Small area variation in diabetes prevalence in Puerto Rico

Edward F. Tierney,[1] Nilka R. Burrows,[1] Lawrence E. Barker,[1] Gloria L. Beckles,[1] James P. Boyle,[1] Betsy L. Cadwell,[1] Karen A. Kirtland,[1] and Theodore J. Thompson[1]

ABSTRACT

Objective. *To estimate the 2009 prevalence of diagnosed diabetes in Puerto Rico among adults ≥ 20 years of age in order to gain a better understanding of its geographic distribution so that policymakers can more efficiently target prevention and control programs.*

Methods. *A Bayesian multilevel model was fitted to the combined 2008–2010 Behavioral Risk Factor Surveillance System and 2009 United States Census data to estimate diabetes prevalence for each of the 78 municipios (counties) in Puerto Rico.*

Results. *The mean unadjusted estimate for all counties was 14.3% (range by county, 9.9%– 18.0%). The average width of the confidence intervals was 6.2%. Adjusted and unadjusted estimates differed little.*

Conclusions. *These 78 county estimates are higher on average and showed less variability (i.e., had a smaller range) than the previously published estimates of the 2008 diabetes prevalence for all United States counties (mean, 9.9%; range, 3.0%–18.2%).*

Key words   Diabetes mellitus; prevalence; public policy; Puerto Rico.

In recent decades, the prevalence of diabetes has increased dramatically, both in the United States of America and worldwide (1). Diabetes prevalence increases with age, and in the United States, is higher among minorities (1). This disease is costly in both individual and societal terms. It is the leading cause of kidney failure, nontraumatic lower extremity amputations, and new cases of blindness among adults (1). In 2007, diabetes was the 7th leading cause of death in the United States and its total cost (direct and indirect) was US$ 174 000 million (2).

While national and state statistics on diabetes prevalence have been available for decades (3), data for all United States counties, has only become available recently (4). The Centers for Disease Control and Prevention (Atlanta, Georgia, United States; CDC) website contains maps and downloadable data for each of the years from 2004–2008 (3). Many public health activities and interventions are conducted at the county or local level (5). Whereas state-level estimates cannot, county-level estimates *can* help policymakers identify and target populations that have the highest burden of disease, plan for health services, monitor progress in prevention and control efforts, and allocate scarce resources.

The United States Government has a long history of research into and use of small area estimation methods. In the 1960s, state estimates of disability from a national health survey were published for decades (3), data for all United States counties, has only become available recently (4). The Centers for Disease Control and Prevention (Atlanta, Georgia, United States; CDC) website contains maps and downloadable data for each of the years from 2004–2008 (3). Many public health activities and interventions are conducted at the county or local level (5). Whereas state-level estimates cannot, county-level estimates *can* help policymakers identify and target populations that have the highest burden of disease, plan for health services, monitor progress in prevention and control efforts, and allocate scarce resources.

(6). Since then several approaches for producing small area estimates from complex sample surveys have been proposed; a recent review is available (7).

A number of researchers have used Behavioral Risk Factor Surveillance System (BRFSS) data to estimate diabetes prevalence for small areas. For example, a study conducted in 2005 (8) developed a multilevel model to produce estimates for 32 000 zip code census tracts. In addition to individual-level risk factors, the approach included spatially structured effects for each state in the United States. Another study (4) used a Bayesian multilevel model to estimate county-level diabetes prevalence from 3 years of pooled BRFSS data; and another (9) emphasized the importance of model validation, while also producing county-level estimates. The approach of the present

[1] United States Centers for Disease Control and Prevention, Division of Diabetes Translation, Atlanta, Georgia, United States of America. Send correspondence to Edward F. Tierney, email: ext5@cdc.gov

study builds on the work that utilized a Bayesian multilevel model (4), with a few differences detailed below.

Puerto Rico's unique history and status distinguish it from the rest of the United States in terms of geography, ethnicity, and poverty. However, Puerto Rico's age-structure is similar to that of the United States; e.g., those 65+ years of age in Puerto Rico represent 14.5% of the population, and in the United States they represent 13.0% (10). The other population age-groups, i.e., < 18 years of age, 18–44 years, and 45–64 years, are also similar. In 2009, 45.0% of the population of Puerto Rico lived in poverty, as opposed to 14.3% of the United States as a whole (11). Most adults (92%) in Puerto Rico reported having health-care coverage (health insurance, prepaid plans, or government plans), and 78% reported a routine visit to a doctor in the past year (12).

The present study applies methods similar to those used by the CDC to create county-level estimates of diagnosed diabetes prevalence in the United States (3, 4) to estimate 2009 prevalence among the population $\geq$ 20 years of age in the 78 "municipios" in Puerto Rico, hereafter referred to as counties.

MATERIALS AND METHODS

Description of the data

The BRFSS is an ongoing, state-based, random-digit-dialed telephone survey of non-institutionalized adults $\geq$ 18 years of age in all 50 states, the District of Columbia, and Puerto Rico (12). It provides widely-used state-level estimates of health status, including prevalence of diagnosed diabetes. Respondents are classified as having diagnosed diabetes if they respond "yes" to, "Have you ever been told by a doctor that you have diabetes?" Those who responded "yes," but were told only during pregnancy (gestational diabetes), or those who responded "no" were considered not to have diabetes. Respondents who did not know or refused to answer the question were considered to have missing diabetes status.

Estimates

The study methods accounted for missing data on diagnosed diabetes by poststratification, as explained in An-

nex 1. The survey respondents' "county of residence" is only available through the internal BRFSS files of the CDC. Due to concerns about confidentiality, the internal BRFSS files with information about county of residence are not available to researchers outside of CDC. For respondents with missing county of residence, the CDC uses the county most likely associated with the respondent's telephone number.

