December 2016

Diabetes State Burden Toolkit

Technical Report

[This page intentionally left blank.]

CONTENTS

Section **Page**

TABLES

This technical report was developed by RTI International under contract number 200-2013-M-53964B from the Centers for Disease Control and Prevention. The section of the report on direct medical costs was developed by the Lewin Group, Inc., under subcontract number 1-312-0213618-51951L with RTI International. 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 Claudia Steiner and Richard W. Jordan from the Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP) for their valuable collaboration to produce reports for states whose data were not publicly available; AHRQ HCUP Data Partners that contributed to the State Inpatient Database HCUP (list of state organizations is available at http://www.hcup-us.ahrq.gov/db/hcupdatapartners.jsp); state partners that contributed to the Behavioral Risk Factor Surveillance System; and the following members of the Expert Advisory Panel who provided advice on the technical approach, development, and data sources to be used for the development of the Burden Toolkit: Barbara Bartman, Jonathan Brown, William Herman, Matt Petersen, and Kristi Pier.

[This page intentionally left blank.]

1. INTRODUCTION

Diabetes is a serious health condition and is a major contributor to heart disease, kidney disease, stroke, vascular/blood vessel disease, and vision loss. Prevalence of diabetes increased dramatically from 1990 through 2009 (leveling off since then), with nearly 26 million Americans currently living with this condition. Given health complications associated with diabetes and its increasing prevalence, the disease imposes a large and growing economic burden on the health care system and society. Besides the individuals with diabetes themselves, other parties are also affected by the financial burden of diabetes, including private insurers, state and federal governments, and employers.

Preventing type 2 diabetes and diabetes complications could result in improved health and reduced downstream health care costs. The Centers for Disease Control and Prevention (CDC) works with state health departments and other stakeholders interested in diabetes prevention and control, including insurers, employers, and community-based organizations, to reduce the health and economic burden associated with diabetes by preventing diabetes and improving diabetes care. These organizations have an emerging need for information on the health and economic burden of diabetes at the state and local levels and the impact of investments in type 2 diabetes prevention efforts.

Toolkits can provide a fast, convenient, and reliable way to generate state-level estimates of the health and economic burden of diabetes in adults and assess the impact of interventions targeting prevention and delay of type 2 diabetes. CDC has contracted with RTI International to develop two online toolkits: one to calculate and report the state-level health and economic burden of diabetes (Diabetes State Burden Toolkit, or the Burden Toolkit, for short) and another to estimate the potential health and economic impacts of implementing evidence-based type 2 diabetes prevention interventions (Diabetes Prevention Impact Toolkit, or the Impact Toolkit, for short) for states, employers, and health insurers. This technical report describes the data and methods used to generate estimates presented in the Burden Toolkit.

[This page intentionally left blank.]

2. DATA AND METHODS

The Burden Toolkit consists of three sections: (1) diabetes health burden, (2) diabetes economic burden, and (3) diabetes mortality. Estimates reported in the Burden Toolkit are for adults only and exclude children younger than age 18. Estimates are reported for type 1 and type 2 diabetes combined because of data limitations.

2.1 Diabetes Health Burden

This section of the Burden Toolkit reports the health burden of diabetes in each state and nationally and reports statistics on diabetes prevalence, diabetes incidence, and diabetes-associated conditions. Reporting of diabetes-associated conditions is based on self-reported data, hospitalization data, and Medicare data.

The following annual estimates are reported in the health burden section of the Burden Toolkit at the state level and nationally:

1. Diabetes prevalence:
    a. Age-adjusted prevalence of diabetes in total and by sex
    b. Estimated number of people with diabetes in total and by age group/sex
    c. Prevalence of diabetes in total and by age group/sex (with 95% confidence intervals [CI])
2. Diabetes incidence:
    a. Crude rate of new cases of diagnosed diabetes per 1,000 (with 95% CI)
    b. Age-adjusted rate of new cases of diagnosed diabetes per 1,000 (with 95% CI)
    c. Number of new cases of diagnosed diabetes (with 95% CI)
3. Associated conditions:
    a. Self-reported data:
        i. Age-adjusted prevalence of conditions among adults with diabetes, in total and by sex
        ii. Number of adults with conditions and diabetes, in total and by age
        iii. Prevalence of conditions among adults with diabetes, in total and by age (with 95% CI)
        iv. Number of condition cases attributable to diabetes, in total and by age
    b. Hospitalization data:
        i. Age-adjusted rate of hospitalizations with conditions among adults with diabetes, in total and by sex
        ii. Number of hospitalizations with conditions among adults with diabetes by age group/sex

iii. Rate of hospitalizations with conditions among adults with diabetes by age group/sex (with 95% CI)

iv. Number of diabetes-attributable hospitalizations with conditions

c. Medicare data

i. Age-adjusted prevalence of conditions among Medicare beneficiaries with diabetes by age group/sex

ii. Number of people with conditions among Medicare beneficiaries with diabetes by age group/sex

iii. Prevalence of conditions among Medicare beneficiaries with diabetes by age group/sex (with 95% CI)

iv. Number of 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 2013 Behavioral Risk Factor Surveillance System (BRFSS) data and followed approaches used by CDC's United States Diabetes Surveillance System (available at http://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html) to estimate prevalence of diabetes. BRFSS is a state-based, cross-sectional telephone interview survey conducted by CDC and state health departments annually (CDC, 2014). The survey represents the civilian noninstitutionalized adult population in each of the 50 states and the District of Columbia (DC). BRFSS is used to collect prevalence data regarding risk behaviors and preventive health practices among U.S. adults. We excluded respondents with missing age or missing diabetes status from the analysis and, using BRFSS sample weights, calculated the percentage of adults who answered "yes" to the question, "Has a doctor, nurse, or other health professional ever told you that you have diabetes?"

This diabetes prevalence was calculated by age group (18–44, 45–64, 65–74, and 75 or older) and sex. We used the rule of sample size greater than 50 and relative standard error (RSE) of less than 30% to evaluate whether each age group/sex cell provided reliable estimates. The data were not reliable by sex in the 18 to 44 age category in DC and eight states: Alaska, Arizona, Delaware, Idaho, Nevada, New Hampshire, Vermont, and Wisconsin. For these locations, we collapsed the data across men and women and reported one estimate for the 18 to 44 age group (without the sex stratification).

We multiplied the percentage of people with diabetes in each age group/sex cell by the weighted number of total respondents in each age group/sex cell to calculate the total number of adults with diabetes in each age group/sex cell. We age-adjusted total and by sex estimates of prevalence of diabetes and conditions to the 2000 U.S. standard population following methodology described by Klein et al. (2001).

