males and females and estimated one RR for CVD deaths for this age group for both sexes. We used a GLM with a Poisson family and a log link and controlling for age and race/ethnicity.

We computed separate mortality rates for all causes and CVD (Table 2-6). Information from the stratification exercise confirmed that the RR for all-cause mortality declined with age.

Table 2-6. Relative Risk Using 2013-2017 Mortality Data

	Age Group	All-Cause Mortality	Cardiovascular Disease Mortality
Sex	**Corresponding ICD-10**	**All**	**I00–I09, I11, I13, I20–I51, I60–I69**
Male	Age 18–44	4.28	3.87
Male	Age 45–64	1.95	2.86
Male	Age 65–74	1.54	1.75
Male	Age 75+	1.43	1.63
Female	Age 18–44	2.77	3.87
Female	Age 45–64	2.36	4.10
Female	Age 65–74	1.94	2.55
Female	Age 75+	1.32	1.22

Source: Relative risks: (diabetes vs no diabetes), by age group/sex.

When reporting the number of estimated diabetes-attributable deaths for all causes and CVD in the toolkit, we multiplied the AF by the total number of deaths from all causes and CVD deaths and rounded the estimate to the nearest 10.

Death certificates provide information on both the immediate cause of death ("the final disease, injury, or complication directly causing death") and the underlying cause of death ("the disease or injury that initiated the chain of morbid events that led directly and inevitably to death") (CDC, 2016). However, diabetes is under-diagnosed and under-reported as an underlying cause of death among adults because (a) diabetes is often not mentioned on death certificates even among persons known to have diabetes, and (b) it is difficult to know whether diabetes caused the fatal outcome, or diabetes was a contributing factor to death (Geiss, Herman, & Smith, 1995). The attributable mortality approach attempts to estimate the number of deaths attributable to diabetes by combining information on the prevalence of diabetes, the RR of death for persons with diabetes relative to persons without diabetes, and the total number of deaths in the entire population.

We also report separately the number of persons with diabetes listed as the underlying cause of death on their death certificates. These are downloadable from CDC WONDER. As noted above, this number underestimates the number of deaths due to diabetes. Nonetheless, it is a number regularly reported by NCHS and can be viewed as a conservative lower bound estimate of the number of deaths due to diabetes. It can also be interpreted within the AF approach where the AF is 1; that is, diabetes is the true cause of death for anyone reported to have diabetes as the underlying cause of death.

For ESRD, we used the 2020 data from the United States Renal Data System (USRDS) (https://www.usrds.org/) to report mortality for individuals diagnosed with diabetes. The USRDS is a national data system that collects, analyzes, and distributes information about CKD and ESRD in the United States. In the toolkit, we report the number of deaths from ESRD reported in death certificates among those with diabetes. This assumes that all deaths from ESRD in this subpopulation are attributable to diabetes. Technically, we are using the AF approach for diabetes and ESRD, but we are assuming that the AF = 1. Similarly to the CDC WONDER mortality data, statistics for the number of deaths with ESRD are suppressed when $n < 10$ for the specified strata.

Because we independently estimated all-cause deaths and the deaths from the three specific causes, there is no guarantee that the estimates will satisfy the following condition:

$$\textit{CVD deaths} + \textit{diabetes underlying cause} + \textit{ESRD deaths} < \textit{all-cause deaths}$$

Mortality estimates used data from four different sources (BRFSS for diabetes prevalence, NHIS for RR, CDC WONDER for deaths, and USRDS for ESRD deaths). Two of the three inputs that go into the AF calculation are estimates: RR and the probability of having diabetes. Because the RR is estimated using national data, rather than state data, there may be a few cases where the sum of state estimates across causes may exceed the actual number of deaths from all causes. Furthermore, while the all-cause, CVD, and diabetes deaths were obtained from 2021 CDC WONDER data, diabetes mortality RRs were calculated based on 2019 data, and ESRD deaths were from the year 2020. Therefore, it is not surprising that the all-cause mortality was less than the sum of mortality for the three specific causes in a small number of cases (59 out of 780). These are outlined in detail in Appendix A.

2.3.2 Years of Potential Life Lost (YPLLs)

YPLLs due to diabetes measure the number of premature deaths due to diabetes. YPLL due to diabetes is an indicator of premature mortality and is calculated by multiplying the number of deaths due to diabetes by the difference in life expectancy between people with and without diabetes. Using the life table approach, we estimated all-cause mortality rates by age and sex and generated a cause-specific life table for diabetes. The cause-specific life table was constructed using prevalence of diabetes by age (5-year bins) and sex (see Appendix Table A-3); all-cause mortality values from NCHS by single year of age and sex; and national-level RR of mortality (Table 2-6) for those with and without diabetes.