For counties with small sample size— in the 2009 BRFSS, 54 of 78 counties had fewer than 50 observations—making direct estimation of diabetes prevalence was impossible. Thus, model-based estimates were derived. A previous study conducted in 2005 of counties in the United States (4) describes methods for deriving Bayesian multilevel model-based estimates of prevalence in small areas, using BRFSS and census data. Except for the differences described below, exactly the same methods were used in the current study of Puerto Rico counties to estimate diabetes prevalence for all counties/county-equivalents in the United States.

Whereas the 2005 study (4) treated age as a three-level categorical variable, here it was treated as a seven-level variable, thus better approximating the true impact of age. Similar to the prior study (4), survey data was treated as observed data collected from a larger set of complete data. However, the other study (4) treated the number of diabetes cases in an area as having a Poisson distribution, while the present study's methods used a binomial distribution. Both methods are valid, although this study's approach results in slightly narrower confidence intervals.

Another difference is that the 2005 study accounted for race/ethnicity. This study of Puerto Rico did not because the commonwealth is much more homogeneous in terms of ethnicity than the United States and accounting for race/ethnicity would have added considerable complexity to the analysis, while having negligible impact on the estimates. The present study did, however, consider several county-level covariates described below that the other study did not; this resulted in somewhat narrower confidence intervals. Finally, boundaries of counties are ultimately arbitrary. Thus, adjacent areas are more likely to be similar than non-adjacent areas. The 2005 study did not explicitly account

for spatial correlation, which the current study did, thereby creating a model that more closely approximates reality.

The study model uses all 11 967 observations from the combined 2008–2010 BRFSS, three county-level covariates, age, sex, and spatially correlated effects (information on which counties were adjacent) to estimate diabetes prevalence for each county. The county-level covariates were obtained from the 2000 United States Department of Agriculture rural-urban continuum codes (range = 1–9; where 1 = metro area with 1 million+ population and 9 = rural with < 2500) (13) and the 2005–2009 Puerto Rico Community Survey (percent of individuals 25 years of age or more who have completed high school and percent below the poverty level in the past 12 months) (14).

The study models allowed the effects of class (14 categories of age and sex combined: age coded 20–29 years of age, 30–39 years, …, 70–79 years, and 80+ for both sexes) to vary by county. Each of the estimates of diagnosed diabetes prevalence is the mean of the posterior predictive distribution for a given county. The 2.5th and 97.5th percentile of the posterior distributions provided the 95% confidence intervals (95%CI). The county estimates were age-standardized to the 2000 United States Standard Population using the direct method. This would eliminate differences in the age structure of the population as a reason for differences in the prevalence of diagnosed diabetes by county.

This study also considered a basic model (7) (with fixed effects for class and a spatially unstructured random effect for county) as a benchmark for assessing an extended model (with random effects that borrowed strength across counties and classes) and explicitly modeled spatial correlation. The two models were compared using the D-statistic (15). This criterion is the sum of two parts: a goodness-of-fit measure or "G" (the squared difference between the data and its posterior predictive mean), plus the expected mean-square predictive error or "P" (the sum over observations of the posterior variance). Models with smaller values of D are preferred, because they tend to have less cumulative systematic and random error.

The study models were assessed for consistency with the data using posterior predictive checking (16). All posterior

distributions were simulated in Win-BUGS (17). A burn-in of 5 000 was used, and then a single chain was monitored for 20 000 iterations (see Annex 1 for a complete description of the Methods). Note that, Bayesian *P* values are not like classical *P* values. Values close to 0 or 1 indicate poor fit and values close to 0.5 indicate a good fit. A more extensive explanation is available (18).

In order to help describe and understand diabetes prevalence distribution in Puerto Rico, natural breaks (also called Jenks) were created to map the point prevalence estimates by county. The Jenks Natural Breaks Classification is a system to break data into different classes (19). This method attempts to reduce the variance within classes and maximize variance between classes.

RESULTS

Description of data

Table 1 provides details on the number of counties and the number of adults ≥ 20 years of age in the 2009 Census (estimated population size), 2009 BRFSS, and 2008–2010 BRFSS for Puerto Rico (numbers of survey participants). The number of respondents in the BRFSS who self-reported diabetes is also given. The BRFSS provides diabetes status for roughly 0.14% (4 143/2 883 981) of the 2009 population ≥ 20 years in Puerto Rico.

For the county-level covariates, the mean percentage in poverty was 49.5% (standard deviation [SD] 8.6) and the mean percentage completing high school was 63.3% (SD 5.7). The median value for rural/urban code was 1 (inter-quartile range 1–3).

Table 2 provides Bayesian *P* values from posterior predictive checking and provides a comparison of the two models using the model selection criterion. Both models are consistent with the data, and the extended model is preferred over the basic model, due to its smaller value of D. Therefore, this study presents estimates from the extended model.

Estimates

Unadjusted and age-adjusted estimates of diabetes prevalence and 95% CIs for all counties appear in Table 3. The mean unadjusted estimate was 14.3

TABLE 1. Descriptive statistics of the 2009 United States Census and the 2009 and 2008–2010 Behavioral Risk Factor Surveillance System (BRFSS) data for Puerto Rico

	Census[a] 2009	BRFSS[b] 2009	BRFSS 2008–2010
Counties	78	78	78
Adults ≥ 20 years of age	2 883 981	4 143	11 967
Adults ≥ 20 years of age with diabetes	NA[c]	843	2 403

[a] Estimated population size.
[b] Numbers of survey participants.
[c] Not applicable; the United States Census does not collect data on diabetes status.

TABLE 2. Model checking and model selection criterion for diabetes prevalence in Puerto Rico by county, 2009

Model	Bayesian *P*-value[a]	Goodness of fit G	Predictive error P	D[b]
Basic	0.90	1 913	2 433	4 345
Extended	0.84	1 449	2 785	4 234

[a] 0.1–0.9 indicates reasonable consistency with the data.
[b] Smaller is better.