2.1.2 Diabetes Incidence

For diabetes incidence, we used crude and age-adjusted rates of newly diagnosed cases of diabetes for 2013 downloaded from the CDC Web site (http://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html). These estimates were derived from the BRFSS. Self-reported diagnosed diabetes was defined by using the survey question, "Has a doctor, nurse, or other health professional ever told you that you have diabetes?" The age at which each person was diagnosed with diabetes was defined by using the survey question, "How old were you when you were told you have diabetes?" We calculated the number of years each person had been diagnosed with diabetes by subtracting the age at which they were diagnosed from their current age. Adults who had a value of zero were identified as having been diagnosed with diabetes within the last year. In addition, half of the adults who had a value of one were classified as having been diagnosed with diabetes within the last year. To calculate incidence, the numerator was the weighted number of adults who were diagnosed with diabetes within the last year, and the denominator was the weighted estimate of the adult population, excluding adults who had been diagnosed with diabetes for more than 1 year and adults who answered "refused," "don't know," or had missing values on the diabetes status question. Three-year averages were used to improve the precision of the annual estimates. States with less than 2 years of data were excluded from the analysis (18 states and DC). States for which incidence rates or numbers were not estimated for 2013 are indicated with "No Data" in the Burden Toolkit.

2.1.3 Diabetes-Associated Conditions

The Burden Toolkit reports statistics on diabetes-associated conditions from three types of data sources: (1) self-reported conditions from BRFSS, (2) hospitalization events from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), and (3) claims from the Medicare data. BRFSS self-reported conditions are hypertension, high cholesterol, severe vision impairment or blindness, mobility limitations, limitations in instrumental activities of daily living, and coronary heart disease (CHD). Estimates from BRFSS represent self-reported lifetime prevalence and are mostly defined using "have you ever been told" questions. Hospitalization events 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 events for which individuals were hospitalized within a given year. Claims from the Medicare data are reported for CHD, CHF, chronic kidney disease (CKD), and peripheral vascular disease. Estimates from Medicare data represent treated events among Medicare beneficiaries. 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 Burden 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 2013 BRFSS data to estimate self-reported diabetes-associated conditions. A list of these conditions and questions used to define each of them are presented in Table 2-1. Although BRFSS also includes limitations in activities of daily living, we did not report this condition in the Burden Toolkit due to data unreliability.

Table 2-1. Definitions of Self-Reported Diabetes-Associated Conditions, 2013 Behavioral Risk Factor Surveillance System

Condition	Definition
Hypertension	Have you ever been told by a doctor, nurse, or other health professional that you have 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?
Severe vision impairment or 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 self-reported condition among people with diabetes by calculating the percentage of people with diabetes who responded “yes” to the corresponding condition question using BRFSS sample weights. We excluded respondents with missing age, missing diabetes status, and/or a missing response to the conditions question. Additionally, we excluded women reporting diabetes or hypertension during pregnancy only (gestational diabetes and gestational hypertension). We report prevalence estimates by age categories (18–44, 45–64, 65–74, and 75 or older). As with diabetes prevalence, we used the data reliability criteria of sample size less than 50 or RSE greater than 30% and aggregated age categories that did not meet both of the criteria. The level of aggregation varied across conditions and across states. For example, for hypertension, we report prevalence for all four age categories in 39 states; in the remaining 11 states and DC, we had to aggregate the two youngest age categories together and report prevalence of hypertension for three age groups: 18–64, 65–74, and 75 or older. We calculated the number of people with each condition and diabetes by age by multiplying prevalence of the 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 risk factor or set of risk factors (Flegal, Graubard, & Williamson, 2004; Rockhill, Newman, & Weinberg, 1998). Because the prevalence of diabetes and its attributable conditions increase with age, the AFs should be estimated separately by age group. Rockhill, Newman, and Weinberg (1998) and Flegal, Graubard, and Williamson (2004) explain that when confounding factors and/or effect modifications are present, the correct formula for calculating the diabetes AF for disease j is shown in Equation 1:

$$AF_j = pd_j\left[\frac{RR_j - 1}{RR_j}\right] \tag{1}$$

where pd_j is the adjusted prevalence of diabetes in the subsample with the condition j, and RR_j is the adjusted relative risk (RR) of condition j in the diabetes subsample relative to the non-diabetes sample. For each age group, we predicted the probability of having diabetes among individuals with the condition (pd_j) using a logit command in Stata and controlling for age (in years), sex, and race/ethnicity (black, Hispanic, other race [including missing race], and white [variable omitted from the regression model]). The model was weighted using BRFSS sample weights to account for the survey design of the data.

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 and 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, Hispanic, other (including missing race), and white (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 the outcome variable 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, other (including missing race and Hispanics), and white (omitted category). If QCS was still present in the more aggregated model, we used results from a model without race/ethnicity controls.

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

$$\begin{gathered}\text{Number of condition}_j \text{ cases attributable to diabetes} = \\ \text{Number of people with condition}_j * \text{AF}_j\end{gathered} \tag{3}$$

Note that in the states where prevalence of the diabetes-associated conditions was estimated for three age groups and in total (instead of four age groups and in total), the number of attributable cases is also reported for three age groups and in total. At the national level, estimates are reported for four age groups and in total; thus, the national number of diabetes-attributable cases summed across four age groups does not add up to the reported 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 or older, and in total (18 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 estimates from Alaska, but they are included in the total count for ages 18 or older. As a result, the number of diabetes-attributable cases by age groups at the national level may be underestimated. We report the number of cases attributable to diabetes in thousands. In the Burden Toolkit, we do not report the number of condition cases attributable to diabetes for age categories where the p-value for the RR was >0.10 (the cells are left blank, and a note indicates that the data were unreliable).

2.1.2.2 Hospitalization Data

We used HCUP SID data sponsored by the Agency for Healthcare Research and Quality (AHRQ) to estimate hospitalizations with diabetes-associated conditions (AHRQ, April 2016). The SID captures hospital inpatient stays in a given state and contains clinical and resource-use information that is included in a typical discharge abstract (AHRQ, June 2016). We analyzed publicly available data for 28 states and directly obtained estimates for 18 states and DC from AHRQ HCUP through active intramural collaboration. The most recently available years of data for SID vary across the states and are listed in Table 2-2. Three states (Alabama, Delaware, and Idaho) do not participate in the HCUP SID; thus, we were not able to report hospitalization data for them. New Hampshire's latest year of available data was 2009, and thus this state was also excluded from our analyses.

Table 2-2. Most Recent Year of Data from the State Inpatient Databases

State	Year of Data from SID
Alaska	2012
Arizona	2013
Arkansas	2012
California	2011
Colorado	2013
Connecticut	2014
District of Columbia	2014
Florida	2013
Georgia	2014
Hawaii	2012
Illinois	2013
Indiana	2014
Iowa	2013
Kansas	2014
Kentucky	2013
Louisiana	2014
Maine	2011
Maryland	2012
Massachusetts	2012
Michigan	2012
Minnesota	2014
Mississippi	2011
Missouri	2013
Montana	2014
Nebraska	2013
Nevada	2012
New Jersey	2013
New Mexico	2012
New York	2012
North Carolina	2012
North Dakota	2014
Ohio	2013
Oklahoma	2014
Oregon	2013
Pennsylvania	2014

(continued)

Table 2-2. Most Recent Year of Data from the State Inpatient Databases (continued)

State	Year of Data from SID
Rhode Island	2012
South Carolina	2013
South Dakota	2012
Tennessee	2014
Texas	2014
Utah	2011
Vermont	2013
Virginia	2014
Washington	2013
West Virginia	2012
Wisconsin	2013
Wyoming	2014

Note: Three states (Alabama, Delaware, and Idaho) do not participate in the Healthcare Cost and Utilization Project State Inpatient Databases; thus, we were not able to report hospitalization data for them. New Hampshire's latest year of available data was 2009, and thus this state was also excluded from our analyses.