YPLL estimates were calculated from the number of deaths for individuals with and without diabetes and the life expectancy at the age at which death occurs, using Pharaoh and Hollingworth's (1996) method for scaling all-cause mortality of those with diabetes relative to those without diabetes. The scale-up factor, θu, takes into account the RR (*r)* and diabetes prevalence (*p*) within the given population:

$$\theta_u = \frac{r}{pr+(1-p)}$$

The scale-up factor ranges between r and 1. When p value is close to zero, θ_u will approximately be equal to r.

The corresponding scale-down factor (θ_d) for mortality for persons without diabetes is

$$\theta_d = \frac{1}{pr+(1-p)}.$$

Using the life table approach, we estimated all-cause mortality rates by age and sex and generated a cause-specific life table for persons with diabetes. The cause elimination life table was constructed from the death rates (the number of deaths per 100,000) by using prevalence of diabetes by state, by age (5-year bins), and by sex; all-cause mortality values from NCHS by state, age, and sex; and national-level RR of mortality for those with and without diabetes.

In our life table approach, we first obtained the probability of dying between a given age x and age x+1 (this probability is commonly denoted as qx). The information on qx for all cause conditions was obtained from the 2021 U.S. life tables produced by NCHS (Arias et al, 2022b).

qx = Number dying between age x and age x+n / number attaining exact age x.

We then estimated the number of person-years lived (denoted as L_x) between age x and x+t, assuming that deaths are evenly distributed, as follows:

$$\text{Number of person years lived } (L_x) = [(\text{Time Interval})/2] * (\text{Number of Persons Alive Age } x = t) \quad (28)$$

Assuming a cohort of 100,000 births, we calculated the total number of person-years that would be lived/being alive after the beginning of the indicated age interval by cumulating the ${}_nL_x$ column from the oldest to the youngest age. The average remaining lifetime (in years) for a person who survives to the beginning of the indicated age interval was calculated by dividing the total number of person-years lived from age x (Tx) by the number of persons alive at age x (lx) (i.e., ex = Tx/lx). For deaths that occurred within the age interval x and x+n, the crude expected YPLL equals the longest life expectancy for each cohort in the absence of diabetes minus the life expectancy with the condition. YPLLs due to diabetes is then averaged across each age group. Total and 18+ estimates represent the weighted average of the age/sex group estimates, where the weights represent the relative share of persons with diabetes accounted for by each sex and age group by state. Because the prevalence estimates by age are not available for the same level of granularity as the life tables (single year of age intervals), we assume the same weight (=1) to each age in the age group (18–44, 45–64, and 65–74 age groups). For the 75+ age group, because the relative age share starts to decline after age 90, we calculated the average YPLLs (and QALYs lost) through age 89–90 only. However, although the average YPLLs for the 75+ strata only include diabetes counts from ages 75–76 through 89–90 in the calculation, the underlying YPLLs (and QALYs for each age) calculation accounts for the full age set, following standard life tables, including losses through age 100.

We conducted validation analyses to assess the impact of using 2019 and 2020 life tables (vs 2021 life tables) on our estimates. For all age groups combined, we found that the shorter life expectancy, which resulted from increased mortality in 2020 and 2021, led to higher YPLLs (increasing by less than 4% each year).

2.3.3 Quality-Adjusted Life Years Lost (QALYs)

QALYs is a measure that combines quality of life (QoL) and life expectancy. QoL is measured on a scale from 0 to 1, where 0 represents death and 1 represents optimum health. The rationale for computing QALYs is to account for mortality and morbidity by assigning patient utility values to health states and then summing utility values for each period over an appropriate time horizon (e.g., a person's remaining life expectancy). We computed QALYs using BRFSS survey data and Jia and Lubetkin's (2008) mapping to obtain preference-based values for the EuroQol five-dimensional (EQ-5D) questionnaire index, based on respondents' answers to the BRFSS Healthy Days questions. This allowed us to estimate average patient utility levels for persons with diabetes and compare that utility to persons without diabetes using BRFSS.

We estimated QALYs following three steps shown below:

1. Aggregated valid responses to the physical and mental Healthy Days (HDs) questions to obtain an overall measure of unhealthy days (UDs). Transformed them into remaining HDs in a month for each participant and aggregated values by age and sex.
2. Mapped HDs into EQ-5D values using Jia and Lubetkin's (2008) table as a reference.
3. Calculated survival probabilities by age and sex.

These three steps are outlined in detail in the sections below.