TABLE 3. Unadjusted and age-adjusted diabetes prevalence in Puerto Rico, by county, 2009

County	Unadjusted %	95% CI[a]	Adjusted %[b]	95% CI
Adjuntas	14.7	11.3,18.6	13.6	10.4,17.2
Aguada	14.6	11.8,17.7	14.3	11.7,17.2
Aguadilla	15.0	12.4,17.7	13.4	11.1,15.9
Aguas Buenas	15.1	12.2,18.5	13.8	11.1,17.0
Aibonito	12.5	10.0,15.3	11.7	9.3,14.3
Añasco	14.8	11.9,18.2	13.5	10.8,16.5
Arecibo	17.9	15.4,20.5	15.8	13.5,18.3
Arroyo	14.1	10.8,18.0	12.9	9.8,16.6
Barceloneta	16.4	13.0,20.5	15.4	12.2,19.3
Barranquitas	13.8	11.0,17.1	13.9	11.1,17.3
Bayamón	12.6	11.0,14.3	11.2	9.7,12.8
Cabo Rojo	15.2	12.5,18.1	13.2	10.8,16.0
Caguas	13.9	12.1,15.9	12.1	10.5,13.9
Camuy	13.6	10.9,16.7	13.0	10.4,16.0
Canóvanas	13.1	10.6,15.9	12.8	10.4,15.5
Carolina	14.0	12.1,16.0	12.3	10.6,14.2
Cataño	13.5	10.5,17.0	12.4	9.6,15.6
Cayey	14.4	11.9,17.1	12.7	10.5,15.2
Ceiba	12.5	9.7,15.8	12.5	9.7,15.8
Ciales	16.1	12.8,20.0	15.2	12.1,18.9
Cidra	13.4	10.9,16.3	13.0	10.6,15.8
Coamo	14.7	11.8,17.9	13.9	11.1,17.0
Comerío	13.9	10.8,17.3	13.3	10.4,16.5
Corozal	12.9	10.2,15.8	12.4	9.8,15.2
Culebra	14.7	8.9,21.9	12.2	7.4,18.4
Dorado	12.1	9.4,15.2	11.3	8.8,14.3
Fajardo	14.0	11.1,17.3	12.3	9.6,15.4
Florida	15.7	12.2,20.1	14.9	11.6,19.1
Guánica	15.2	11.9,18.9	13.5	10.5,16.9
Guayama	14.2	11.3,17.8	13.9	11.0,17.4
Guayanilla	15.5	12.0,19.6	14.4	11.1,18.4
Guaynabo	11.1	9.0,13.2	9.8	8.0,11.8
Gurabo	12.8	10.2,15.8	13.0	10.5,16.0
Hatillo	14.0	11.2,17.1	13.4	10.7,16.4
Hormigueros	14.5	11.2,18.3	11.9	9.1,15.3
Humacao	13.2	10.7,15.9	11.9	9.6,14.3
Isabela	15.6	12.8,18.7	13.9	11.5,16.8
Jayuya	13.6	10.3,17.6	13.4	10.1,17.3
Juana Diaz	14.6	11.9,17.9	14.4	11.8,17.6
Juncos	13.1	10.5,16.0	13.0	10.5,15.9
Lajas	18.0	14.6,21.7	15.0	12.1,18.4
Lares	15.2	12.5,18.4	13.7	11.2,16.6

(Continues)

TABLE 3. (Continued)

County	Unadjusted %	95% CI [a]	Adjusted %[b]	95% CI
Las Marías	14.7	10.9,19.1	13.8	10.1,18.0
Las Piedras	14.8	12.0,18.3	14.6	11.8,18.0
Loíza	12.4	9.5,15.7	12.9	9.9,16.2
Luquillo	14.3	11.1,18.1	12.8	9.9,16.3
Manatí	15.9	13.1,19.3	14.6	11.9,17.8
Maricao	14.7	11.0,18.8	13.9	10.4,17.9
Maunabo	16.9	12.8,21.7	15.0	11.3,19.5
Mayagüez	16.0	13.8,18.3	13.4	11.4,15.7
Moca	13.8	11.3,16.7	13.6	11.1,16.4
Morovis	13.6	10.7,16.8	14.3	11.3,17.6
Naguabo	12.7	9.8,15.9	11.9	9.2,15.0
Naranjito	12.9	10.2,15.7	12.5	9.9,15.3
Orocovis	14.4	11.3,17.9	14.0	11.0,17.4
Patillas	16.1	12.9,19.8	14.1	11.3,17.5
Peñuelas	14.5	11.3,18.6	14.8	11.4,18.9
Ponce	14.8	12.9,16.9	13.3	11.4,15.3
Quebradillas	15.7	12.4,19.5	15.1	11.9,18.8
Rincón	17.0	13.2,21.3	14.4	11.2,18.2
Río Grande	14.1	11.4,17.3	13.2	10.6,16.3
Sabana Grande	15.2	12.3,18.5	13.3	10.7,16.3
Salinas	14.2	11.1,17.7	13.3	10.4,16.7
San Germán	16.4	13.6,19.6	13.8	11.3,16.7
San Juan	12.4	11.0,13.8	10.9	9.6,12.3
San Lorenzo	14.1	11.2,17.4	13.6	10.8,16.8
San Sabastián	16.9	14.0,20.2	14.4	11.9,17.3
Santa Isabel	13.2	9.7,17.3	12.8	9.5,16.8
Toa Alta	9.9	7.8,12.2	10.9	8.7,13.4
Toa Baja	11.9	9.6,14.4	11.3	9.1,13.8
Trujillo Alto	12.8	10.4,15.6	12.4	10.1,15.1
Utuado	16.6	13.6,20.0	14.6	12.0,17.7
Vega Alta	14.2	11.4,17.6	13.2	10.5,16.3
Vega Baja	12.9	10.3,15.7	12.0	9.6,14.6
Vieques	17.0	12.0,22.8	13.4	9.3,18.2
Villalba	11.9	9.2,14.9	12.5	9.7,15.7
Yabucoa	14.1	11.3,17.3	13.5	10.9,16.7
Yauco	13.4	10.8,16.2	12.2	9.8,14.9

[a] Confidence Interval.
[b] Adjusted-Age adjusted to 2000 United States Standard Population.