In SID, we identified persons with diabetes based on ICD-9 code 250 listed in any order diagnosis. Diabetes-associated conditions were defined using a primary (i.e., first-order) diagnosis code (with the exception of LEAs). Conditions, for which we report hospitalization data, and the ICD-9 codes used to identify each of them are listed in Table 2-3.

Table 2-3. Definitions of Hospitalizations with Diabetes-Associated Conditions, State Inpatient Databases

Condition	ICD-9 Code
Congestive heart failure	428
Stroke	430–434, 436–438
Myocardial infarction	410
Lower extremity amputation	84.10–84.19[a] (but exclude 895–897)
Hyperosmolar hyperglycemic nonketotic syndrome	250.2
Diabetic ketoacidosis	250.1
Hypoglycemia	251.0, 251.1, 251.2, 962.3, 250.8 (250.8 is only counted if it is without 259.8, 272.7, 681–682, 686.9, 707.1, 707.8, 707.9, 709.3, 730.0, 730.1, 730.2, 731.8)

[a] Based on procedure codes.

We calculated hospitalization rate with each condition per 1,000 adults with diabetes by age group/sex using Equation 4:

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

The number of people hospitalized with the condition and diabetes was obtained from the SID, and the number of people with diabetes was obtained from the 2013 BRFSS data.

For CHF, MI, and LEAs, we calculated the number of hospitalizations with each condition attributable to diabetes using the AF approach presented in Formula 1 where pd_j is the adjusted prevalence of diabetes among those hospitalized with condition j, and RR_j is the adjusted RR of hospitalization with condition j among those hospitalized with and without diabetes. For each age/sex group, we used a logit model to predict the probability of having diabetes among people with the condition 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 and set it to the middle point of the 5-year interval. We followed the same approach for dealing with QCS as described for the BRFSS conditions. We then used a 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 Burden Toolkit, we rounded the estimate to the nearest 10. We used the following rules to replace unreliable or insufficient results:

1. Replace the number of hospitalizations with the condition and diabetes with 11 if the original number is <11 (these replacements occurred in 152 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 116 out of 1,128 state/condition/age/sex categories).

We did not report the number of stroke hospitalizations attributable to diabetes in any of the states because of high frequency of unreliable data or insufficient sample size. We assumed that all hospitalizations with HHNS, DKA, or hypoglycemia were attributed to diabetes.

We calculated the number of hospitalizations with conditions and diabetes and the number of diabetes-attributable hospitalizations with conditions at the national level by applying the hospitalization rates and AFs aggregated across the 46 states, for which we had SID results, and DC to the total population counts across all states. Readers should also note that, although the national estimates are reported at the annual level, the years of data vary across states (see Table 2-2).

2.1.2.3 Medicare Data

We used data from the Centers for Medicare & Medicaid Services (CMS) 2013 Master Beneficiary Summary File (MBSF) to estimate diabetes-associated conditions among Medicare beneficiaries with diabetes. We merged data from the MBSF Base Segment, the Chronic Conditions Warehouse (CCW) Segment, the Other Chronic or Potentially Disabling Conditions Segment, and the Cost and Utilization Segment (ResDAC, 2016). We restricted our analysis to beneficiaries aged 65 or older; we also removed individuals aged 90 or older with no health care 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. We used already-defined variables from the 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 inpatient and outpatient claims (Chronic Conditions Data Warehouse, 2016). For all of the conditions used in our analysis, CMS uses a reference period of 2 years to identify presence of a condition. We used variables called the end-of-year indicators to identify beneficiaries with the conditions, which means that the algorithm criteria were applied using December 31, 2013, 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 (neither claims, no coverage 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.

For each state, we calculated the prevalence of conditions among beneficiaries with diabetes as the percentage of beneficiaries with diabetes who also have the condition. Estimates were generated by age group/sex with two age groups (65 to 74 and 75 or older).

For each state, we calculated the number of condition cases attributable to diabetes using the AF approach presented in Equation 1 where pd_j is the adjusted prevalence of diabetes among those with condition j, and RR_j is the adjusted RR of condition j among those with and without diabetes. For each age/sex group, we used a logit 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 condition and diabetes and the number of diabetes-attributable cases to the entire Medicare population in the 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 2013 dollars.

The following annual estimates are reported in the economic burden section of the Burden Toolkit at the state level and nationally:

1. Total costs attributable to diabetes, in total and by age group/sex:
   a. Direct costs
   b. Indirect costs
   c. Total costs, in total and per person with diabetes
2. Medical costs attributable to diabetes:
   a. All Payers, in total 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 payers
   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 payers

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, in total and by age group/sex:
        i. Number of work days 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 work days 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, in total 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, in total 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, in total 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 diabetes-attributable medical costs and diabetes-attributable 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, disability that prevents people from working, or early mortality. Methods for estimating indirect costs are described in Section 2.2.3. In the Burden Toolkit, total costs are reported 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, costs by payer, and costs by payer, age group, and sex. We used an AF approach to estimate state health expenditures attributable to diabetes. National and state health expenditures are regularly compiled by CMS (CMS, 2014). To implement this approach, we first estimated the fraction of medical spending for various services, such as ambulatory services, prescription drugs, hospital care, and other services, attributable to diabetes. The general formula used for AF is presented in Equation 5:

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

where AF represents the estimated fraction of medical costs attributable to diabetes, *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 amount of medical costs attributable to diabetes is then calculated as AF multiplied by total medical costs or expenditures.

We applied this general approach to estimate the costs attributable to diabetes for medical services used by the noninstitutionalized and for nursing home costs. To estimate state health expenditures attributable to diabetes, we used the 2008 NHEA from CMS. We used the 2008 data file, because it was the most recent year for which we had access to both NHEA and State Health Expenditure Account (SHEA) data. From the SHEA, we used total medical expenditures by state of residence, including administrative costs and medical spending. National expenditures are available by age group, 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 group and sex.

To obtain state expenditure estimates by age group, sex, payer, and service type, we allocated state aggregate expenditures across age group, sex, payer, and service type categories. Specifically, we started with data from the SHEA. We then estimated diabetes-attributable costs by age group, sex, payer, and service type and summed these to state and national levels for reporting in the Burden Toolkit by using the information from the NHEA. We organized our approach around four major tasks:

1. Estimate state expenditures by age group, 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 health care 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. At the state level, however, only total and service-level spending are available by payer. An algorithm is required to estimate state health expenditures by age group, sex, payer, and service type, as described briefly in this section.

Table 2-4. National Health Expenditure Accounts and State Health Expenditure Accounts Data Available by Category

Level/Payer	Total Spending	Spending by Age Group, Sex	Spending by Service Type	Spending by Age Group, 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.

