SUPPLEMENTARY DATA

Estimating annual medical expenditures

1. A Two-part Model

We used a two-part model to estimate the annual medical expenditures in persons with and without diabetes. In the first part, logistic regression was used to model the probability of an individual having a positive medical expenditure. In the second part, a generalized linear model with log-link was used to model the annual medical expenditure given that it is positive. To estimate the cost of persons with diabetes, we used the models to predict the probability of an individuals having a positive medical spending and the mean medical spending if one has any spending (conditional mean), among the sample of individuals who had diabetes. The predicted mean medical spending was calculated as the product of the predicted probability and the predicted conditional mean. For the cost of persons without diabetes, we followed the same procedure, except that we assumed those individuals had no diabetes. The standard error of the mean was estimated using 1000 iteration of bootstrapping.

In both parts of the regression models, the key explanatory variables of interest are self-reported diabetes status, years of having diagnosed diabetes and age. A square term of age and years of having diabetes were added to provide a better fitting cost curve.(Supplementary Figure 1) To allow for potential different age effect among persons with diabetes and without diabetes, interactions between diabetes status and age and age square were also added. By including both diabetes years and the interaction of diabetes and age, the model allows for testing the differential effects of diabetes and age on medical spending by age of diagnosis. In the base-case specification, we exclude diabetes-related cardiovascular diseases, i.e., angina, myocardial infarction and stroke. In the sensitivity analysis, we include them.

2. Model Specification Test

For the second part of the two-part model, we performed a Modified Park test and determined that a gamma variance function was an appropriate variance specification for the model. This specification accounts for an increasing variation associated with higher medical spending.[1] The result of the test is presented in Supplementary Table 1. Compared to other specifications, the model specification performs better in terms of calibration of predictions, mean square errors, absolute prediction error and cross-validated forecast error. [2] The results of the test are presented in Supplementary Table 1. The coefficient on xbetahat is close to 2, which indicates that the variance exceeds mean and thus a gamma family is a more appropriate fit to specify the variance function.

Supplementary Table 1. Modified Park Test on the variance structure of the two-part model

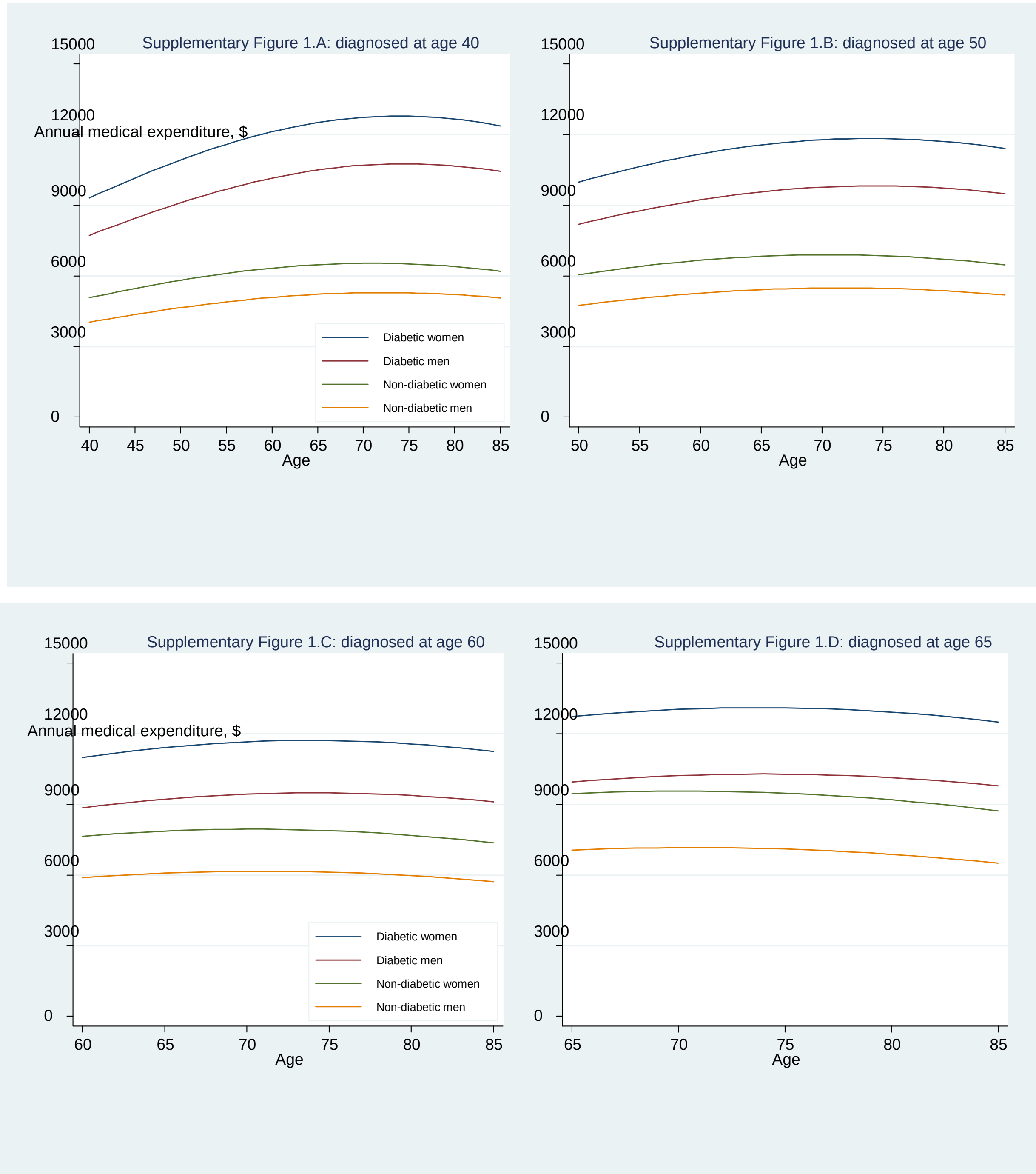
r2	Coef.	Robust Std. Err.	z	P>z	[95% Conf. Interval]	
xbetahat	1.817885	0.1068332	13.27	0	1.208496	1.627275
_cons	6.44735	0.9537121	6.76	0	4.578109	8.316592