(range, 9.9%–18.0%). The average width of the 95% CI was 6.2% (range, 2.8%–13.0%). The mean age-adjusted estimate was 13.3% (range, 9.8%–15.8%). Adjusted and unadjusted estimates differed little, probably due to similarity of the age-structure between Puerto Rico and the United States. The unadjusted diabetes prevalence by county for Puerto Rico appears in Figure 1.
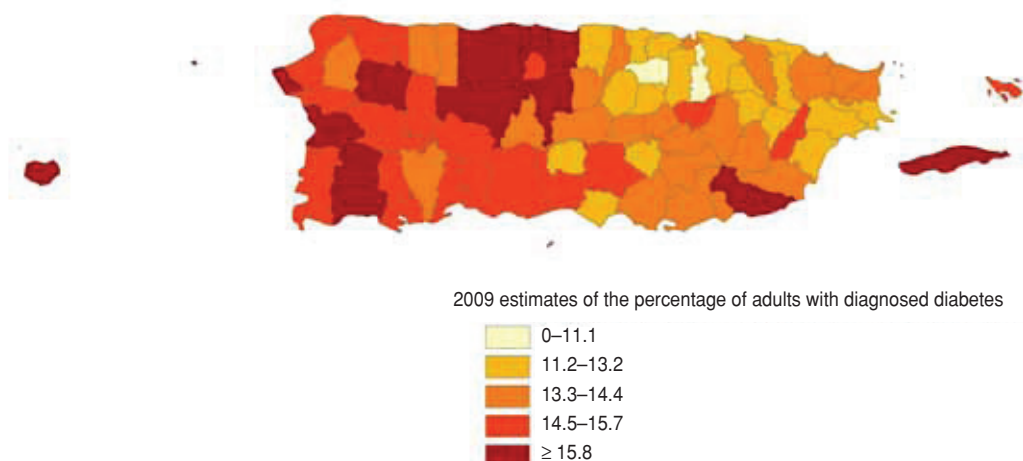
DISCUSSION

This study is believed to hold the first estimates of diabetes prevalence by county in Puerto Rico. The 78 county estimates are higher on average and showed less variability (i.e., had a smaller range), than the previously published estimates of 2008 diabetes prevalence for all United States counties (3.0%–18.2%; mean, 9.9%) (3). This smaller range likely reflects the greater homogeneity of Puerto Rico's population with respect to ethnicity and the smaller number of estimates (78 counties versus 3 141 counties/county-equivalents).

A visual inspection of the map in Figure 1 suggests that, in general, low prevalence counties were concentrated on the eastern half of Puerto Rico and high prevalence counties on the western half. Counties in Puerto Rico's western half tend to have a higher percentage of the population below poverty status, a higher average age household size, a lower percentage of those 25 years of age and over who have completed high school, and lower median household income (14). These factors may have contributed to the observed geographic variation in diabetes prevalence.

In the United States, during 1999–2009, the crude prevalence of diagnosed diabetes increased 61% (4.1% to 6.6%), while the age adjusted prevalence increased 48% (4.2% to 6.2%). During the same period in Puerto Rico, the crude prevalence increased 27% (9.2% to 12.6%), while the age-adjusted prevalence increased 25% (10.1% to 12.6%) (3).

One explanation for the high prevalence of diagnosed diabetes on the island of Puerto Rico is the high rate of inci-

FIGURE 1. Unadjusted prevalence of diabetes among adults 20 years of age and older in Puerto Rico by county, 2009

2009 estimates of the percentage of adults with diagnosed diabetes

- 0–11.1
- 11.2–13.2
- 13.3–14.4
- 14.5–15.7
- ≥ 15.8

dent cases of diagnosed diabetes (20). In 2005–2007, of all states within the United States, only West Virginia had a higher average annual unadjusted incidence of diabetes (13.3/1 000) than Puerto Rico (11.9/1 000). The age-adjusted rate in Puerto Rico (12.8/1 000) was the highest of all states and territories reported (20). Though Puerto Rico has a high prevalence of overweight and obesity (12), these conditions are unlikely to explain the high prevalence of diabetes, since many states had equal or greater levels of overweight and obesity.

Although Puerto Rico's population is overwhelmingly of Hispanic ethnicity, there is also a unique mixture of people of African, Chinese, European (French, German, Irish, Italian, Scottish, and Spanish), Indigenous American, and even, Lebanese descent; some of these populations are at high risk for diabetes (21). Furthermore, Puerto Ricans are unique among inhabitants of the Caribbean in that they have a higher than expected prevalence of type 1 diabetes among children under 15 years of age (6.4/10 000 children versus Trinidad and Tobago's 2.8) (22).

A diet high in carbohydrates and fried foods combined with obesity may contribute to the risk of type 2 diabetes and add to Puerto Rico's high prevalence of diabetes. Fast-food restaurants began appearing in Puerto Rico in the late 1950s and early 1960s. In 2005, a newspaper article reported that 77% of locals visited the 2 000 fast-food restaurants in Puerto Rico often, and restaurant sales reached over US$ 1 000 million (23).