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

First, we obtained NHEA and SHEA data on aggregated Personal Health Care expenditures for 2008 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 2008 state expenditures by age group, sex, payer, and service type, we used a multi-step process, as summarized below:

1. Calculated per-capita 2008 NHEA costs by age group, sex, payer, and service type.
2. Developed per-capita 2008 NHEA cost estimates by payer and service type.
3. Created an adjustment index equal to 2008 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 2008 NHEA spending by age group, sex, payer, and service type.

We collapsed categories to the following stratifications to provide 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 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 2008 state expenditures at the age group, sex, payer, and service type levels so that the aggregated cost estimates matched the 2008 actual total expenditures from SHEA. We then inflated the state health spending from 2008 to 2013 using historical expenditure growth from NHEA and calibrated estimates to ensure that the sum across all 2013 state estimates matched 2013 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 group 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 age and sex. To estimate 2013 state Medicaid costs, we used publicly available 2011 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 (KFF, 2011). Because the KFF state Medicaid enrollment and spending data by enrollment group did not differentiate spending by service type, we combined spending across the four types of services in the 2008 Medicaid spending data from SHEA. We inflated the Medicaid costs to 2013 using NHEA projected growth in expenditures.

2.2.2.2 State Diabetes Prevalence

We used BRFSS data to estimate state diabetes prevalence by age group, sex, and payer, using an approach similar to that described in Section 2.1 (CDC, 2014). For this estimation, we first assigned payers using data from the 2013 BRFSS core question and module questions on health care access. All BRFSS respondents were asked the core question, “Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, government plans such as Medicare, or Indian Health Service?” We combined this information with responses from the Health Care Access BRFSS module question, “Are you CURRENTLY covered by any of the following types of health insurance or health coverage plans?,” to identify respondents who reported having Medicare, Medicaid, private insurance, and/or other third party or out-of-pocket payments for health care. Because only 37 states and DC asked the module question on health care, we imputed payer type of Medicare and Medicaid for the other 13 states using a chained imputation approach. We then estimated diabetes prevalence by payer, by age group (19 to 64 and 65 or older), and by sex for each state. We used overall state diabetes prevalence by age group and sex to represent prevalence among those with payers other than Medicare or Medicaid.

2.2.2.3 Diabetes Cost Ratios

We generated diabetes cost ratios to estimate the impact of having diabetes on annual health care spending by type of service. We generated a diabetes cost ratio, which is the ratio of predicted costs for people with diabetes over predicted costs for people with diabetes under the scenario in which they do 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 health care expenditures by types of service, as well as the state indicator that allows for state-specific cost ratios to be calculated. Although the MEPS restricted file has the majority of these required variables and the state indicator, it requires an involved application process and contains only 29 large states that may also encounter sample size issues 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 MEPS publicly available file to calculate cost ratios by age group, sex, payer, and type of service (all services combined for Medicaid) (AHRQ, 2009).

We used the 2008 to 2012 MEPS 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 asks detailed questions on 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 responded as having Medicare or Medicaid as their primary insurance for at least 1 month during the 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 health care 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	Medicare Paid	Medicaid Paid	Other Paid	Private Plan Paid
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

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

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 not able 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 need 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 private paid anyway along with Medicare, Medicaid, and Other for flexibility.

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

Using multivariate regression analysis, we estimated the cost ratios by age group (19 to 64 or 65 or older), sex, and payer. 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, including age, age squared, sex, race/ethnicity (white, black, Hispanic, Asian, other), poverty status (poor, near poor, middle income, and high income), 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 the regressions for Medicare and Medicaid to adjust for lengths of coverage. As recommended by the Expert Advisory Panel, we did not control for comorbidities.

The cost data are highly skewed and include many nonusers of the health care system with zero spending as well as users who have 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 the cost variables, looking at model goodness-of-fit statistics, as well as analyzing the results of the family link test.

The cost ratio for individuals in age group a, sex s, payer p, and, service type t was estimated as follows (Equation 6):

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

where $RR_{a,s,p,t}$ 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.

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

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

The usual AF formula is as follows:

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

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

Equation 7 can be rewritten as follows:

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

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

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

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

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

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" nondiabetes-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 (Equation 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:

$$Y_j = \frac{pd_j \times \left(RR_j - 1\right) \times C_{noDM} \times \left(1 + \phi\right)}{pd_j \times \left(RR_j - 1\right) \times C_{noDM} \times \left(1 + \phi\right) + pd_j \times C_{noDM} \times \left(1 + \phi\right) + \left(1 - pd_j\right) \times C_{noDM}} \tag{11}$$

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

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

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

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

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, age group (19 to 64 and 65 or older), sex, payer, and type of service (all services combined for Medicaid) 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 2.2.2.2 to estimate state-level diabetes-attributable nursing home costs by age group, sex, and payer (CMS, 2016).

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

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

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 (CMS, 2009). The number of nursing home residents by diabetes status was estimated from the MDS data using a data reference period of April 2013 to May 2015. 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. Our estimates of diabetes prevalence in the community are from BRFSS 2013 and use the same estimation approach as described in Section 2.1.2.

Estimate Nursing Home Costs by Age Group, Sex, and Payer. As described in Section 2.2.2.1, we estimated 2013 state-level nursing home costs by age group, sex, and payer. Because continuing care retirement communities (CCRCs) are included in the "Nursing Care Facilities and Continuing Care Communities" cost category of SHEA but are not considered to provide ongoing nursing care, we discounted state-level nursing homes costs by the national percentage of payments from CCRCs.

Calculate Diabetes-attributable Nursing Home Costs. We multiplied the nursing home AF by state nursing home costs to estimate diabetes-attributable nursing home costs by state, age group, sex, and payer.

2.2.2.6 Total Medical Costs

We added the diabetes-attributable nursing home costs to the diabetes-attributable medical costs among the noninstitutionalized population to estimate medical costs of diabetes by age group, sex, and payer. These medical cost estimates are provided in the medical cost section of the Burden Toolkit and reflect medical spending and administrative costs for ambulatory care, hospital care, prescription drugs, nursing home care, durable medical equipment, and other care.

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 of work absenteeism, work presenteeism, household productivity losses, and inability to work. We 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 count of people with diabetes includes the noninstitutionalized 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 workdays 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 estimates of the number of employed people with diabetes in each state, by age group, and 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 (CDC, June 2016). NHIS is a cross-sectional household interview survey administered by CDC that was designed with a goal 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 2009 through 2013 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 in the NHIS public use data files. In NHIS, we identified persons with diabetes using the question "Have you ever been told that you have 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 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 four different models for best fit: one-part negative binomial model, two-part truncated negative binomial model with a logit, two-part GLM 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 last 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. We estimated diabetes-attributable 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. We estimated predicted values using coefficients from both the logit and GLM models. Note that 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) Table Creator to estimate daily earnings (CPS, 2016). We used 2015 CPS earnings data, which reflect annual earnings from 2014 and include income from wages, salaries, and self-employment. Average earnings were estimated by 5-year age groups and sex and included full-year full-time workers, part-year full-time workers, full-year part-time workers, and part-year part-time workers. For consistency with the direct medical cost estimates, we used the Bureau of Economic Analysis (BEA) Price Indexes for GDP to deflate the annual earnings from 2014 to 2013 dollars (U.S. BEA, 2016).