3. Prediction of Annual Medical Spending

Supplementary Figure 1 presents the predicated mean annual medical expenditures in persons with diabetes and without. It shows that diabetic patients had significantly higher annual medical expenditures than their non-diabetic equivalents at all ages. The annual incremental medical expenditures attributed to diabetes, as measured by the distance between the lines of diabetes and non-diabetes, are greater in patients diagnosed when younger. In diabetic men diagnosed at age 40, 50, 60, and 65, the mean annual incremental expenditure across the lifespan of diabetes is

$6,000,$5,100,$4,100, and $3,900 respectively. In diabetic women diagnosed at the respective age, the mean incremental expenditure is $6,800, $5,700, $4,600 and $4,500.

Supplementary Figure 1. Predicated mean annual medical expenditures (undiscounted) by diabetes status.



Source: estimated by authors using the Medical Expenditures Panel Survey 2006-2009 linked with the National Health Interview Survey 2005-2008.
Note: the annual medical expenditures are expressed in 2012 dollar but not discounted.

SUPPLEMENTARY DATA

Differential mortalities by diabetes status

1. Relative risk of mortality

The relative risk of mortality was derived from two previous studies. Using a 12-16 years mortality follow-up of the National Health and Nutritional Examination Survey sample of year 1976-1980, Saydah and colleagues compared the all-cause mortality and cause-specific mortality among persons with and without diabetes.[3] Their estimated relative hazards between the two groups are summarized in Supplementary Table 2. In a more recent study by Gregg et al found that, the relative hazards have declined in recent years. [4] Their estimated hazard rate ratios were used to update the relative hazards.

Supplementary Table 2.

Age group at baseline	Adjusted relative hazard	HRR cohort 2003-04 vs. 1997-98	updated relative risk
	Men		
30-49	2.57(1.15,5.81)	0.45	1.15(1.00,2.61)
50-64	1.81(1.33,2.38)	0.79	1.43(1.05,1.88)
65-85	1.71(1.42,2.18)	0.77	1.32(1.09,1.68)
	Women		
30-49	2.81(1.25,6.36)	0.45	1.26(1.00,2.86)
50-64	1.98(1.46,2.61)	0.79	1.56(1.15,2.06)
65-85	1.87(1.56,2.39)	0.77	1.44(1.20,1.84)