In 2009, Puerto Rico had a greater estimated prevalence of not engaging in physical activity (another factor that contributes to developing type 2 diabetes) than any United States state: 72% responding "no" to whether they engaged in 30 minutes or more of moderate physical activity 5 or more days per week, or vigorous physical activity for 20 minutes or more 3 or more days per week (12).

Family history of diabetes may also contribute to the high prevalence of diabetes in Puerto Rico. For example, among Puerto Ricans living in Chicago in 2002–2003, there was a high overall frequency of having a family history of diabetes (43.2%). Diabetes prevalence among those with a family history of diabetes was 32.5%, and among those without a family history, only 12.3% (24). It is unclear what role the high health-

care coverage in Puerto Rico may play in the high prevalence of diagnosed diabetes through access to care and earlier diagnosis.

In general, the prevalence of diabetes among Puerto Ricans living on the mainland United States is higher than those living on the island of Puerto Rico. For example, among Puerto Ricans 18–75 years of age, living in Chicago in 2002–2003 (24), the unadjusted prevalence was 20.8%, i.e., almost twice the unadjusted prevalence among those 18 years of age or more living in Puerto Rico (10.4 in 2002 and 10.7 in 2003) (3). It was higher among those born in the United States (25.1%) than those born in Puerto Rico (15.7%) (24). There was also a high overall prevalence of obesity (33.2%) among Puerto Ricans living in Chicago (24). Since obesity is a strong risk factor for diabetes, it is not surprising that the prevalence of diabetes was 33.4% among obese individuals and 11.8% among nonobese individuals in this study.

In a study conducted in 1991–1997, the prevalence of diagnosed diabetes was 34.3% among Puerto Ricans 60–96 years of age residing in Massachusetts (25). This compares with prevalence rates that ranged from 24.7%–28.1% among Puerto Rican island residents ≥ 65 years of age in 1994–1997 (3). Another study found a higher prevalence of diabetes among Puerto Ricans living in New York City in 1999–2000 (14.9%) than among those living in Puerto Rico (10.5%) (26).

In a study using data from the 2000–2005 National Health Interview Survey, the overall prevalence of diabetes among Puerto Ricans was 11%, with a prevalence of 6% among those born in the United States, and 15% among those born in Puerto Rico (27). This compares with yearly rates of diabetes prevalence for 2000–2005 that ranged from 9.3%–11.7% among those residing in Puerto Rico (3). In this study, those born in Puerto Rico were, on average, 12.8 years older than those born in the United States.

In the Hispanic Health and Nutrition Examination Survey (HHANES), conducted in 1982–1984, the prevalence of diagnosed diabetes among Puerto Ricans living in the United States was 2% for those 20–44 years of age and 14.3% for those 44–74 years of age. This was comparable to the survey conducted in 1984 by the Puerto Rican Department of Health that found a 1.9% prevalence among those 25–44 years of age; 12.8%

for those 45–64 years of age, and 18.9% for those ≥ 65 years (28). Since the first two age groups in HHANES and the Puerto Rican survey were so similar, the disparity in diabetes prevalence between mainland and island Puerto Ricans may have begun in the early/mid 1980s. This would be consistent with the increase in obesity in the United States and its contribution to the rise in type 2 diabetes.

The findings above show that diabetes and obesity (a major risk factor for diabetes) are likely to be more common among migrants to the mainland United States than among residents of Puerto Rico. Some migrants may be exposed to living conditions associated with unhealthy lifestyles and reduced access to health care (29). Such migrants may return home when they develop severe illnesses, such as diabetes (29). However, the extent to which return migration contributes to the high prevalence of diabetes observed in Puerto Rico is not known.

The present study had a number of strengths. First, the Bayesian methods were applied to estimate the posterior distribution of diabetes prevalence by county, which allowed for a more efficient use of the available data than that of competing methods. The study used model-based estimates with random effects that overcome the limitations of other estimation approaches, such as demographic methods, synthetic estimators, and composite estimators. This allowed information from the Census and the BRFSS to be combined to predict diabetes status for all individuals in the 2009 Census who had not participated in the 2009 BRFSS. The use of random effects allowed the analysis to borrow strength over county and class (combination of sex and age). That is, the estimate for a certain county is improved by using information (data) from other counties and classes. Second, by combining 3 years of BRFSS data, the sample size was larger.

Study limitations

First, this study used estimates of diabetes prevalence that depended on self-reported diagnosed diabetes. This is a potentially severe limitation, if one is concerned with total diabetes—diagnosed plus undiagnosed—rather than simple diagnosed diabetes. While the fraction of total cases of diabetes in Puerto Rico is unknown, 27% of the total

cases of diabetes in the United States are undiagnosed (2). However, the estimates provided here can be treated as a lower bound to the prevalence of diabetes, whether diagnosed or not. Moreover, the high health care coverage in Puerto Rico may influence the un-diagnosed fraction on the island.

Second, the study did not distinguish between type 1 and type 2 diabetes. While the etiologies of type 1 and type 2 diabetes are totally different, recommendations to control blood sugar, blood pressure and cholesterol, and engage in physical activity are the same for both types. Thus, it is not clear how much of a limitation this is.

Third, telephone surveys may be biased in populations with low telephone coverage or high cellphone use. However, the study estimates were arrived at by weighting by population totals (post-stratification), and simultaneously corrects for non-response (see Annex 1).

Finally, telephone surveys may be subject to recall and social desirability bias. While these sources of bias are a minor factor in reporting diabetes in the United States (30, 31), any such study in Puerto Rico does not exist to the best of the authors' knowledge.