Second, we used 2014 national and state-level occupational employment mean wage estimates from the BLS to estimate a state-to-national wage ratio (OES, 2016). 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 2014 Census population estimates (U.S. Census Bureau, 2015). We calculated mean earnings per day of work by dividing annual mean earnings by 250, which is the typical number of week days 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 ($\text{Per\_Cap\_Missed\_Work\_Cost}_{sag}$) as follows:

$$\text{Per\_Cap\_Missed\_Work\_Cost}_{sag} = \text{Work\_loss\_Diab}_{rag} * \text{Daily\_Earn}_{sag}, \qquad (16)$$

where $\text{Work\_loss\_Diab}_{rag}$ 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 2011 through 2013 NHIS data to calculate the percentage of people with diabetes who are employed by region, age group, and sex (CDC, June 2016). 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:

$$\text{Diab\_work}_{sag} = \text{Perc\_employ|diab}_{rag} * \text{Num\_diab}_{sag}, \qquad (17)$$

where $\text{Perc\_employ|diab}_{sag}$ 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 Burden 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:

$$\text{Absenteeism\_tot}_{sag} = \text{Per\_Cap\_Missed\_Work\_Cost}_{sag} * \text{Diab\_work}_{sag} \quad (18)$$

For this calculation, we multiplied the per-capita cost of missed work attributable to diabetes ($\text{Per\_Cap\_Missed\_Work\_Cost}_{sag}$) 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 Burden 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 on the job productivity losses. 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) analysis, the authors assumed that, on average, 6.6% of annual productivity is lost while people are at work as a result of diabetes (ADA, 2013). We multiplied this reduced productivity estimate by state-level daily earnings by age group and sex ($\text{Daily\_Earn\_per\_cap}_{sag;}$ 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:

$$\text{Present\_Cost\_per\_cap}_{sag} = 0.066 * \text{Daily\_Earn\_per\_cap}_{sag} * (250 - \text{Days\_Missed\_Work}_{ag}) \quad (19)$$

where 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. $\text{Days\_Missed\_Work}_{ag}$ 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 2011–2013. 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 Burden 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 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 NHIS and the same methodology that we used to estimate the number of workdays lost for the absenteeism cost analysis. However, 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 does not result in double counting of costs.

Step 1: Estimated bed days attributable to diabetes. We defined number of days spent in bed using 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 patient in a hospital)." Using 2009–2013 NHIS data, we tested four different regression models for best fit: one-part negative binomial, two-part truncated negative binomial with a logit, two-part GLM 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 a 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 represents the presence of other comorbidities, Region_t represents region of residence (Census region), and X_t represents other sociodemographic characteristics. We used a GLM with a gamma distribution and a log link to estimate 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 controls: age, age squared, race/ethnicity, education, family income, health insurance, and employment status. We estimated diabetes-attributable per person number of days spent in bed, calculated by age and sex at the regional level, as the mean difference between predicted number of bed days for each person with diabetes and the predicted number of bed days assuming no diabetes. We estimated predicted values 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, 2013" publication (Expectancy Data, 2014). 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 2015 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 total household productivity costs. We calculated total household productivity losses by age group and sex for each state ($HH\_prod\_loss\_tot_{sag}$) as follows:

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

where $HH\_prod\_loss\_PC_{sag}$ 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 Burden Toolkit. In the Burden Toolkit, the per person household productivity costs are reported per person with diabetes where persons with diabetes include the noninstitutionalized population from BRFSS and 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 to stop working. 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, applying this probability to state estimates of the number of people with diabetes by age group and sex, and then valuing 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: Estimated probability of being unable to work attributable to diabetes. We estimated the probability of being unable to work because of diabetes using 2011–2013 NHIS data (CDC, June 2016). We defined a person as being unable to work if he or she answered "Disabled" to the 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, using a logistic regression model, as follows:

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

where Unable_to_work_t represents whether an individual reports being unable to work because of a health condition; Diab_t denotes whether the individual has diabetes; Comorbid_t represents the presence of other comorbidities (e.g., arthritis, asthma), Region_t represents region of residence (Census region), and X_t represents other sociodemographic characteristics (age, sex). 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 because of 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, or Num_diab_{sag}, was from Section 2.1.2.

Step 3: Calculated total 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 in Section 2.2.3.2), as follows:

$$Unable\_to\_Work\_Diab\_Cost_{sag} = Num\_unable\_to\_work_{sag}\_ * Annual\_Earn_{sag}, \quad (26)$$

where $Unable\_to\_Work\_Diab\_Cost_{sag}$ 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 Burden 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. 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). Future costs were discounted by the probability of surviving to each year of age at which the expected production occurs. We used 2010 U.S. life tables from the National Vital Statistics Report to calculate compounded survival rates for each age group (Arias, 2014). 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 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). 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.

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 stakeholder groups with estimates of costs or losses that they incur as a result of diabetes. The costs reported in this portion of the Burden Toolkit are estimates that may be useful to stakeholders 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. Stakeholders that 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 Prevention Impact Toolkit.

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 health care 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 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. 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 or older) and sex. We used 2013 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, where we imputed payer status for states that did not ask the health insurance module questions. We then estimated diabetes prevalence among the privately insured by age group and sex for each state, also using 2013 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. To estimate employer costs, we applied per-person diabetes-attributable private insurance cost estimates to the number of employees with diabetes.

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. 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 the state and average cost per employee with diabetes in the state.

2.3 Diabetes Mortality

This section of the Burden Toolkit reports diabetes mortality estimates in each state and nationally and consists of diabetes-attributable deaths, years of life lost (YLLs), and quality-adjusted life years (QALYs) lost due to diabetes.

The following annual estimates are reported in the mortality section of the Burden Toolkit at the state level and nationally:

1. Mortality
    a. Number of diabetes-attributable deaths, in total, 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, in total, by sex, by age group, and by sex/age group
        i. Diabetes as the underlying cause of death, defined as "the disease or injury that initiated the chain of morbid events that led directly and inevitably to death"
        ii. All causes and cause-specific deaths, including CVD deaths and ESRD deaths attributable to diabetes
2. YLLs, in total and by age group/sex
    a. Estimated average YLLs attributable to diabetes
    b. Number of persons with diabetes (in thousands)
    c. Total YLLs attributable to diabetes (in thousands)
3. QALYs lost, in total 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 measure of the mortality estimates is described in detail in the following subsections.

2.3.1 Mortality Data

The mortality section of the Burden 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 or older) by sex and state using 2013 CDC Wide-ranging Online Data for Epidemiologic Research (CDC WONDER) mortality data (CDC, July 2016). 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 Burden Toolkit reports number and rate of deaths with diabetes as the underlying cause of death and diabetes-attributable deaths for all causes, CVD, and ESRD for 780 (52*5*3) different combinations of states and DC (51 plus the United States as a whole), age categories (4 plus overall), and sex (2 plus overall).

Aggregating up to four age groups to match diabetes prevalence calculated in Section 2.1.2 drastically decreases the percentage of suppressed or unreliable cohorts. Originally, we planned to report deaths attributable to heart disease and stroke/cerebrovascular disease, but we could not obtain sufficiently precise estimates of the relative risk for those causes separately. So we combined deaths from heart disease and stroke into one group: CVD deaths.

