August 2014

Lifetime Probability of Developing Diabetes in the United States

Samuel Preston [1], Ezra Fishman [1], and Andrew Stokes [2]

[1] Population Studies Center, University of Pennsylvania
[2] Department of Global Health, Boston University

The prevalence of diabetes among adults in the United States has been rising from one period to the next (Cheng et al. 2013; Bullard et al. 2013) and from one birth cohort to the next (Fishman et al. 2014). The increase is evident whether assessed using self-reported data or measures of glycosylated hemoglobin (HbA1c) and fasting plasma glucose (FPG).

Estimates of diabetes incidence and prevalence are typically shown by age interval. By cumulating relevant data across age intervals, it is possible to estimate the likelihood of developing diabetes in the course of life, a readily understood concept that translates aggregate processes down to the level of an individual. Recent estimates of the lifetime probability of developing diabetes were presented in Gregg et al. (2014), following a strategy developed by Narayan et al. (2003). The estimates were based on self-reported data from the National Health Interview Surveys (Gregg et al. 2014). Individuals were asked if they had ever received a diagnosis of diabetes from their physician and if so at what age.

In this paper, we estimate the lifetime probability of developing diabetes in the United States by using HbA1c values, rather than self- reports, as the main criterion for identifying diabetes. These values are derived from National Health

and Nutrition Examination Surveys through 2010. This approach has the advantage of including undiagnosed cases of diabetes, which represent approximately 19% of total diabetes cases (Cowie et al. 2010). It has the additional advantage of using a stable criterion for the identification of diabetes. Ascertainment biases from increased screening, more permissive diagnostic criteria, and greater public awareness of the disease have been cited as contributors to sharply increased diabetes prevalence in the United States, United Kingdom and in Ontario (McBean et al. 2004; Holden et al. 2013; Lipscombe and Hux 2007). Using records from Medicare A and B, McBean et al. (2004) estimate that about a quarter of the 36% increase in diabetes prevalence over the period 1992-93 to 2000-01 in the United States were attributable to changes in the algorithm used to detect individuals with diabetes. Incidence values are more vulnerable to ascertainment bias than are prevalence values because they include among new cases the influx of individuals who were not considered diabetic under old criteria but who are under new.

Data and Methods

We employ data from NHANES III, 1988 to 1994 and from the Continuous NHANES (1999-2010). NHANES is a complex, multi-stage probability sample of the U.S. civilian non-institutionalized population. Participants complete a home interview and then are examined in a mobile examination center, which includes sampling participants' blood for laboratory tests. The National Center for Health Statistics (NCHS) provides extensive documentation of NHANES survey, examination, and laboratory procedures on its website (NCHS 2012).

Characteristics of the NHANES study sample are reported elsewhere (Cheng et al. 2013; Lee et al. 2010).

There were 88,224 individuals examined during our study periods. We exclude individuals below age 20, above age 80, or who were pregnant. We also exclude subjects with missing HbA1c values. The final analytic sample for HbA1c-based measures consists of 40,130 observations, with 7,011 observations from Phase 1 of NHANES III, 7,427 from Phase 2 of NHANES III, 7,778 from NHANES 1999-2002, 7,755 from NHANES 2003-2006, and 10,159 from NHANES 2007-2010.

Definition of diabetes

We define diabetes using the HbA1C criterion, which was first measured in NHANES III. This measure reflects average glycemia over a prolonged period and thus has more intra-subject stability than the leading alternative, measures of fasting plasma glucose (FPG) (Bonora et al. 2011). Furthermore, HbA1c-based measures of diabetes are more strongly associated with cardiovascular disease and death than are FPG-based measures (Selvin et al. 2010). Finally, only 54% as many observations of diabetes status are available in NHANES using FPG as using HbA1c.

In accord with recommendations of the American Diabetes Association (2012), we define diabetes as HbA1c ≥ 6.5% (48 mmol/mol). Because diabetes medication is expected to reduce glycemia, the HbA1c values of medicated persons might not capture their diabetes status correctly. Therefore, individuals who reported taking diabetes medication are also considered diabetic. There were 896 individuals, or 19.2% of the group with diabetes, who reported taking diabetes medication and who had HbA1c ≤ 6.5%.