2.3.3.1 Unhealthy Days and EQ-5D

The BRFSS included the HD measures that asked respondents to report the number of days during the past 30 days when they felt physically and/or mentally unhealthy (physically unhealthy days [PUDs] and mentally unhealthy days [MUDs]). The questions are phrased as follows:

- Thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?
- Thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?

We used 2019-2021 BRFSS data for this analysis. Physical and mental HD questions were available for all 50 states and DC in the 2019-2021 BRFSS. Both questions required the respondent to answer by referring to any number between 0 and 30. The overall UD measure was calculated by adding together a respondent's PUDs and MUDs with a logical maximum value of 30 UD (formula 29) following guidance from Jia and Lubetkin (2008).

$$\text{Unhealthy Days} = \text{minimum}\ (30, PUD + MUD) \tag{29}$$

To assess the heath-related QoL, we transformed our UD estimates to EQ-5D scores. The EQ-5D is the most widely used generic preference-based measure of health-related QoL. The EQ-5D is a descriptive system covering five dimensions—mobility, self-care, usual activity, pain/discomfort, and anxiety/depression—that each has three levels: no problem, some problems, and extreme problems. We used the mapping algorithm provided by Jia and Lubetkin (2008) to translate HDs into EQ-5D scores as shown in Appendix A. We calculated HD from BRFSS UD by subtracting respondents' PUDs and MUDs from 30 days, logical maximum value of 30 HDs (Equation 30). HDs were calculated by state, age category, and sex. We used the same age categories as the Health Burden section and the Diabetes Mortality section (18–44, 45–64, 65–74, and 75 or older).

$$\text{Healthy Days} = 30 - \text{minimum}\,(30, PUD + MUD) \tag{30}$$

Total and 18+ QoL estimates represent the weighted average of the age/sex group estimates, where the weights represent the relative share of persons with diabetes accounted for by each group.

2.3.3.2 QALY and Quality-Adjusted Life Expectancy

The quality-adjusted survival estimate was obtained as follows:

$$QAS = \sum_{i=1}^{k} \frac{(Q_i + Q_{i+1})}{2} \frac{(S_i + S_{i+1})}{2} (t_{i+1} - t_i) \tag{31}$$

where Q_i is the mean QoL at time t_i, and S_i is an estimate of the survival probability at time t_i. Survival probabilities are estimated via life tables and differ by age, sex, and diabetes status. We used published national-level life expectancy for our QALY and YPLL calculations so as not to confound state-level effects in life expectancy with diabetes prevalence by state. Average QALYs lost due to diabetes is averaged across each age in the age group. Total and 18+ estimates represent the weighted average of all the age/sex group estimates, where the weights represent the relative share of persons with diabetes accounted for by each group.

References

Ackermann, R. T., Finch, E. A., Brizendine, E., Zhou, H., & Marrero, D. G. (2008). Translating the Diabetes Prevention Program into the Community. The DEPLOY Pilot Study. *American Journal of Preventive Medicine, 35*(4), 357–363. doi: 10.1016/j.amepre.2008.06.035

Agency for Healthcare Research and Quality. (2015a). Inpatient Quality Indicators #16 (IQI #16) Heart Failure Mortality Rate. AHRQ Quality Indicators™ (AHRQ QI™) ICD-9-CM and ICD-10-CM/PCS Specification Enhanced Version 5.0. Retrieved from https://qualityindicators.ahrq.gov/Downloads/Modules/IQI/V50-ICD10/TechSpecs/IQI%2016%20Heart%20Failure%20Mortality%20Rate.pdf

Agency for Healthcare Research and Quality. (2015b). Inpatient Quality Indicators #15 (IQI #15) Acute Myocardial Infarction Mortality Rate. AHRQ Quality Indicators™ (AHRQ QI™) ICD-9-CM and ICD-10-CM/PCS Specification Enhanced Version 5.0. Retrieved from https://qualityindicators.ahrq.gov/Downloads/Modules/IQI/V50-ICD10/TechSpecs/IQI%2015%20Acute%20Myocardial%20Infarction%20Mortality%20Rate.pdf

Agency for Healthcare Research and Quality. (2015c). Prevention Quality Indicators #16 (PQI #16) Lower-Extremity Amputation Among Patients with Diabetes Rate. AHRQ Quality Indicators™ (AHRQ QI™) ICD-9-CM and ICD-10-CM/PCS Specification Enhanced Version 5.0. Retrieved from https://qualityindicators.ahrq.gov/Downloads/Modules/PQI/V50-ICD10/TechSpecs/PQI%2016%20Lower-Extremity%20Amputation%20among%20Patients%20with%20Diabetes%20Rate.pdf