In CDC WONDER, mortality statistics are suppressed when $n < 10$ for the specified strata and are considered unreliable when $n < 20$; thus, we are not able to report the data for these strata in the Burden Toolkit. To minimize the amount of suppressed and unreliable data at the state level, aggregation of data was needed for some subgroups. We used the following set of rules to aggregate the data.

For reporting the number of deaths from CDC WONDER, we used the following rules, ranking from most desirable to least desirable:

1. Use 2013 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% of CVD deaths; and 83% of diabetes as the underlying cause of deaths).
2. When #1 is suppressed or not reliable, pool state data through 2011–2013 and divide by 3 to calculate an average annual death rate (2% of deaths with diabetes 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 2013 regional death rates and apply to state cohort population (0.2% of CVD deaths; 0.9% of diabetes deaths).
4. If #1–3 all yield suppressed numbers, we report “suppressed” in the Burden Toolkit (1.5% of CVD deaths; 6% of diabetes deaths).
5. If #1–3 all yield unreliable numbers, we report the 2013 state value (#1), but caution users that such estimates may be unreliable (1.3% of CVD deahs; 8.1% of diabetes deaths).

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. The AF approach estimates 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. To do so, we used Miettinen’s formula for calculating AF (1974):

$$AF_j = p\left[\frac{RR_j - 1}{p(RR_j - 1) + 1}\right] \tag{27}$$

where p is diabetes prevalence, and RR_j is the adjusted RR of disease *j* (in our case, from all causes or CVD) in the diabetes subsample relative to the non-diabetes sample. We used diabetes prevalence from BRFSS 2013, stratified by age group, sex, and state. For more information on how the diabetes prevalence was estimated for each state, see Appendix A. We applied Miettinen’s formula for all-cause and CVD mortality attributable to diabetes because CVD is a major diabetes-related causes of death.

One concern about Miettinen’s formula is that it is not appropriate if adjusted RRs are included in the presence of confounding. 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 age, RRs decrease with age, and overall deaths increase with age. Still, some concerns about confounding may remain because our RR estimates 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 decline. RRs were estimated using a logistic regression model, in which the dependent variable was a binary variable for death and the independent variables were diabetes indicator, age, sex, and race. We validated the RR estimates by noting that results from the logistic regression with discrete survival time are not statistically different when using and not using race/ethnicity as a control.

To calculate RR, we used NHIS data as in Gregg et al. (2012) but included more recent NHIS base years (2005–2009) and follow-up (using mortality data up to 2011). We estimated a discrete logistic model stratified by age group (18–44, 45–64, 65–74, and 75 or older) and by sex and adjusted for race/ethnicity.

We computed separate mortality rates for all causes and CVD (Table 2-6). Information from the stratification exercise showed that the RR for all-cause mortality declined with age. When RRs were not available because no males aged 18 to 44 with diabetes died during the study period compared with individuals without diabetes, our best estimate was to assume that deaths were equal to zero.

Table 2-6. Relative Risk (Logit Model) Using 2005–2011 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	3.15	—
Male	Age 45–64	2.90	3.89
Male	Age 65–74	1.90	2.43
Male	Age 75+	1.26	1.12[a]
Female	Age 18–44	2.16	9.62
Female	Age 45–64	2.99	3.33
Female	Age 65–74	2.13	2.68
Female	Age 75+	1.36	1.37

Source: Relative risks: (diabetes vs no diabetes), by age group/sex were computed by Yiling Cheng (CDC).

Note: Dash (—) indicates no relative risk available (no individuals with the condition died in the sample).

[a]This cell (cardiovascular disease, Male 75+) indicates that relative risk is not statistically significant at alpha = 0.1.

We multiplied the AF by the total number of deaths from all causes and CVD deaths to estimate the number of diabetes-attributable deaths for all causes and CVD 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 or was a contributing factor to deaths (Geiss, Herman, & Smith, 1995).

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 CDC's National Center for Health Statistics (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 2013 data from the United States Renal Data System (USRDS) (http://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 Burden Toolkit, we reported 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 used the AF approach for diabetes and ESRD, but we are assumed that the AF = 1. Because we independently estimated all-cause deaths and the deaths from the three specific causes, there is no guarantee that the estimates 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 diabetes prevalence. Because the relative risk 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. 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 (48 out of 780). These are outlined in detail in Appendix A.

The 48 cases are limited to ages 15 to 44 and 75 or older. The specific reasons for the discrepancy between the sum of parts and the total cause of death are as follows:

- In 41 states (plus the United States as a whole), for females aged 15 to 44, the CVD RR for this group is much bigger (9.62) than the all-cause RR for the same group (2.16). Despite the difference in actual deaths between all-cause and CVD, once we apply the AF formula, the numbers of diabetes-attributable CVD and all-cause deaths are comparable. For example, the total number of CVD deaths for females aged 15 to 44 in Ohio equals 288. According to our AF estimations, 53 of these CVD deaths are attributable to diabetes. Total deaths from all causes in this cohort equal 1,974. Using Equation 1, 59 of these total deaths are attributable to diabetes.

- In Minnesota, the “excess” deaths among females aged 15 to 44 are enough to push the overall (across males and females) cause-specific deaths above all-cause deaths.
- In Hawaii, New Mexico, and Vermont, for males aged 75 or older and in North Dakota for women aged 75 or older, the crude counts of deaths with diabetes as the underlying cause are relatively large, while the RR for the computed all cause of deaths is relatively low (i.e., close to 1) for older individuals.
- In New Mexico, the “excess” deaths among males aged 75 or older are enough to push the overall cause-specific deaths among adults aged 75 or older above all-cause deaths (by 6 deaths).

2.3.2 Years of Life Lost (YLLs)

YLL due to diabetes is an indicator of premature mortality and is calculated by multiplying the number of persons with 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 CDC’s NCHS by single year of age and sex; and national-level RR of mortality (Table 2-6) for those with and without diabetes.

We estimated YLLs using 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 existing population as shown in Equation 28:

$$\theta_u = \frac{r}{pr + (1 - p)} \qquad (28)$$

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

The corresponding scale-down factor (θ_d) for mortality for persons without diabetes is shown in Equation 29:

$$\theta_d = \frac{1}{pr + (1 - p)} \qquad (29)$$

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 (number of deaths per 100,000) by using prevalence of diabetes by state, age (5-year bins), and sex; all-cause mortality values from CDC’s NCHS by state, age, and sex; and national-level relative risk 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 published 2010 CDC tables.[3]

$$qx = \text{Number dying between age x and age x+n / number attaining exact age x.} \quad (30)$$

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:

$$\begin{gathered}\text{Number of person years lived } (L_x) = \\ [(\text{Time Interval})/2] * (\text{Number of Persons Alive Age } x = t)\end{gathered} \quad (31)$$

Assuming a cohort of 100,000 births, we calculated the total number of person-years that would be lived after the beginning of the indicated age interval by cumulating the number of person-years lived 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 YLL equals the longest life expectancy for each cohort in the absence of diabetes minus the life expectancy with the condition. YLLs due to diabetes is then averaged across each age group. Total and 18+ age group 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 prevalence estimates by age are not available for the same level of granularity as the life tables (1 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 YLLs through age 89–90 only. However, although the average YLLs for the 75+ strata only include diabetes counts from ages 75–76 through 89–90 in the calculation, the underlying YLLs (and QALYs for each age) calculation accounts for the full age set, following standard life tables, including losses through age 100.