Conclusions

The most noteworthy finding of this report was the high prevalence of diagnosed diabetes throughout Puerto Rico. The counties with the lowest estimated prevalence still have a very high estimated prevalence. Public health officials in Puerto Rico need to be aware of this, because people with diabetes often develop expensive and debilitating complications, such as blindness, lower extremity amputation, and kidney disease, as they age (1, 2). Though the study did not measure incidence on a county level, given the high county-level prevalence and high Puerto Rico-wide prevalence, it may be supposed that incidence is also high in many or all counties.

Future research is needed to identify potentially modifiable county-level characteristics of the built environment known to influence levels of diabetes-related health behaviors (32). These structural characteristics might include availability of grocery stores and facilities for physical exercise (presence of sidewalks, low levels of traffic, neighborhood safety, and level of crime). Identifying these characteristics will have utility for policymakers as they aim to prevent future cases of diabetes in Puerto Rico by encouraging healthier diets and less sedentary lifestyles.

Poverty is also a well-established risk factor for diabetes (33–36), and is extremely widespread in Puerto Rico. To prevent future suffering from diabetes in Puerto Rico and the United States as a whole, steps to reduce poverty might be necessary. This is an ambitious goal, but without reducing poverty, an even higher human price in years to come should be expected.

Disclaimer. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the United States Centers for Disease Control and Prevention.

REFERENCES

1. United States Centers for Disease Control and Prevention. Diabetes report card 2012. Atlanta, Georgia: CDC, United States Department of Health and Human Services; 2012. Available from: http://www.cdc.gov/diabetes/pubs/pdf/DiabetesReportCard.pdf Accessed 20 November 2012.

2. United States Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, Georgia: CDC, United States Department of Health and Human Services. Available from: http://www.cdc.gov/diabetes/pubs/factsheet11.htm. Accessed 20 November 2012.

3. United States Centers for Disease Control and Prevention. National Diabetes Surveillance System. Atlanta, Georgia: CDC, United States Department of Health and Human Services; 2011. Available from: http://www.cdc.gov/diabetes/surveillance/index.htm. Accessed 20 November 2012.

4. Cadwell BL, Thompson TJ, Boyle JP, Barker, LE, 2010. Bayesian small area estimates of diabetes prevalence by U.S. county, 2005. JDS. 2010;8:173–88.

5. United States Institute of Medicine, Committee for the Study of the Future of Public Health. The future of public health: summary and recommendations. Washington, D.C.: National Academy Press; 1988.

6. Levy PS, French DK. Synthetic estimation of state health characteristics based on the Health Interview Survey. Vital Health Stat 2. 1977;75:1–22. Available from: http://www.cdc.gov/nchs/data/series/sr_02/sr02_075.pdf. Accessed 20 November 2012.

7. Rao JNK. Small area estimation. Hoboken, New Jersey: Wiley Interscience; 2003.

8. Congdon P, Lloyd P. Estimating small area diabetes in the U.S. using the Behavioral Risk factor Surveillance system. JDS. 2010;8:235–252.

9. Srebotnjak T, Mokdad AH, Murray CJL. A novel framework for validating and applying standardized small area measurement strategies. Popul Health Metr. 2010;8:26.

10. United States Government. U.S. Census: age and sex composition: 2010–2010 census briefs. Available from: http://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf Accessed 20 November 2012.

11. United States Government. American community survey–briefs: poverty: 2008 and 2009. Available from: http://www.census.gov/prod/2010pubs/acsbr09-1.pdf. Accessed 20 November 2012.

12. Li C, Balluz LS, Okoro CA, Strein TW, Lin J-MS, Town M, et al. Surveillance of certain health behaviors and conditions among states and selected local areas–Behavioral Risk Factor Surveillance System (BRFSS), United States, 2009. MMWR. 2011;60(No. SS #9):23–194. (Tables: 4, 7, 22, 40, 43).

13. United States Department of Agriculture, Economic Research Service. Rural-urban continuum codes. Available from: http://www.ers.usda.gov/Data/RuralUrbanContinuumCodes/. Accessed 20 November 2012.

14. American community survey, 2005–2009.5-year estimates. Available from: http://factfinder.census.gov/servlet/DatasetMainPageServlet?_program=ACS&_submenuId=datasets_1&_lang=en&_ts. Accessed 20 November 2012.

15. Gelfand AE, Ghosh SK. Model choice: A minimum posterior predictive loss approach. Biometrika. 1998;85(1):1–11.

16. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchial models. Cambridge, United Kingdom: Cambridge University Press; 2007.

17. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS–a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput. 2000;10(4):325–37.

18. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. Second ed. Washington, D.C.: Chapman & Hall; 2004.

19. Jenks GF, Caspall F. Errors on choroplethic maps: definition, measurement, reduction. Ann Assoc Am Geogr. 1971;61:217–44. Available from: http://www.tandfonline.com/

doi/pdf/10.1111/j.1467–8306.1971.tb00779.x. Accessed 4 December 2012.

20. Kirtland KA, Li YF, Geiss LS, Thompson TJ. State-specific incidence of diabetes among adults–participating states, 1995–1997 and 2005–2007. MMWR. 2008;57(43):1169–73.

21. Haddock L, Torres de Conty I. Prevalence rates for diabetes mellitus in Puerto Rico. Diabetes Care. 1991;14:676–84.

22. Barcelo A, Rajpathak S. Incidence and prevalence of diabetes mellitus in the Americas. Rev Panam Salud Publica. 2001;10(5):300–8.

23. Rosa T. Puerto Rico's appetite for convenience. Puerto Rico Herald, 28 July 2005. Available from: http://www.puertorico-herald.org/issues2/2005/vol09n30/CBFast Food.html. Accessed 20 November 2012.

24. Whitman S, Silva A, Shah AM. Disproportionate impact of diabetes in Puerto Rican communities in Chicago. J Community Health. 2006;31(6):521–31.