American Diabetes Association. (2023). Diabetes and Coronavirus (COVID-19). How COVID-19 impacts people with diabetes. Retrieved from https://diabetes.org/coronavirus-covid-19/how-coronavirus-impacts-people-with-diabetes

American Diabetes Association. (2018). Economic costs of diabetes in the US in 2017. *Diabetes Care, 41*(5), 917-928. https://doi.org/10.2337/dci18-0007

Arias, E., Xu, J. (2022a). United States life tables, 2020. *National Vital Statistics Report, 71*(1), 1–63. Retrieved from https://www.cdc.gov/nchs/data/nvsr/nvsr71/nvsr71-01.pdf

Arias, E., Xu, J. (2022b). Provisional Life Expectancy Estimates for 2021. *Vital Statistics Rapid Release Report,* 23, 1–16. Retrieved from https://www.cdc.gov/nchs/data/vsrr/vsrr023.pdf

Bernard, D., Cowan, C., Selden, T., Cai, L., Catlin, A., & Heffler, S. (2012). Reconciling medical expenditure estimates from the MEPS and NHEA, 2007. *Medicare Medicaid Research Review, 2*(4), mmrr.002.04.a09. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4006479/

Billingham, L. J., Abrams, K. R., & Jones, D. R. (1999). Methods for the analysis of quality-of-life and survival data in health technology assessment. *Health Technology Assessment, 3*(10), 1–152.

Boyle, J. P., Honeycutt, A. A., Venkat Narayan, K. M., Hoerger, T. J., Geiss, L. S., Chen, H., & Thompson, T. J. (2001). Projection of diabetes burden and cost through 2050: Impact of changing demography and disease prevalence in the U.S. *Diabetes Care, 24*, 1936–1940. doi: 10.2337/diacare.24.11.1936

Boyle, J. P., Thompson, T. J., Gregg, E. W., Barker, L. E., & Williamson, D. F. (2010). Projection of the year 2050 burden of diabetes in the US adult population: Dynamic modeling of incidence, mortality, and prediabetes prevalence. *Population Health Metrics*, *8*(1), 29. doi: 10.1186/1478-7954-8-29

Bozack, A., Millstein, S., Garcel, J.M., Kelly, K., Ruberto, R., & Weiss, L. (2014) Implementation and outcomes of the New York State YMCA diabetes prevention program: A multisite community-based translation, 2010–2012. *Preventing Chronic Disease*, *11*, E115. doi: 10.5888/pcd11.140006

Centers for Disease Control and Prevention (CDC). (2013). Awareness of prediabetes—United States, 2005–2010. *MMWR. Morbidity and Mortality Weekly Report, 62,* 209–211.

Centers for Disease Control and Prevention (CDC). (2015). *Diabetes Prevention Recognition Program standards and operating procedures*. Retrieved from https://www.cdc.gov/diabetes/prevention/pdf/dprp-standards.pdf

Centers for Disease Control and Prevention (CDC). (2016). Instructions for Completing the Cause-of-Death Section of the Death Certificate. Retrieved from https://www.cdc.gov/nchs/data/dvs/blue_form.pdf

Centers for Medicare & Medicaid Services (CMS). (2015, October). Long-Term Care Facility Resident Assessment Instrument 3.0 User’s Manual. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/MDS-30-RAI-Manual-V113.pdf

Chronic Conditions Warehouse. (2022). Impact of Transition from 27 to 30 CCW Chronic Conditions. Retrieved from: https://www2.ccwdata.org/documents/10280/19002256/ccw-condition-categories-impact-of-transition-from-27-to-30.pdf

Chronic Conditions Data Warehouse. (2023). Condition Categories. Retrieved from: https://www.ccwdata.org/web/guest/condition-categories

Dall, T. M., Halder P., Pang, B., Massoudi, M., Wintfield, N., Semilla, A., Franz, J., & Hogan P. (2014). The economic burden of elevated blood glucose levels in 2012: Diagnosed and undiagnosed diabetes, gestational diabetes mellitus, and prediabetes. *Diabetes Care, 37*, 3172–3179. doi: 10.2337/dc14-1036

De Geest, S., & Sabaté, E. (2003). Adherence to long-term therapies: Evidence for action. *European Journal of Cardiovascular Nursing*, *2*(4), 323.

Dugan J, Shubrook J. International classification of diseases, 10th revision, coding for diabetes. Clinical Diabetes. 2017 Oct 1;35(4):232-8.

Expectancy Data. (2021). *The dollar value of a day: 2020 dollar valuation*. Shawnee Mission, Kansas: Expectancy Data.