We calculate each individual's birth year using the equation *Birth cohort = Period - Age*. To ensure large enough age-cohort cells, we analyze cohorts born in ten-year-wide intervals (1910 to 1919, 1920 to 1929, etc.). Using this approach, we obtain a total of 8 ten-year cohorts between 1910-1919 and 1980-1989. For details, see Fishman et al. (2014).

Prevalence is calculated as the proportion of individuals in the given age-cohort cell with diabetes as defined above. Calculations are adjusted for complex survey design using strata and primary sampling units provided by the National Center for Health Statistics, along with survey weights. For HbA1c, we use the final examination weight provided by NCHS; because we pool adjacent data release cycles of Continuous NHANES, we divide the examination weights in Continuous NHANES by 2, as recommended by National Health and Nutrition Examination Survey (2006).

Constructing a Multiple Decrement Life Table for the Diabetes-Free Population

In order to estimate the lifetime probability of developing diabetes, two data series are required: the age-specific rate of developing diabetes for a diabetes-free person; and the age-specific rate of death for a diabetes-free person. From these series, a multiple-decrement life table can be produced that tracks a diabetes-free birth cohort into either death or development of diabetes (Preston, Heuveline, and Guillot 2001: chapter 4).

A detailed description of our estimates of incidence is presented in Fishman et al. (2014). The logic of our approach was to infer incidence by observing changes in diabetes prevalence in a birth cohort as it progressed from one age to the next in successive waves of NHANES. Changes in prevalence in a birth cohort

from one age to the next were attributed to new cases of diabetes (incidence) and mortality differences between those with diabetes and those without. We thus assumed that migration into and out of the cohort does not affect the changes in prevalence; when we tested this assumption by including only native-born individuals, results were similar. Those who develop diabetes are assumed remain in that state until death.

Rather than making separate estimates for each birth cohort, we allowed birth cohorts to "borrow strength" from one another by estimating an age-cohort model of the following type:

$$ln(Y_{ia}) = \alpha + \beta_a X_a + \beta_i X_i ,$$

where Y_{ia} is the proportion of the population in cohort *i* at age *a* with diabetes, X_a is a dummy variable indicating that the observation pertains to age *a*, and X_i is a dummy variable indicating that the observation pertains to cohort *i*. This model implies that there is a standard age-pattern of diabetes prevalence, indicated by the B_a's, that is scaled upwards or downwards by a scalar appropriate to each birth cohort and represented by the B_i's. We used least squares to estimate the parameters of this model, which produced an R^2 of 0.949.

For the current analysis, we use this model to predict prevalence at each age for each birth cohort and use the predicted prevalence to infer age-specific incidence. Our estimate of incidence uses the following formula:

$${}_{10}\bar{\delta}^{O}_{x} = -\frac{1}{5}\ln\left[\frac{{}_{5}\Pi_{x+5}}{{}_{5}\Pi_{x}} \frac{{}_{5}L_{x+5} / {}_{5}L_{x}}{{}^{M}_{5}L^{O}_{x+5} / {}^{M}_{5}L^{O}_{x}}\right], \text{ where}$$

${}_{10}\bar{\delta}^{O}_{x}$ = rate of developing diabetes for a non-diabetic person in the age interval x to x+10,

${}_{5}\Pi_{x}$ = prevalence of non-diabetes at ages x to x+5

$_5L_x$ = person-years lived between ages x and x+5 in a life table for the population (method described below)

$^M_5L^O_x$ = person-years lived between ages x and x+5 in a life table for persons free of diabetes (method described below).

We interpret $_{10}\bar{\delta}^O_x$ as pertaining to the age interval x+2.5 to x+7.5, i.e. the five-year age span at the middle of the ten-year age interval x to x+10 (e.g. ages 22.5 to 27.5, age 27.5 to 32.5, etc.). Values of $_5\Pi_x$ pertaining to persons without diabetes are obtained as the complement of the prevalence values by age and cohort calculated from fitted values in the age-cohort model of prevalence in Fishman et al. (2014). We smooth the incidence series using a three-term moving average of $_{10}\bar{\delta}^O_x$ with the exception of the first (youngest) and last (oldest) values. We estimated these values for age groups 22.5 - 27.5 to 72.5- to 77.5. From ages 20 to 22.5, we used the incidence value for ages 22.5-27.5. For ages above 77.5, we used the incidence value for ages 72.5 to 77.5. We assumed that the incidence rate was constant within each five-year interval. For "boundary" ages (e.g. 27, 32) we used the average of the two adjacent incidence values.