Sources: Saydah et al.2002 ; Gregg et al. 2012

The updated relative risks of mortality were applied to the general all-cause mortality rates estimated by the Centers for Disease Control and Prevention [5] to obtain the age-specific mortality rates in persons with and without diabetes. The calculation is shown as below:

$$\begin{aligned} M_{s,a} &= M_{s,a}^{dm} \cdot P_{s,a}^{dm} + M_{s,a}^{ndm} \cdot P_{s,a}^{ndm} \\ &= R_{s,a} \cdot M_{s,a}^{ndm} \cdot P_{s,a}^{dm} + M_{s,a}^{ndm} \cdot P_{s,a}^{ndm} \\ &= M_{s,a}^{ndm} \cdot \left(R_{s,a} \cdot P_{s,a}^{dm} + P_{s,a}^{ndm}\right) \end{aligned} \qquad \text{Eq(1)}$$

Whereas $M_{s,a}$ is sex (s) and age (a) -specific mortality rate in general population. $M_{s,a}^{dm}$ is sex and age-specific mortality rate in diabetes (dm) population. And $P_{s,a}^{dm}$ is sex and age-specific prevalence rate of diabetes. $M_{s,a}^{ndm}$ is sex and age-specific mortality rate in non-diabetes (ndm) population. $P_{s,a}^{ndm}$ is sex and age-specific prevalence rate of non-diabetes. $R_{s,a}$ is the sex and age-specific relative risk of mortality.

Based on Eq(1), $M_{s,a}^{ndm} = M_{s,a} / \left(R_{s,a} \cdot P_{s,a}^{dm} + P_{s,a}^{ndm}\right)$ Eq(2)

2. Life expectancies and life-years lost

With the sex and age-specific mortality rates, we derived the expected life-expectancies for average American adults with diabetes and without diabetes using a traditional period life table method. The formula of calculating life expectancy is shown as below. Life-years lost were defined as the differences between the life expectancies in persons with and without diabetes.

$$LE_a = \sum_0^{\infty} tp_a^t q_{a+t}$$ Eq(3)

Whereas LE_a is the life expectancy at age a. The probability of surviving from age a to age a+t is denoted p_a^t and the probability of dying during age a (i.e. between ages a and a+1) is denoted q_a.

Supplementary Table 3. Presents the expected life expectancies and life-years lost.

Age	Life expectancy in diabetes	Life expectancy in non-diabetes	Life-years lost
	Female		
30	75.2	82.8	7.5
40	76.2	83.2	6.9
50	77.9	83.9	6.0
60	79.9	85.1	5.2
65	81.2	85.9	4.7
70	82.7	87.0	4.3
	Male		
30	71.0	78.6	7.5
40	72.5	79.3	6.7
50	74.8	80.3	5.5
60	77.5	82.2	4.6
65	79.3	83.4	4.1
70	81.1	84.8	3.6

Calculating lifetime excess medical expenditures

The lifetime excess medical expenditure attributed to diabetes was defined as the difference between the lifetime medical expenditure of an average person with diabetes and the one of the identical person but with no diabetes. The calculation is shown in Eq(4).

$$LC_{DM} - LC_{NDM} = \sum_{t=0}^{85-a} C_{a,t}^{dm} \cdot S_{a,t}^{dm} d_{a,t} - \sum_{t=0}^{85-a} C_{a,t}^{ndm} \cdot S_{a,t}^{ndm} d_{a,t}$$ Eq(4)