25. Tucker KL, Bermundez OI, Castaneda C. Type 2 diabetes is prevalent and poorly controlled among Hispanic elders of Caribbean origin. Am J Public Health. 2000;90(8):1288–93.

26. Ho GYF, Qian H, Kim MY, Melnik TA, Tucker KL, Jimenez-Valasquez IZ. , et al. Health disparities between island and mainland Puerto Ricans. Rev Panam Salud Publica. 2006;19(5):331–9.

27. Pabon-Nau LP, Cohen A, Meigs JB, Grant RW. Hypertension and diabetes prevalence among U.S. Hispanics by country of origin: The National Health Interview Survey, 2000–2005. J Gen Intern Med. 2010;25(8):847–52.

28. Flegal KM, Ezzati TM, Harris MI, Haynes SG, Juarez RZ, Knowler WC. Prevalence of diabetes in Mexican Americans, Cubans, and Puerto Ricans from the Hispanic Health and Nutrition Examination Survey, 1982–1984. Diabetes Care. 1991;14(Suppl 3):628–38.

29. Davies AA, Borland RM, Blake C, West HE. The dynamics of health and return migration. PLoS Medicine. 2011;8(6):e1001046.

30. Bowlin SJ, Morrill BD, Nafziger AN, Lewis C, Pearson TA. Reliability and changes in validity of self-reported cardiovascular disease risk factors using dual response: the Behavioral Risk Factor Survey. J Clin Epidemiol. 1996;49(5):511–7.

31. Saydah SH, Geiss LS, Tierney E, Benjamin SM, Engelgau M, Brancati F. Review of the performance methods to identify diabetes cases among vital statistics, administrative, and survey data. Ann Epidemiol. 2004;14(7):507–16.

32. Castro FG, Shaibi GQ, Boehm-Smith E. Ecodevelopmental contexts for preventing type 2 diabetes in Latino and other racial/ethnic minority populations. J Behav Med. 2009;32(1):89–105.

33. Dinca-Panaitescu S, Dinca-Panaitescu M, Bryant T, Daiski I, Pilkington B, Raphael D. Diabetes prevalence and income: results of the Canadian Community Health Survey. Health Policy. 2011;99(2):116–23.

34. Maty SC, Everson-Rose SA, Haan MN, Raghunathan TE, Kaplan GA. Education, income, occupation, and the 34-year incidence (1965–99) of type 2 diabetes in Alameda County Study. Int J Epidemiol. 2005;34(6):1274–81.

35. Kanjilal Sanjat, Gregg EW, Cheng YJ, Zhang P, Nelson D, Mensah G, Beckles GLA. Socioeconomic status and trends in disparities in 4 major risk factors for cardiovascular disease among US adults, 1971–2002. Arch Intern Med. 2006;166(21):2348–55.

36. Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. Int J Epidemiol. 2011;40(3):804–18.

37. Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst of Stat Math. 1991;43(1):1–59.

38. Gelfand AE, Hills SE, Racine-Poon A, Smith AFM. Illustration of Bayesian inference in normal data models using Gibbs sampling. J Am Stat Assoc. 1990;85(412):972–85.

ANNEX 1. Supplementary material for the study methods

1. Data

Data from the 2008, 2009, and 2010 BRFSS surveys were combined. For each of the 78 counties in Puerto Rico, sampled persons were cross-classified by age group (20–29 years of age, 30-39, …, 70–79, 80+) and sex (male, female). This cross-classification resulted in 14 classes per county. The number of people sampled in each class that have diabetes can be determined. Specifically, let:

n_{ij} = the number of sampled people in county, i class $j = 1, …, 14$
y_{ij} = the number of sampled people with diagnosed diabetes in county i, class j

In some years, the n_{ij} is some counties will = 0. For these, the corresponding y_{ij} will also = 0. The United States Census Bureau publishes population estimates by demographic characteristics (unit-level auxiliary information) for all counties; the Census provides no information on diabetes status.

The 2009 Census county projections were used to obtain estimates for the number of persons in each age and sex group used to cross-classify the BRFSS data. Let, N_{ij} = the estimated number of people in county i, class $j = 1, . . . , 14$, in 2009. Variability in Census projections was ignored.

The county-level covariates were obtained from the 2000 United States Department of Agriculture (USDA–population density) (13) and the 2005–2009 Puerto Rico Community Survey (percent of population 25 years of age and older who have completed high school; percent of population below poverty level in past 12 months) (14). The county-level covariates was centered and scaled by subtracting the overall mean from each and dividing the result by twice the standard error (SE).

2. Regression model

A Bayesian multilevel model was fit to the combined BRFSS data. This model relates observed quantities to other variables of interest. In particular:

$$y_{ij} \sim \text{Binomial}(p_{ij}, n_{ij}); \ i = 1,...,78 \text{ and } j = 1,...,14$$

where p_{ij} = the prevalence of diagnosed diabetes in county i, class j. The regression model includes the following terms:

(a) logit link function: $\log (p_{ij}/(1 - p_{ij}))$
(b) a separate intercept for each class (age by sex group) α_j; $j = 1,...,14$
(c) effects of county-level predictors by sex δ_{ls}; $l = 1,2,3$ and $s = 1,2$. Predictors include rural-urban continuum code xi1, percent of people 25 years and older who have completed high school xi2, and percent of people below poverty level in past 12 months xi3.
(d) spatially correlated effects by county and class: v_{ij}; $i = 1,...,78$ and $j = 1....,14$
(e) spatially unstructured effects by county and class: μ_{ij}; $i = 1,...,78$ and $j = 1....,14$

Parameters under (b) and (c) are "fixed" effects, while (d) and (e) are random effects that borrow strength over county and class. Parameters under (d) are modeled via multivariate normal conditional autoregressive priors (of dimension 14) (37). These parameters allow spatial correlation of a county with its neighboring counties. Parameters under (e) are modeled via multivariate normal priors (of dimension 14) (38). These parameters allow for correlated effects across class without any form of spatial correlation over county. Thus, the regression model is:

$$\text{logit}(p_{ij}) = \alpha_j + \delta_{1[j]}x_{i1} + \delta_{2[j]}x_{i2} + \delta_{3[j]}x_{i3} + v_{if} + \mu_{ij}$$

Where $[j]=1$ if class j contains males and $[j]=2$ if class j contains females.