Our mortality estimates are based on life tables for birth cohorts in the United States produced by the Social Security Administration (Bell and Miller 2005). The basic life table that we use for the cohort born in years t to t+10 is the average of male and female life tables in years t and t+10. To adapt this life table to the population without diabetes, we use NHANES to estimate the ratio of $_nL_x$ columns for persons without diabetes to those of the entire population. Life tables for individuals with and without diabetes are estimated using pooled data from NHANES III and Continuous NHANES (1999-2004 waves) cohorts linked to

deaths in the National Death Index through 2006. A discrete hazards model on a person-month file is employed to generate the underlying risks for predicting mortality rates. The model is implemented on baseline ages 20-74. There were 2,903 deaths among 25,971 respondents. Death rates in single-year intervals are derived from the ${}_1L_x$ column of the composite life table for persons without diabetes in a particular birth cohort.

We complete the life table by adding up the two forces of decrement:

${}_1m_x = {}_1m_x^{death} + {}_1m_x^{diabetes}$, and estimating the probability of leaving the diabetes-free population between ages x and x+1 as

${}_1q_x = 1 - \exp(-{}_nm_x)$. The probability of exiting from diabetes is

${}_1q_x^{diabetes} = {}_1q_x * ({}_1m_x^{diabetes}/{}_1m_x)$ and the number of decrements is

$d_x^{diabetes} = {}_1q_x^{diabetes} * l(x)$

Finally, the lifetime probability of developing diabetes for a diabetes-free person age 20 is

$$\frac{\sum_{x=20}^{\omega} d_x^{diabetes}}{l_{20}}.$$

Results

Table 1 and Figure 1 present estimates of the probability that an individual from various birth cohorts will develop diabetes in the course of life. We estimate that a non-diabetic twenty-year old born during 1930-39 has a 30.4% chance of developing diabetes. If diabetes had not developed by age 50 for someone in this cohort, the probability of subsequently developing diabetes declines to 24.2%. Lifetime probabilities rise for later-born cohorts. For someone born a decade later, during 1940-49, the probability that a non-diabetic 20-year old will develop diabetes increases to 37.3%. Finally, for those individuals born between 1960-69,

the most recent cohort investigated, we estimate that the lifetime probability of developing diabetes is greater than half (54.8%).

It should be noted that the number of observed diabetes cases on which the cohort probabilities are based declines as the year of birth advances. The values for the cohort born 1960-69 are based on observations from ages 20 to 50. This is the age interval in which incidence rates are lowest and consequently estimates for this cohort are subject to the greatest error. In contrast, estimates for the cohort born during 1940-49 are based on observations spanning ages 40 to 70, and most of the cases of diabetes that we estimate will occur to this cohort have already been observed.

Reductions in mortality play a role in the increased probability of developing diabetes. If the 1960-69 cohort were subjected to mortality conditions of the 1930-39 cohort instead of to their own conditions, we estimate that 50.9% of 20-year olds in the cohort would develop diabetes, compared to the actual cohort value of 54.8%. So reductions in mortality account for about .039 (16%) of the .244 increase in the lifetime diabetes probability between the 1930-39 and 1960-69 cohorts

Discussion

Estimates by Gregg et al. use retrospective self-reported diagnoses of diabetes recorded in National Health Interview Surveys up to 2011. They estimate that a 20-year old male subject who experiences the incidence and mortality rates observed during 2000-2011 for all of his life had a 40.2% chance of developing diabetes, nearly identical to the 39.6% chance for a 20-year old woman. These values are based on “period rates” applied to a hypothetical

cohort. The most nearly comparable figures from estimates for actual birth cohorts apply to the cohort born during 1940-49, who spent the period 2000-2010 in the ages of highest diabetes incidence from 50 to 70 (Fishman et al. 2014; Gregg et al. 2014). We estimate that 37.3% of this cohort will develop diabetes, close to the two-sex mean of 39.9% estimated by Gregg et al. Our figure is slightly lower than Gregg et al.'s, consistent with incidence rates being somewhat lower in earlier decades. Thus, it appears that estimates of the lifetime probability of developing diabetes based upon self-reports and a clinical measure of diabetes are mutually reinforcing.