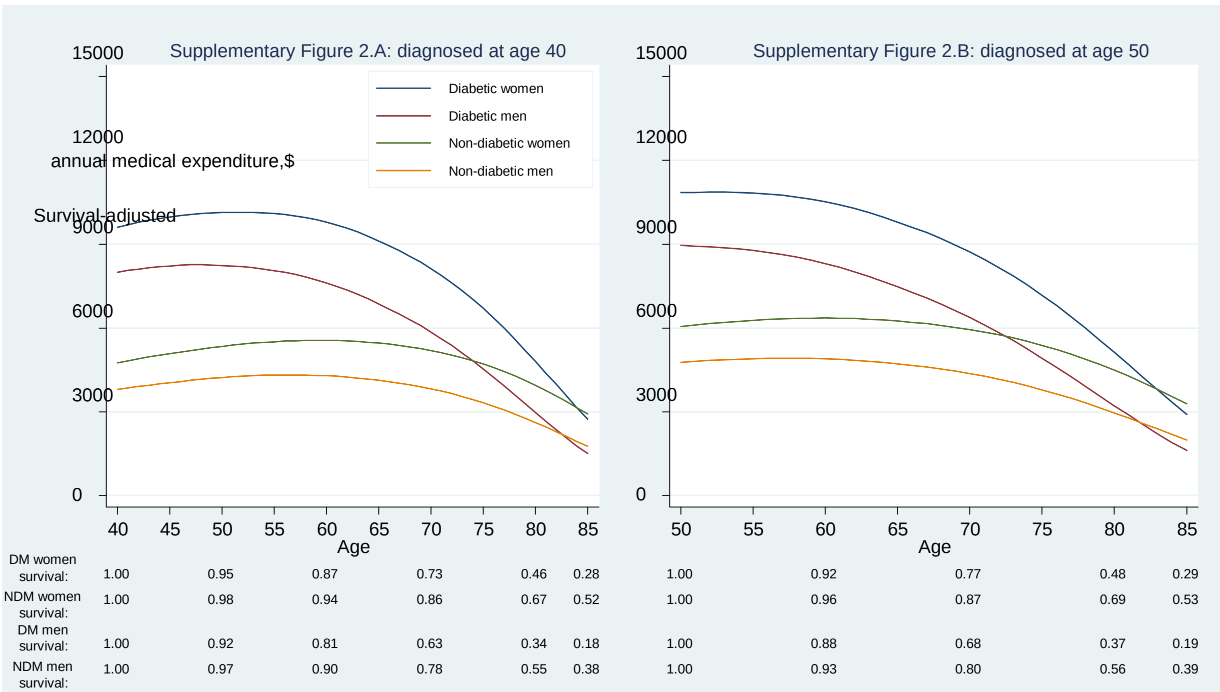
In the Eq(4), LC_{DM} and LC_{NDM} denotes the lifetime medical cost of persons with diabetes and without diabetes respectively. Lifetime medical expenditure can be expressed as the aggregated product of annual medical expenditure and survival rate. $C_{a,t}^{dm}$ denotes the annual medical expenditure at t years after diagnosis of diabetes at age a. The survival rate at that time point is denoted by $S_{a,t}^{dm}$. For non-diabetic person, $C_{a,t}^{ndm}$ denotes the annual medical expenditure at t year since age a. The survival rate at that time point is denoted by $S_{a,t}^{ndm}$. $d_{a,t}$ represents the annual discounting factor, which is equal to 3% per year.