2.3.3 Quality-Adjusted Life Years Lost

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 perfect 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'

[3] At the time of the analysis, the most recent published numbers were from 2010: http://www.cdc.gov/nchs/data/nvsr/nvsr63/nvsr63_07.pdf

answers to the BRFSS Healthy Days (HD) 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 using the following three steps:

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

These steps are outlined in detail in the sections below.

2.3.3.1 Unhealthy Days and EQ-5D

The BRFSS included the HD measures between 1993 and 2013. The HD measures asked respondents to report the number of days in the past 30 days when they had 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 2013 BRFSS data. Physical and mental HD questions were available for all 50 states and DC in the 2013 BRFSS. Both questions required the respondent to answer with a number between 0 and 30. The overall UD measure was calculated by adding together a respondent's PUDs and MUDs with a logical maximum of 30 UD (Equation 32):

$$\text{Unhealthy Days} = \text{minimum}\left(30, PUD + MUD\right) \tag{32}$$

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 have 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, with a logical maximum of 30 HDs (Equation 33). 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).

$$\text{Healthy Days} = 30 - \text{minimum}\left(30, PUD + MUD\right) \tag{33}$$

Total and 18+ age group 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 QALYs

We used the integrated quality-survival product method to calculate QALYs. The method involves separate estimation of the survival function and the utility score. The population-based QALY model was obtained by using the quality-adjusted survival curve (QAS). QAS is formed by plotting against time t, the product of the mean QoL score of the population alive at time t and the probability of surviving at time t. In our case, we considered k discrete time points (annual intervals from birth to age 100) to estimate the expected quality-adjusted survival time and hence QALYs (Billingham et al., 1999):

$$QAS = \sum_{i=1}^{k} \frac{\left(Q_i + Q_{i+1}\right)}{2} \frac{\left(S_i + S_{i+1}\right)}{2} \left(t_{i+1} - t_i\right) \tag{34}$$

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 (i.e., ex = Tx/lx as described Section 2.3.2) and differ by age, sex, and diabetes status. We used published national-level life expectancy for our QALY and YLL calculations so as not to confound state-level effects in life expectancy with diabetes prevalence by state. Average QALYs lost due to diabetes are averaged across each age in the age group. Total and 18+ age group 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.

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 (AHRQ). (April 2016). HCUP Databases. Healthcare Cost and Utilization Project (HCUP). Retrieved from: www.hcup-us.ahrq.gov/databases.jsp

Agency for Healthcare Research and Quality (AHRQ). (June 2016). SID Database Documentation. Healthcare Cost and Utilization Project (HCUP). Retrieved from: www.hcup-us.ahrq.gov/db/state/siddbdocumentation.jsp

Agency for Healthcare Research and Quality (AHRQ). (August 2009). Medical Expenditure Panel Survey (MEPS), Data Overview. Retrieved from: https://meps.ahrq.gov/mepsweb/data_stats/data_overview.jsp

American Diabetes Association (ADA). (2008). Economic costs of diabetes in the U.S. in 2007. *Diabetes Care*, *31*(3), 596–615. doi: 10.2337/dc08-9017

American Diabetes Association (ADA). (2013). Economic costs of diabetes in the U.S. in 2012. *Diabetes Care*, *36*(4), 1033–1046. doi: 10.2337/dc12-2625

Arias, E. (2014). United States life tables, 2010. *National Vital Statistics Report, 63*(7), 1–63. Retrieved from: http://www.cdc.gov/nchs/data/nvsr/nvsr63/nvsr63_07.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: http://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. *Morbidity and Mortality Weekly Report, 62,* 209–211.

Centers for Disease Control and Prevention (CDC). (August 2014). Overview: BRFSS 2013. Retrieved from: http://www.cdc.gov/brfss/annual_data/2013/pdf/overview_2013.pdf

Centers for Disease Control and Prevention (CDC). (2015). *Diabetes Prevention Recognition Program standards and operating procedures*. Retrieved from: http://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 Disease Control and Prevention (CDC). (June 2016). National Health Interview Survey (NHIS) Data, Questionnaires and Related Documentation. Retrieved from: http://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm

Centers for Disease Control and Prevention (CDC). (July 2016). WONDER. Retrieved from: https://wonder.cdc.gov/

Centers for Medicare & Medicaid Services (CMS). (2009). Medicare Program; Prospective Payment System and Consolidated Billing for Skilled Nursing Facilities for FY 2010; Minimum Data Set, Version 3.0 for Skilled Nursing Facilities and Medicaid Nursing Facilities; Finale Rule. *Federal Register*, 74(153), 40288-40395.

Centers for Medicare & Medicaid Services (CMS). (September 2016). Long Term Care Minimum Data Set (MDS). Retrieved from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/IdentifiableDataFiles/LongTermCareMinimumDataSetMDS.html

Centers for Medicare and Medicaid Services (CMS). (May 2014). National Health Expenditure Data. Retrieved from: https://www.cms.gov/Research-Statistics-Data-and-systems/Statistics-Trends-and-reports/NationalHealthExpendData/index.html

Centers for Medicare & Medicaid Services (CMS). (October 2015). 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 Data Warehouse. (2016). Condition Categories. Retrieved from: https://www.ccwdata.org/web/guest/condition-categories

Current Population Survey (CPS). (2016) Table Creator. Suitland, MD: U.S. Census Bureau. Retrieved from: http://www.census.gov/cps/data/cpstablecreator.html

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.

Expectancy Data. (2014). *The dollar value of a day: 2013 dollar valuation*. Shawnee Mission, KS: 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.* 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*, *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.). Washington, DC: 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. Retrieved from: http://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

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

KFF State Health Facts. (2011). Data Source: Medicaid Statistical Information System. *Distribution of Medicaid Enrollees by Enrollment Group*. Retrieved from: http://kff.org/medicaid/state-indicator/distribution-of-medicaid-enrollees-by-enrollment-group/

Klein, R. J., Proctor, S. E., Boudreault, M. A., & Turczyn, K. M. (2002). Healthy People 2010 criteria for data suppression. *Healthy People 2010 Statistical Notes*, 24.

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 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. Retrieved from: http://www.ncbi.nlm.nih.gov/books/NBK53341/

Occupational Employment Statistics (OES). (August 2016). Washington, DC: Bureau of Labor Statistics. Retrieved from: http://www.bls.gov/oes/tables.htm

Pharoah, P. D. P., & Hollingworth, W. (1996). Cost effectiveness of lowering cholesterol concentration with statins in patients with and without pre-existing coronary heart disease: Life table method applied to health authority population. *BMJ*, *312*(7044), 1443–1448.