Flegal, K., Graubard, B., & Williamson, D. (2004). Methods of calculating deaths attributable to obesity. *American Journal of Epidemiology*, *160*(4), 331–338. doi: 10.1093/aje/kwh222

Ford, E., Murphy, L., Khavjou, O. A., Giles, W. H., Holt, J. B., & Croft, J. (2014). Total and state-specific medical and absenteeism costs of chronic obstructive pulmonary disease among adults aged >=18 years in the United States for 2010 and projections through 2020. *Chest.* Online First, 0–0. doi:10.1378/chest.14-0972

Garrett-Mayer, E., Parmigiani, G., Zhong, X., Cope, L., & Gabrielson, E. (2008). Cross-study validation and combined analysis of gene expression microarray data. *Biostatistics (Oxford, England)*, *9*(2), 333–354. doi: 10.1093/biostatistics/kxm033

Geiss, M. A., Herman, W. H., & Smith, P. J. (1995). Mortality in non-insulin-dependent diabetes. In *Diabetes in America* (2nd ed.). National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.

Gregg, E. W., Cheng, Y. J., Saydah, S., Cowie, C., Garfield, S., Geiss, L., & Barker, L. (2012). Trends in death rates among U.S. adults with and without diabetes between 1997 and 2006: Findings from the National Health Interview Survey. *Diabetes Care*, *35*(6), 1252–1257. https://dx.doi.org/10.2337/dc11-1162

Gregg, E. W., Zhuo, X., Cheng, Y. J., Albright, A. L., Venkat Narayan, K. M., & Thompson, T. J. (2014). Trends in lifetime risk and years of life lost due to diabetes in the USA, 1985–2011: A modelling study. *The Lancet Diabetes & Endocrinology, 2*(11), 867–874. doi: 10.1016/S2213-8587(14)70161-5

Haddix, A. C., Teutsch, S. M., & Corso, P. S. (2003). *Prevention effectiveness: A guide to decision analysis and economic evaluation* (2nd ed.). New York, NY: Oxford University Press.

Hernan, W. H., Brandle, M., Zhang, P., Williamson, D. F., Matulik, M. J., Ratner, R. E., Lachin, J. M., & Engelgau, M. M.; Diabetes Prevention Program Research Group. (2003). Costs associated with the primary prevention of type 2 diabetes mellitus in the diabetes prevention program. *Diabetes Care*, *26*(1), 36–47.

Hogan, P., Dall, T., & Nikolov, P. (2003). Economic costs of diabetes in the U.S. in 2002. *Diabetes Care*, *26*(3), 917–932. doi: 10.2337/diacare.26.3.917

Honeycutt, A. A., Boyle, J. P., Broglio, K. R., Thompson, T. J., Hoerger, T. J., Geiss, L. S., & Narayan, K. (2003). A dynamic Markov model for forecasting diabetes prevalence in the United States through 2050. *Healthcare Management Sciences, 6*(3), 155–164. doi: 10.1023/A:1024467522972

Ingram DD, Franco SJ. (2013). NCHS urban–rural classification scheme for counties. National Center for Health Statistics. Vital Health Stat 2(166). 2014. Retrieved from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf

Jia, H., & Lubetkin, E. (2008). Estimating EuroQol EQ-5D scores from Population Healthy Days data. *Medical Decision Making*, *28*(4), 491–499. doi: 10.1177/0272989X07312708

Jia, H., Zack, M. M., & Thompson, W. W. (2011a). State quality-adjusted life expectancy for U.S. adults from 1993 to 2008. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *20*(6), 853–863. doi: 10.1007/s11136-010-9826-y

Jia, H., Zack, M., Moriarty, D., & Fryback, D. (2011b). Predicting the EuroQol Group's EQ-5D Index from CDC's ''Healthy Days'' in a US Sample. *Medical Decision Making*, *31*(1), 174–185. doi: 10.1177/0272989X10364845

Jia, H., Zack, M. M., & Thompson, W. W. (2013). The effects of diabetes, hypertension, asthma, heart disease, and stroke on quality-adjusted life expectancy. *Value in Health, 16*(1), 140–147. doi: 10.1016/j.jval.2012.08.2208

Jiang, L., Manson, S. M., Dill, E. J., Beals, J., Johnson, A., Huang, H., Acton, K. J., & Roubideaux, Y. (2015). Participant and site characteristics related to participant retention in a diabetes prevention translational project. *Prevention Science*, *16*(1), 41–52. doi: 10.1007/s11121-013-0451-1

Karter AJ, Warton EM, Moffet HH, Ralston JD, Huang ES, Miller DR, Lipska KJ. Revalidation of the hypoglycemia risk stratification tool using ICD-10 codes. Diabetes Care. 2019 Apr 1;42(4): e58-9.

