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.

Manuscript received on 5 May 2012. Revised version accepted for publication on 17 December 2012.

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: ν_{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:

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

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 $v_i = (v_{i1}, v_{i2}, \ldots, v_{i14})'$. Then:

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

where $v_{(-i)}$ equals the matrix v' with the i^{th} 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.