Research Data Assistance Center (ResDAC). 2016. Master Beneficiary Summary File. Overview. Retrieved from: https://www.resdac.org/cms-data/files/mbsf

Rice, D. P. (1966). *Estimating the cost of illness*. Health Economics Series No. 6. PHS Pub. No. 947-6. Washington, DC: U.S. Government Printing Office.

Rice, D. P., Hodgson, T. A., & Kopstein, A. N. (1985). The economic costs of illness: A replication and update. *Health Care Finance Review*, *7*(1), 61–80.

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

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: http://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.

U.S. Bureau of Economic Analysis (BEA). (November 2016). *Table 1.1.4. Price Indexes for Gross Domestic Product.* Retrieved from: http://www.bea.gov/itable/

U.S. Census Bureau. (June 2015). Annual Estimates of the Resident Population for Selected Age Groups by Sex: April 1, 2010 to July 1, 2014. Retrieved from: http://www.census.gov/popest/data/national/asrh/2015/index.html

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, YLL, and QALYs

Table A-1 shows the nine states with unreliable data on diabetes prevalence in the Behavioral Risk Factor Surveillance System (BRFSS) for the 18 to 44 age group. We defined reliable data as data where the number of respondents is greater than 50 and the relative standard error (RSE) of the weighted mean is less than 30% (Klein et al., 2002). We calculated the RSE using the following formula:

$$\%RSE = \frac{100 * \text{Standard error of mean}}{[\text{mean}]}$$

For these states, prevalence across sexes was combined to obtain reliable estimates. For each state except Alaska and Wisconsin, the RSE was less than 30% for females but higher than 30% for males.

Table A-1. Unreliable Values of Diabetes Prevalence Age 18 to 44 by State

	Female		Male	
State	**N**	**RSE (%)**	**N**	**RSE (%)**
Alaska	888	33.8	841	31.0
Arizona	581	26.4	493	37.6
Delaware	774	20.9	612	32.7
District Of Columbia	730	24.9	553	40.7
Idaho	911	22.3	683	30.4
Nevada	849	28.1	689	38.5
New Hampshire	871	24.2	679	31.4
Vermont	870	26.2	709	36.0
Wisconsin	957	31.4	863	38.4

Note: The number of diabetes diagnosis (N) was aggregated across the sexes because male prevalence in the state/age group were unreliable.

Table A-2 shows all cases (48 out of 780) where CVD deaths + diabetes underlying cause + kidney deaths < all-cause deaths. Diabetes as the underlying cause 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 cardiovascular disease (CVD) deaths attributable to diabetes are estimated obtained using the attributable fraction approach. Because we also independently estimated all-cause deaths and the deaths from the three specific causes, using different data sources, there is no guarantee that the cause-specific estimates will be less than the all-cause deaths attributable to diabetes.

Table A-2. Cases Where Cardiovascular Disease Deaths + Diabetes Underlying Cause + Kidney Deaths < All-Cause Deaths

State	Age	Sex	All Cause	CVD	UCD	Kidney
Alabama	15–44	F	64	56	27	12
Arizona	15–44	F	34	14	41	15
Arkansas	15–44	F	26	25	17	4
California	15–44	F	177	108	97	48
Colorado	15–44	F	18	9	15	7
Connecticut	15–44	F	9	8	0	2
Delaware	15–44	F	6	6	0	2
District of Columbia	15–44	F	2	2	0	1
Florida	15–44	F	99	63	65	36
Georgia	15–44	F	81	70	36	37
Hawaii	15–44	F	4	4	0	5
Hawaii	75+	M	77	12	67	20
Illinois	15–44	F	58	41	55	26
Indiana	15–44	F	34	25	21	12
Iowa	15–44	F	13	8	12	3
Kentucky	15–44	F	42	33	23	10
Louisiana	15–44	F	46	33	30	20
Maine	15–44	F	5	4	0	2
Maryland	15–44	F	40	32	21	9
Massachusetts	15–44	F	22	9	12	6
Michigan	15–44	F	60	44	34	22
Minnesota	15–44	O	58	8	38	13
Minnesota	15–44	F	12	8	15	6
Mississippi	15–44	F	44	31	25	20
Missouri	15–44	F	43	36	20	7
Nebraska	15–44	F	6	5	0	2
Nevada	15–44	F	15	12	13	1
New Jersey	15–44	F	25	19	20	12
New Mexico	75+	O	482	122	309	58
New Mexico	75+	M	172	24	128	29

(continued)

Table A-2. Cases Where Cardiovascular Disease Deaths + Diabetes Underlying Cause + Kidney Deaths < All-Cause Deaths (continued)

State	Age	Sex	All Cause	CVD	UCD	Kidney
New York	15–44	F	86	64	57	20
North Carolina	15–44	F	71	52	45	18
North Dakota	75+	F	114	38	68	14
Ohio	15–44	F	59	53	47	28
Oklahoma	15–44	F	38	29	24	12
Oregon	15–44	F	15	6	15	4
Pennsylvania	15–44	F	65	42	33	23
Rhode Island	15–44	F	3	2	0	1
South Carolina	15–44	F	40	38	12	17
Tennessee	15–44	F	65	45	43	15
Texas	15–44	F	138	97	95	75
United States	15–44	F	1741	1231	1054	597
Utah	15–44	F	10	3	14	6
Vermont	75+	M	59	8	41	11
Virginia	15–44	F	41	29	28	9
Washington	15–44	F	26	15	26	12
West Virginia	15–44	F	20	14	0	9
Wisconsin	15–44	F	16	10	16	8

Notes: CVD = cardiovascular disease, F = female, M = male, O = overall (both males and females), UCD = Diabetes as the underlying cause. UCD and Kidney (death) represent population-based averages.

Table A-3. Diabetes Prevalence at the National Level Used in the Computation of Years of Life Lost

Sex	Age Group	Diabetes Prevalence (BRFSS 2013)	RSE
Male	Age 18–24	1.0%	15.01
Male	Age 25–29	1.7%	17.83
Male	Age 30–34	2.1%	10.86
Male	Age 35–39	3.4%	8.39
Male	Age 40–44	5.9%	6.43
Male	Age 45–49	9.5%	5.03
Male	Age 50–54	13.3%	3.95
Male	Age 55–59	16.3%	3.43
Male	Age 60–64	19.7%	3.08
Male	Age 65–69	24.4%	2.65
Male	Age 70–74	25.1%	2.82
Male	Age 75–79	26.5%	3.89
Male	Age 80+	21.7%	3.56
Female	Age 18–24	1.4%	14.84
Female	Age 25–29	1.6%	11.49
Female	Age 30–34	2.9%	9.24
Female	Age 35–39	4.1%	6.92
Female	Age 40–44	6.4%	5.82
Female	Age 45–49	8.9%	4.49
Female	Age 50–54	10.5%	3.94
Female	Age 55–59	15.0%	3.39
Female	Age 60–64	17.4%	2.81
Female	Age 65–69	20.6%	2.92
Female	Age 70–74	21.4%	2.64
Female	Age 75–79	21.8%	3.30
Female	Age 80+	18.6%	2.98

Note: RSE = relative standard error.

Source: 2013 Behavioral Risk Factor Surveillance System (BRFSS).

Table A-4. 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.

[This page intentionally left blank.]