Klein, R. J., & Schoeborn, C. A. (2001). Age adjustment using the 2000 projected US population. *Healthy People Statistical Notes*, *20*, 1–10.

Knowler, W. C., Fowler, S. E., Hamman, R. F., Christophi, C. A., Hoffman, H. J., Brenneman, A. T., . . . Nathan, D. M., & the Diabetes Prevention Program Research Group. (2009). 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet*, *374*(9702), 1677–1686. doi: 10.1016/S0140-6736(09)61457-4

Kramer, M. K., Kriska, A. M., Venditti, E. M., Miller, R. G., Brooks, M. M., Burke, L. E., . . . Orchard, T. J. (2009). Translating the Diabetes Prevention Program: A comprehensive model for prevention training and program delivery. *American Journal of Preventive Medicine*, *37*(6), 505–511. doi: 10.1016/j.amepre.2009.07.020

Li, R., Zhang, P., Chattopadhyay, S., & Hopkins, D. (2015). *Economic evaluation of translational programs to prevent type 2 diabetes among high risk adults*. Unpublished presentation by the Centers for Disease Control and Prevention Division of Diabetes Translation.

Miettinen, O. S. (1974). Proportion of disease caused or prevented by a given exposure, trait or intervention. *American Journal of Epidemiology*, *99*(5), 325–332.

Narayan, K. M., Boyle, J. P., Geiss, L. S., Saaddine, J. B., & Thompson, T. J. (2006). Impact of recent increase in incidence on future diabetes burden: U.S., 2005–2050. *Diabetes Care*, *29*(9), 2114–2116. doi: 10.2337/dc06-1136

National Research Council (US) Panel to Advance a Research Program on the Design of National Health Accounts. (2010). Accounting for health and health care: Approaches to measuring the sources and costs of their improvement. *3, Allocating medical expenditures: A treatment-of-disease organizing framework*. Washington (DC): National Academies Press (US). Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK53341/

Parker JD, Talih M, Malec DJ, Beresovsky V, Carroll MD, Gonzalez JF, Hamilton BE, Ingram DD, Kochanek KD, McCarty F, Moriarity CL. (2017). National Center for Health Statistics data presentation standards for proportions.

Research Data Assistance Center (ResDAC). 2023. Master Beneficiary Summary File Base. Retrieved from: https://resdac.org/cms-data/files/mbsf-base

Rice DP, Hodgson TA, Kopstein AN. 1985. The Economic Costs of Illness: A Replication and Update. *Health Care Financ Rev*;7(1):61-80.

Rice DP, 1966. Estimating the Cost of Illness: Health Economics Series No. 6. PHS Pub. No. 947-6 (Washington, D.C.: U.S. Government Printing Office).

Rockhill, B., Newman, B., & Weinberg, C. (1998). Use and misuse of population attributable fractions. *American Journal of Public Health*, *88*(1), 15–19. doi: 10.2105/AJPH.88.1.15

Shrestha SS, Honeycutt AA, Yang W, Zhang P, Khavjou OA, Poehler DC, Neuwahl SJ, Hoerger TJ. Economic costs attributable to diabetes in each US state. Diabetes Care. 2018 Dec 1;41(12):2526-34.

Shultz JM, Novotny TE, Rice DP. Quantifying the disease impact of cigarette smoking with SAMMEC II software. Public health reports. 1991 May;106(3):326.

Sing, M., Banthin, J. S., Selden, T. M., Cowan, C. A., & Keehan, S. P. (2006). Reconciling medical expenditure estimates from the MEPS and NHEA, 2002. *Health Care Financing Review, 28*(1), 25–40. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194973/

Steenland, K., & Armstrong, B. (2006). An overview of methods for calculating the burden of disease due to specific risk factors. *Epidemiology*, *17*(5), 512–519. doi: 10.1097/01.ede.0000229155.05644.43

Tangka, F. K., Trogdon, J. G., Richardson, L. C., Howard, D., Sabatino, S. A., & Finkelstein, E. A. (2010). Cancer treatment cost in the United States: Has the burden shifted over time? *Cancer*, *116*(14), 3477–3484. doi: 10.1002/cncr.25150

Trogdon, J. G., Finkelstein, E. A., & Hoerger, T. J. (2008). Use of econometric models to estimate expenditure shares. *Health Services Research, 43*(4), 1442–1452. doi: 10.1111/j.1475-6773.2007.00827.x

Trogdon, J. G., Tangka, F. K., Ekwueme, D. U., Guy, G. P., Jr., Nwaise, I., & Orenstein, D. (2012). State-level projections of cancer-related medical care costs: 2010 to 2020. *American Journal of Managed Care*, *18*(9), 525–532.

Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A., Beaton, A. Z., Bittencourt, M. S., ... & American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. (2022). Heart disease and stroke statistics—2022 update: a report from the American Heart Association. Circulation, 145(8), e153-e639.

Ward, B. W. (2019). kg_nchs: a command for Korn–Graubard confidence intervals and National Center for Health Statistics' data presentation standards for proportions. *The Stata journal*, 19(3), 510-522.

Zhuo, X., Zhang, P., Baker, L., Albright, A., Thompson, T. J., & Gregg, E. (2014). The lifetime cost of diabetes and its implications for diabetes prevention. *Diabetes Care*, *37*, 2557–2564. doi: 10.2337/dc13-2484

Appendix A:
Estimating Mortality, YPLL, and QALYs

Table A-1 show all deceased cases (59 out of 780) where cardiovascular disease (CVD) deaths + diabetes underlying cause + kidney deaths were higher than all-cause deaths. Diabetes as the underlying cause of death (UCOD) and the number of deaths from kidney disease among individuals with diabetes are population-based counts. The number of deaths attributable to diabetes (all-cause deaths) and the number of CVD deaths attributable to diabetes are estimated using the attributable fraction approach and are rounded to the nearest ten. Because we also independently estimated all-cause deaths and the deaths from the three specific individual causes, using different data sources and different data years, there is no guarantee that the cause-specific estimates will be less than the all-cause deaths attributable to diabetes.

Table A-1. Cases Where Cardiovascular Disease (CVD) Deaths + Diabetes Underlying Cause + Kidney Deaths > All-Cause Deaths

State	Age	Sex	All-Cause	CVD	Diabetes	Kidney	CVD+Diabetes+ Kidney
Alaska	75+	F	50	10	32	17	59
Arkansas	15-44	F	50	10	34	17	61
Arkansas	45-64	M	700	310	279	115	704
California	65-74	M	4,480	1,490	1,730	1,637	4,857
California	75+	M	8,970	3,810	2,585	2,953	9,348
Colorado	65-74	M	490	140	216	137	493
Colorado	75+	M	750	280	265	208	753
District of Columbia	15+	M	270	110	85	77	272
District of Columbia	45-64	M	80	40	32	24	96
District of Columbia	45-64	O	170	80	51	48	179
District of Columbia	65-74	M	70	30	26	29	85
District of Columbia	65-74	O	150	60	47	53	160
District of Columbia	75+	F	80	20	33	60	113
District of Columbia	75+	M	90	40	27	64	131
Florida	15-44	F	210	20	131	62	213
Florida	75+	F	3,590	800	1,672	1,235	3,707
Hawaii	15+	F	440	130	146	183	459
Hawaii	15+	M	640	270	196	239	705
Hawaii	15+	O	1080	390	342	422	1154
Hawaii	45-64	F	110	40	37	65	142
Hawaii	45-64	M	170	80	57	78	215
Hawaii	45-64	O	280	110	94	143	347
Hawaii	65-74	F	110	40	33	58	131
Hawaii	65-74	M	180	70	51	80	201
Hawaii	65-74	O	290	110	84	138	332
Hawaii	75+	M	250	110	88	121	319

(continued)

Table A-1. Cases Where Cardiovascular Disease (CVD) Deaths + Diabetes Underlying Cause + Kidney Deaths > All-Cause Deaths (continued)

State	Age	Sex	All-Cause	CVD	Diabetes	Kidney	CVD+Diabetes+ Kidney
Illinois	65-74	M	1,530	560	544	460	1,564
Illinois	75+	F	1,890	410	733	903	2,046
Maryland	75+	F	940	210	397	372	979
Maryland	75+	M	1,310	570	364	441	1,375
Massachusetts	75+	M	1,250	460	370	483	1,313
Michigan	15-44	F	120	20	61	48	129
Michigan	75+	F	1,720	410	728	633	1,771
Missouri	75+	M	1,430	590	416	481	1,487
Montana	65-74	M	120	40	55	30	125
New Jersey	15-44	F	50	10	22	20	52
New Jersey	45-64	M	1,060	430	354	361	1,145
New Jersey	75+	F	1,530	340	506	693	1,539
New Jersey	75+	M	1,980	860	482	939	2,281
New Mexico	65-74	M	310	100	113	108	321
New Mexico	75+	M	480	180	198	117	495
New York	65-74	M	2,180	790	762	769	2,321
New York	75+	M	4,590	1,990	1,101	1,818	4,909
North Dakota	45-64	F	50	20	22	22	64
North Dakota	65-74	M	90	30	33	31	94
Oregon	65-74	M	440	130	284	114	528
South Dakota	65-74	M	90	30	35	37	102
Texas	45-64	M	4,810	1,860	1,415	1,580	4,855
Utah	45-64	M	290	100	148	78	326
Utah	65-74	M	250	70	130	65	265
Utah	75+	M	480	190	212	97	499
Vermont	45-64	M	70	40	26	12	78
Vermont	75+	F	80	20	46	19	85
Virginia	15-44	F	90	20	40	37	97
Washington	15-44	F	50	10	26	19	55
Washington	45-64	M	780	300	325	181	806
Washington	65-74	M	830	270	398	182	850
Washington	75+	F	880	170	479	273	922
Wisconsin	75+	F	780	160	390	277	827
Alaska	75+	F	50	10	32	17	59

