Diabetes in Ontario

Determination of prevalence and incidence using a validated administrative data algorithm

JANET E. HUX, MD, FRCPC[1,2,3]
FRANK IVIS, BA[1]
VIRGINIA FLINTOFT, MSc[1]
ADINA BICA, MSc[1]

OBJECTIVE — Accurate information about the magnitude and distribution of diabetes can inform policy and support health care evaluation. We linked physician service claims (PSCs) and hospital discharge abstracts (HDAs) to determine diabetes prevalence and incidence.

RESEARCH DESIGN AND METHODS — A retrospective cohort was constructed using administrative data from the national HDA database, PSCs for Ontario (population 11 million), and registries carrying demographics and vital statistics. All HDAs and PSCs bearing a diagnosis of diabetes (ICD9-CM 250) were selected for 1991–1999. Two previously reported algorithms for identification of diabetes were applied as follows: "1-claim" (any HDA or PSC showing diabetes) and "2-claim" (one HDA or two PSCs within 2 years showing diabetes). Incident cases were defined as individuals who met the criteria for diabetes for the first time after at least 2 years of observation. For validation, diagnostic data abstracted from primary care charts (n = 3,317) of 57 randomly selected physicians were linked to the administrative data cohort, and sensitivity and specificity were calculated.

RESULTS — In 1998, 696,938 individuals met the 1-claim criteria and 528,280 met the 2-claim criteria. Sensitivity for diabetes was 90 and 86%; for the 1- and 2-claim algorithms, specificity was 92 and 97%, respectively, and positive predictive values were 61 and 80%, respectively. Using the 2-claim algorithm, the all-age prevalence increased from 3.2% in 1993 to 4.5% in 1998 (6.1% in adults). Incidence remained stable.

CONCLUSIONS — Administrative data can be used to establish population-based incidence and prevalence of diabetes. Diabetes prevalence is increasing in Ontario and is considerably higher than self-reported rates.

Diabetes is a common, chronic condition that imposes a heavy burden of morbidity and early mortality on affected patients (1–3). Diabetes and its complications drive a substantial portion of medical resource use for those who fund and deliver health care. At the same time, research findings now provide unprecedented levels of evidence regarding the prevention of diabetes complications (4–9). In this context, accurate, population-based assessments of the prevalence of diabetes become important for policy-makers and for those mounting and evaluating disease management strategies.

Previous work evaluating the prevalence of diabetes has been largely based on national population-based surveys (10–12), registries (13), and cohort studies in highly selected populations (14). National health interview programs may facilitate population-based estimates; however, there is evidence that they underestimate the prevalence of diagnosed diabetes, as revealed in medical record reviews (15–17). Surveys also provide insufficient data to define prevalence at the level of geographically small areas and are inefficient for ongoing surveillance.

Blanchard et al. (18) in Manitoba have used comprehensive databases of physician service claims (PSCs) and hospital discharge abstracts (HDAs) to identify individuals diagnosed with diabetes in the province and to estimate rates over time. Their validation against a diabetes education registry suggested that the sensitivity of the algorithm was in excess of 95%. However, such a validation strategy does not assess its sensitivity for identifying individuals who do not pursue diabetes education (who may use health services differently), nor does it assess the specificity of the algorithm.

The purpose of this study was to use linked PSCs and HDAs to determine the prevalence and incidence of diabetes in Canada's largest province. We sought to evaluate the sensitivity and specificity of the detection algorithm by comparison to an independent administrative database and by primary chart abstraction.

RESEARCH DESIGN AND METHODS

Data sources and definitions
Administrative data were used to assemble a cohort of individuals who had been diagnosed with diabetes in Ontario. Discharge abstracts prepared by the Canadian Institute for Health Information (CIHI) were used to identify patients admitted to a hospital with a diagnosis of

•••

diabetes (any of 16 diagnostic fields showing a diagnosis of diabetes: ICD9 250.x). Ontario Health Insurance Plan (OHIP) records were used to identify PSCs for which the diagnosis recorded was diabetes (ICD8 250.x). All relevant records from these two data sources from fiscal year 1991 (1 April 1991 to 31 March 1992) through fiscal year 1999 were extracted. All identified CIHI and OHIP records for individuals were linked through a reproducibly scrambled unique health care identifier.