References

American Diabetes Association. 2012. Diagnosis and classification of diabetes mellitus. *Diabetes Care*;35 Suppl 1:S64-71.

Bell, Felicitie C. and Michael L. Miller. 2005. Life tables for the Unites States Social Security Area 1900-2100. Actuarial Study No. 120. Office of the Actuary, Social Security Administration.

Bonora E, Tuomilehto J. 2011. The pros and cons of diagnosing diabetes with A1C. *Diabetes Care.* 34 Suppl 2:S184-190.

Bullard KM, Saydah SH, Imperatore G, et al. Secular Changes in U.S. 2013. Prediabetes Prevalence Defined by Hemoglobin A1c and Fasting Plasma Glucose: National Health and Nutrition Examination Surveys, 1999-2010. *Diabetes Care:*36(8):2286-2293.

Cheng YJ, Imperatore G, Geiss LS, et al. 2013. Secular Changes in the Age-Specific Prevalence of Diabetes Among U.S. Adults: 1988-2010. *Diabetes Care.*

Cowie CC, Rust KF, Byrd-Holt DD, et al. 2010. Prevalence of diabetes and high risk for diabetes using A1C criteria in the U.S. population in 1988-2006. *Diabetes Care.* 33(3):562-568.

Fishman, Ezra I., Andrew Stokes and Samuel H. Preston. 2014. The Dynamics of Diabetes among Birth Cohorts in the United States. *Diabetes Care*. Published online February 10, 2014: 1-8.

Gregg, E.W., X. Zhuo, Y.J. Cheng, A.L. Albright, G.M. Venkat Narayan, and T.J. Thompson. 2014. Trends in lifetime risk and years of life lost due to diabetes in the USA, 1985-2011: A modelling study. Lancet-Diabetes-Endocrinology. .Published online August 13, 2014 http://dx.doi.org/10.1016/S2213-8587(14)70161-5.

Holden, S.E., A.H. Barnett, J.R.Peters, S.Jenkin-Jones, C.D. Poole, C.L.Morgan, and C.J.Currie. 2013. The incidence of type 2 diabetes in the United Kingdom from 1991 to 2010. Diabetes, Obesity and Metabolism 15: 844-52.

Lee JM, Pilli S, Gebremariam A, et al. 2010. Getting heavier, younger: trajectories of obesity over the life course. *Int J Obes.* 34(4):614-623.

Lipscombe, Lorraine L. and Janet E. Hux. 2007. Trends in diabetes prevalence, incidence, and mortality in Ontario, Canada 1995-2005: A population-based study. *Lancet* 369: 750-56.

McBean, A. Marshall, D.T. Gilbertson, S. Li, and A.J. Collins. 2004. Differences in Diabetes Prevalence, Incidence, and Mortality among the Elderly of four Racial/Ethnic Groups: Whiles, Blacks, Hispanics, and Asians. *Diabetes Care* 27 (10): 2317-24.

Narayan, K.M.Venkat, James P. Boyle, Theodore J. Thompson, Stephen W. Sorensen, and Dvid F. Williamson. 2003. Lifetime Risk for Diabetes Mellitus in the United States. *Journal of the American Medical Association* 290(14): 1884-1890.

National Health and Nutrition Examination Survey.2006. *Analytic and reporting guidelines*: National Center for Health Statistics.

National Center for Health Statistics, National Health and Nutrition Examination Survey. Survey questionnaires, datasets, and related documentation. [Internet]. 2012; http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm. Accessed July 1, 2013.

Preston, S.H., P. Heuveline and M. Guillot. 2001. *Demography: Measuring and Modeling Population Processes*. Blackwell. London.

Selvin E, Steffes MW, Zhu H, et al. Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. 2010. *N Engl J Med.*;362(9):800-811.

Table 1. Probability of eventually developing diabetes at various ages

	Birth cohort			
Age	1930-39	1940-49	1950-59	1960-69
20	0.304	0.373	0.386	0.548
25	0.302	0.371	0.383	0.545
30	0.298	0.366	0.379	0.539
35	0.292	0.359	0.371	0.528
40	0.283	0.347	0.359	0.512
45	0.266	0.327	0.339	0.484
50	0.242	0.298	0.309	0.442
55	0.211	0.258	0.268	0.384
60	0.174	0.213	0.221	0.317
65	0.140	0.170	0.177	0.254
70	0.110	0.134	0.140	0.200