Notes: F = female, M = male, O = overall (both males and females), All cause and CVD deaths are rounded to the nearest ten as they are estimated using an attributable fraction approach. Diabetes and Kidney deaths represent population-based averages. Diabetes deaths represent deaths where diabetes was reported as the underlying cause of death.

Table A-2. Diabetes Prevalence at the National Level Used in the Computation of Years of Potential Life Lost

Sex	Age Group	Diabetes Prevalence (BRFSS 2021)	RSE
Male	Age 18–24	1.3%	16.97
Male	Age 25–29	1.8%	13.82
Male	Age 30–34	2.9%	10.12
Male	Age 35–39	4.2%	9.42
Male	Age 40–44	7.1%	6.96
Male	Age 45–49	10.6%	5.48
Male	Age 50–54	13.2%	4.46
Male	Age 55–59	18.2%	3.80
Male	Age 60–64	21.5%	3.12
Male	Age 65–69	22.8%	3.08
Male	Age 70–74	25.7%	3.49
Male	Age 75–79	28.2%	3.48
Male	Age 80+	25.2%	4.27
Female	Age 18–24	1.4%	13.27
Female	Age 25–29	2.0%	12.68
Female	Age 30–34	2.5%	10.35
Female	Age 35–39	4.4%	9.82
Female	Age 40–44	6.8%	8.11
Female	Age 45–49	10.7%	6.06
Female	Age 50–54	11.4%	4.42
Female	Age 55–59	16.0%	3.70
Female	Age 60–64	17.4%	3.67
Female	Age 65–69	19.6%	3.61
Female	Age 70–74	20.9%	3.26
Female	Age 75–79	22.5%	4.09
Female	Age 80+	18.8%	4.64

RSE= relative standard error. Source: 2021 BRFSS (Behavioral Risk Factor Surveillance System (2021)).

Table A-3. Estimated EQ-5D Index from the Number of Healthy Days by Age Category

Healthy Days	EQ-5D				
	18–24 Years	25–44 Years	45–64 Years	65–74 Years	75+ Years
30	0.999	0.998	0.968	0.905	0.883
29	0.998	0.995	0.834	0.823	0.811
28	0.997	0.949	0.827	0.817	0.806
27	0.994	0.842	0.823	0.809	0.795
26	0.992	0.833	0.818	0.802	0.782
25	0.914	0.827	0.809	0.796	0.778
24	0.843	0.824	0.803	0.784	0.776
23	0.839	0.821	0.800	0.779	0.773
22	0.832	0.816	0.797	0.776	0.770
21	0.829	0.811	0.795	0.776	0.769
20	0.826	0.804	0.787	0.773	0.764
19	0.824	0.801	0.778	0.770	0.758
18	0.823	0.800	0.777	0.769	0.756
17	0.821	0.799	0.776	0.768	0.753
16	0.817	0.798	0.773	0.765	0.716
15	0.805	0.793	0.767	0.740	0.708
14	0.800	0.781	0.761	0.711	0.706
13	0.799	0.776	0.759	0.711	0.706
12	0.797	0.773	0.757	0.710	0.705
11	0.797	0.771	0.755	0.710	0.705
10	0.794	0.767	0.717	0.708	0.704
9	0.789	0.763	0.709	0.707	0.702
8	0.779	0.76	0.708	0.706	0.701
7	0.773	0.758	0.708	0.706	0.701
6	0.771	0.754	0.707	0.706	0.700
5	0.768	0.716	0.706	0.705	0.699
4	0.766	0.710	0.705	0.705	0.695
3	0.765	0.709	0.705	0.705	0.694
2	0.763	0.708	0.704	0.704	0.692
1	0.760	0.706	0.704	0.703	0.689
0	0.528	0.479	0.464	0.453	0.441

Source: Table 2, Jia and Lubetkin (2008). EQ-5D = EuroQol five dimensions questionnaire. The EQ-5D is a standardized instrument for measuring generic health status.