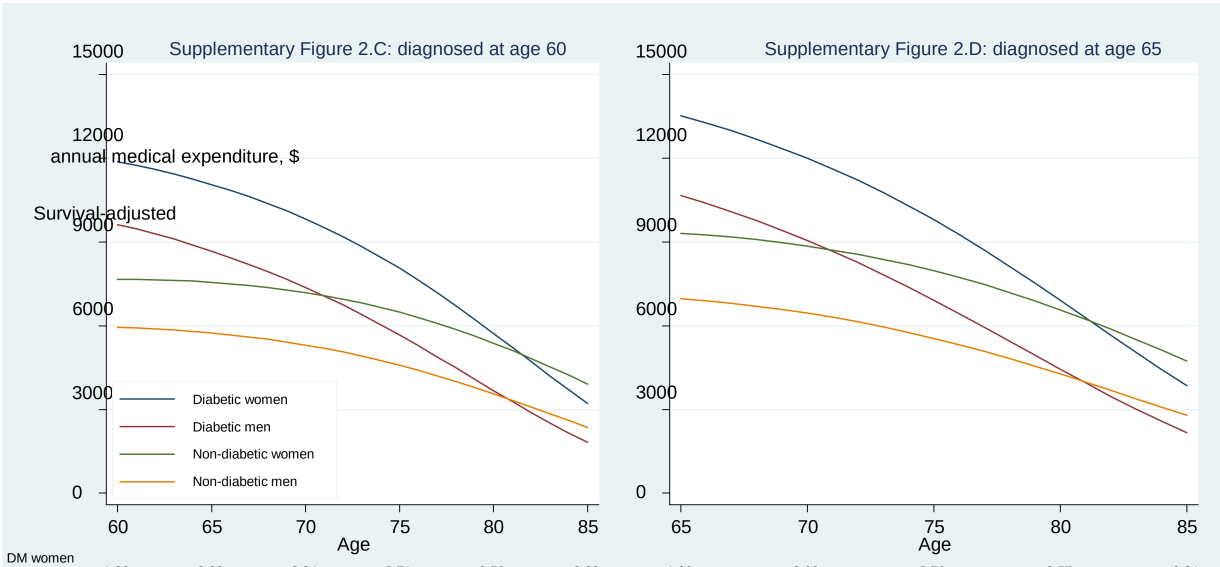
SUPPLEMENTARY DATA

Sensitivity analysis

A more extensive regression specification attenuates the incremental effect of diabetes on expenditures and produced lower excess annual medical expenditure estimates, which leads to lower estimates of lifetime excess medical expenditures. Based on the specification, we re-estimated the annual medical spending and lifetime medical spending. Supplementary Figure 2 shows the re-estimated annual medical expenditures after adjusting for the different rates of survival by diabetes status. Supplementary Figure 3 compare the re-estimated lifetime medical spending with the base-case lifetime estimates.

Supplementary Figure 2. Estimated Annual Medical Expenditures Based on An Alternative Model Specification.

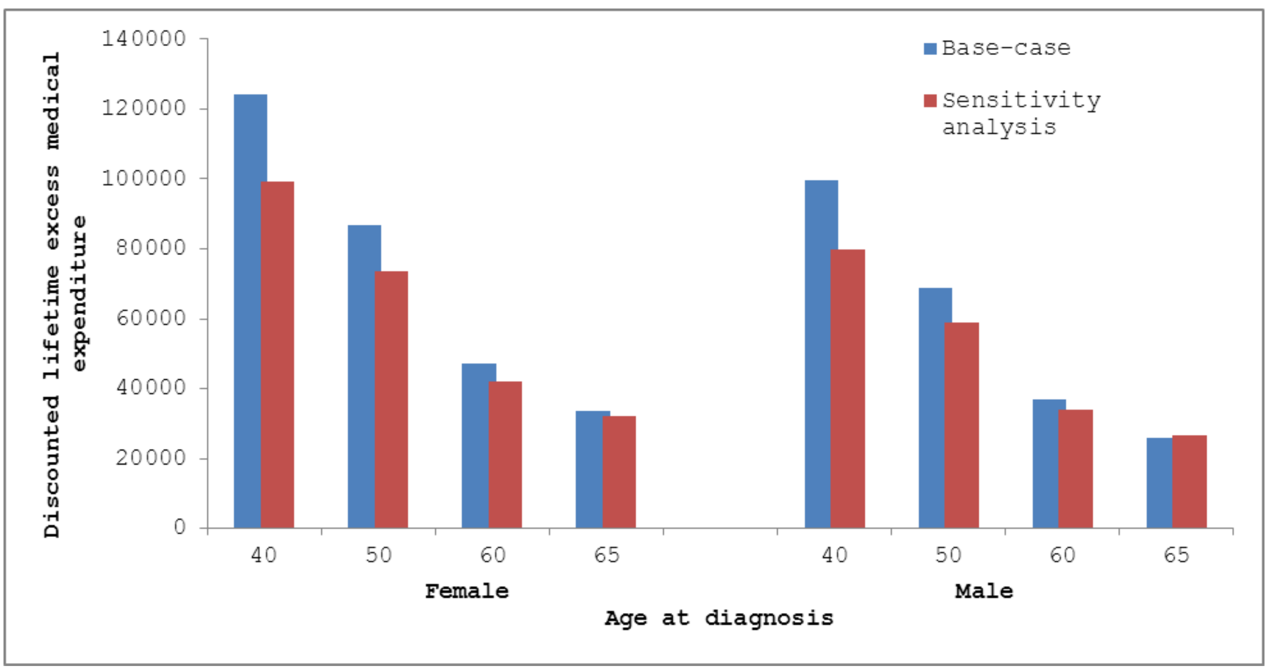




Age (diagnosed at 60)	60	65	70	75	80	85
DM women survival:	1.00	0.93	0.84	0.71	0.53	0.32
NDM women survival:	1.00	0.96	0.91	0.83	0.71	0.55
DM men survival:	1.00	0.90	0.77	0.61	0.42	0.22
NDM men survival:	1.00	0.94	0.86	0.76	0.61	0.42