A basic model was considered as a benchmark to assess the study's extended model. The basic model includes fixed effects for class and a spatially-unstructured random effect for county. The basic regression model is:

$$\text{logit}(p_{ij}) = a_j + \varepsilon_i$$

Where ε_i are modeled via a normal prior with mean zero.

3. Estimates of diabetes prevalence

The study's prevalence estimates of diagnosed diabetes in each county are the means of the posterior predictive distributions of the p_i's:

$$p_i = \frac{\sum_j p_{ij} N_{ij}}{\sum_j N_{ij}}$$

This weighting by population totals is called poststratification and simultaneously corrects for nonresponse. Age-adjusted prevalence for county i is given by:

$$\sum_k w_k \left[\frac{\sum_{j \in \delta_k} p_{ij} N_{ij}}{\sum_{j \in \delta_k} N_{ij}} \right]$$

where k indexes age group; w is a vector of standard population weights and δ_k contains the subset of classes belonging to age group k. The United States population in the year 2000 was used as the standard.

All posterior distributions were simulated in WinBUGS (17). The 2.5th and 97.5th percentiles of the posterior distributions of the p_i's provided the 95%CI for county prevalence of diagnosed diabetes. A burn-in of 5 000 was used, and a single chain for 20 000 iterations was then monitored.

4. Model comparison

Models were compared using the criterion developed by prior researchers (15). This criterion (D) is the sum of two parts, a goodness of fit measure (G), and the expected mean-square predictive error (P). Calculating D requires replicating the entire data set for each posterior draw of the model parameters. Using these replicates the posterior predictive mean and variance for each observation were computed. G is the sum over observations of the squared difference between the data and its posterior predictive mean; P is the sum over observations of the posterior predictive variances and D = G + P. Models with smaller values of D are preferred.

5. Model checking

Model checking includes answering the question, "Is the model consistent with the data?" Posterior predictive checks were implemented to examine the consistency of the model with the data (18). In posterior predictive checking, the entire data set is replicated for each posterior draw of the model parameters. A discrepancy or test measure that reflects relevant aspects of the model is calculated for each replicate. A Bayesian P-value associated with the test measure is calculated. A value of 0.1–0.9 indicates reasonable model fit. The Pearson χ^2 measure, which has been used for a long time, was used for model checking.

6. Prior Assumptions

The intercepts by class are assigned improper flat priors, $\alpha < 1$. The fixed effects, δ, of county level predictors are assigned diffuse normal priors with mean 0 and variance 1 000.

The spatially correlated effects by county and class, v, are assigned a multivariate normal (MVN) conditional autoregressive prior. Let $v_i = (v_{i1}, v_{i2}, ..., v_{i14})'$. Then:

$$\overline{v}_i \mid v_{(-i)} \sim \text{MVN}(\overline{v}_i, \Sigma_v)$$

where $v_{(-i)}$ equals the matrix v' with the ith column removed and $\overline{v}_i = \sum_{j \in \delta_i} v_j \mid n_i$. δ_i and n_i denote the set of labels of the neighbors of county i and the number of neighbors, respectively. The inverse of Σ_v is assigned a Wishart prior with scale matrix S_v and 14 degrees of freedom. The matrix S_v has ones along the main diagonal and 0.001 for all other elements (7).

The spatially unstructured effects by county and class, μ, are assigned a multivariate (of dimension 14) normal prior with mean zero and variance matrix Σ_μ. The inverse of Σ_μ is assigned a Wishart prior with scale matrix S_μ and 14 degrees of freedom. The matrix S_μ has ones along the main diagonal and 0.001 for all other elements (7).

The error terms, ε, in the basic model are assigned a proper half-Cauchy (16) prior distribution with median equal to one. This is a weakly informative prior distribution that, for this model, greatly speeds convergence.

RESUMEN

Variación en un área pequeña de la prevalencia de la diabetes en Puerto Rico

Objetivo. Calcular la prevalencia en el año 2009 de casos con diagnóstico de diabetes en Puerto Rico en adultos de 20 años de edad o mayores, para conocer mejor su distribución geográfica con objeto de que los responsables políticos puedan encauzar más eficientemente los programas de prevención y control.

Métodos. Se ajustó un modelo multinivel bayesiano a la combinación de datos del Sistema de Vigilancia de Factores de Riesgo del Comportamiento 2008–2010 y del Censo de los Estados Unidos del 2009 para calcular la prevalencia de la diabetes en cada uno de los 78 municipios de Puerto Rico.

Resultados. El cálculo del valor medio no ajustado para todos los municipios fue de 14,3% (intervalo por municipio de 9,9 a 18,0%). La amplitud promedio de los intervalos de confianza fue de 6,2%. Hubo poca diferencia entre los cálculos ajustados y los no ajustados.

Conclusiones. Los valores obtenidos mediante estos cálculos correspondientes a 78 municipios fueron por término medio más elevados y mostraron menor variabilidad (es decir, el intervalo era más pequeño) que los cálculos anteriormente publicados sobre la prevalencia de la diabetes en todos los municipios de los Estados Unidos en el 2008 (media, 9,9%; intervalo de 3,0 a 18,2%).

Palabras clave Diabetes mellitus; prevalence; política social; Puerto Rico.