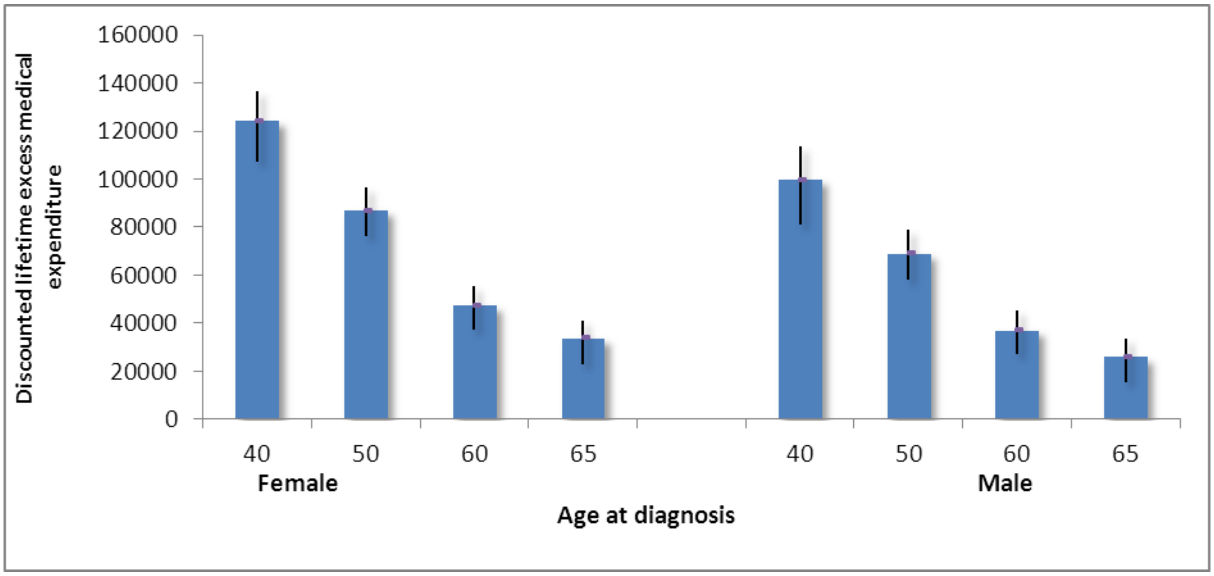
Age (diagnosed at 65)	65	70	75	80	85
DM women survival:	1.00	0.90	0.76	0.57	0.34
NDM women survival:	1.00	0.95	0.87	0.74	0.57
DM men survival:	1.00	0.86	0.68	0.47	0.25
NDM men survival:	1.00	0.92	0.80	0.64	0.45

Supplementary Figure 3. Estimated Lifetime Medical Expenditures Based on An Alternative and Base-case Model Specification.



When the relative mortality risks were varied from the lower bound to the upper bound of the 95% confidence interval, the lifetime excess medical expenditures changed modestly. The lifetime incremental medical expenditures were increased by up to $14,000 when the lower-bound estimates of the relative mortality risk were used. When upper bound estimates were applied, the medical expenditures were reduced by up to $19,000. The results are shown in Supplementary Figure 4.

Supplementary Figure 4 Estimated Lifetime Medical Spending Associated with Diabetes Using the Lower Bound and Upper Bound Estimates of Relative Mortality Risks Associated with Diabetes



Lifetime medical cost by service component

We further stratified the lifetime medical cost by type of medical service, using the same analytical method for total medical cost. The medical services were categorized into four groups: inpatient care, outpatient care, prescription medications, and other medical services. Supplementary Figure 4 presents the percentages of 4 medical components in the lifetime medical cost associated with diabetes by the age of diagnosis.

Supplementary Table 4. Lifetime Medical Spending by Type of Medical Service.

Age at diagnosis	Inpatient care (%)	Prescription medications (%)	Outpatient care (%)	Other medical services (%)	Total
40	36.7	41.7	17.8	3.8	124,600 (100%)
50	35.1	44.3	17.2	3.5	91,200 (100%)
60	33.0	49.7	14.5	2.8	53,800 (100%)
65	30.6	54.8	12.5	2.1	35,900 (100%)

Reference:

1. Manning, W.G. and J. Mullahy, *Estimating log models: to transform or not to transform?* Journal of Health Economics, 2001. **20**(4): p. 461-494.
2. Buntin, M.B. and A.M. Zaslavsky, Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. J Health Econ, 2004. 23(3): p. 525-42.
3. Saydah, S.H., et al., Age and the burden of death attributable to diabetes in the United States. Am J Epidemiol, 2002. 156(8): p. 714-9.
4. Gregg, E.W., et al., 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, 2012. 35(6): p. 1252-7.
5. Arias. E, United States life tables, 2007, in National Vital Statistics Reports, 2011 Edition: Hyattsville, MD.