Among the individuals identified with a diagnostic code for diabetes, those with diabetes were identified following the algorithm of Blanchard et al. (18), developed using administrative data in Manitoba. Their algorithm specified that any patient with two PSCs bearing a diagnosis of diabetes within a 2-year period or one hospitalization with a diagnostic code for diabetes would be identified as having diabetes. A similar algorithm requiring only a single PSC was also examined. Algorithms using a longer window to improve sensitivity were rejected because they were impractical for ongoing surveillance, and algorithms requiring more claims to improve specificity were rejected because of the unacceptable reduction in sensitivity. To exclude women from the diabetes database who had only gestational diabetes, any record bearing a diabetes diagnostic code followed by a PSC or hospital discharge record within 5 months, indicating an obstetrical event, were eliminated. The resultant administrative data cohort was titled the Ontario Diabetes Database (ODD).

Individuals identified as having diabetes were linked by a unique identifier to the Registered Persons Database (RPDB), the annual registry of all individuals eligible for coverage under the provincial health plan. The RPDB provided each patient's sex, year of birth, date of death (where applicable), and postal code. Individuals for whom no death record was identified remained in the diabetes database, regardless of whether they had claims with a diagnosis of diabetes in subsequent years.

The diabetes database was initially validated using two independently derived cohorts of individuals with diabetes identified from the Ontario Drug Benefit (ODB) Program database and the National Population Health Survey (NPHS). The ODB provides prescription drug benefit coverage for all individuals >65 years of age. Individuals with claims for either insulin or oral hypoglycemics were labeled as having diabetes. The NPHS is a self-report survey of a stratified, random sample of the population in which respondents were specifically asked whether they had diabetes that had been diagnosed by a physician. Both the NPHS and the ODB database carry each patient's unique numeric identifier and thus could be linked to the ODD in order to assess the algorithm's sensitivity to detect diagnosed diabetes.

Primary data collection
The ODD was further validated by primary data collection from physician office charts. To simplify data collection, the individuals selected for review were nested within the practices of randomly selected primary care physicians who practiced within 50 km of Toronto and who consented to participate. Eligible physician claim data were screened to ensure that they had seen at least 30 patients with diabetes in the past year. We did not require physicians whose practice demographics were representative of the province because the validation was based on adjudication of the assignment of an individual as having diabetes or not, rather than on the calculation of prevalence within the chart abstraction sample.

The sample size calculation was based on a need to detect high sensitivity in the algorithm and high positive predictive value (PPV). Assuming a prevalence of diabetes of 5%, a sample of 3,000 primary care patients would give a lower limit for a one-sided 95% binomial confidence interval of 98% for the sensitivity. A trained abstractor collected information regarding the diagnosis, duration, and type of diabetes. Diagnosis was based on a diagnosis of diabetes listed in clinic notes and/or consult letters and/or prescriptions for antidiabetic medications. In the absence of such evidence for disease, the patient was labeled as not having diabetes.

The study protocol was approved by the Research Ethics Board of the Sunnybrook and Women's College Health Sciences Center.

Analysis
The appropriate algorithm for identifying cases of diabetes from administrative data was determined by comparing the ODD cohorts to those derived from the ODB, NPHS, and primary chart review. Two algorithms were tested: one required only one PSC or hospitalization with a diagnosis of diabetes, and another previously reported algorithm required either two PSCs within a 2-year period or one hospitalization bearing a diagnosis of diabetes. We sought an algorithm that maximized sensitivity while providing at least 80% of PPV. For cases where there was disagreement between the administrative data algorithm and the reference population, descriptive statistics were prepared to elucidate the discrepancy.

Using the preferred algorithm, age- and sex-standardized annual prevalence and incidence rates were determined for 1994–1998. Beginning in 1994, any patient identified as having diabetes but not previously identified as such was labeled as an incident case. Note that patients classified as incident cases in 1994 have a maximum of 3 previous years of data free of diabetes codes, while in 1998 there is a 7-year period. Accordingly, incident rates in the early years may be inflated because of the inclusion of a small number of misclassified prevalent cases. Denominators for incidence and prevalence rates were drawn from Statistics Canada census data and intercensal estimates for the province of Ontario. Rates were compared over time and across counties ($n = 49$, median population 73,406). Standard small area variation statistics were prepared to test for variation between counties.

RESULTS

Validation of the administrative data algorithm
Application of the algorithm is illustrated in Fig. 1. Validation of both the 1- and 2-claim rules against a cohort of individuals receiving antidiabetic agents through the ODB demonstrated a high level of sensitivity (94 and 91%, respectively). Because individuals appearing in the ODB database must have obtained a prescription, this validation exercise does not measure sensitivity for individuals who use health services infrequently.

Comparison to self-reported diabetes status in the NPHS showed reasonable sensitivity (90 and 85% for 1- and 2-claim algorithms, respectively) but high levels of apparent false positives (PPV 44 and 64% for 1- and 2-claim algorithms, respectively). Descriptive analysis of the ap-
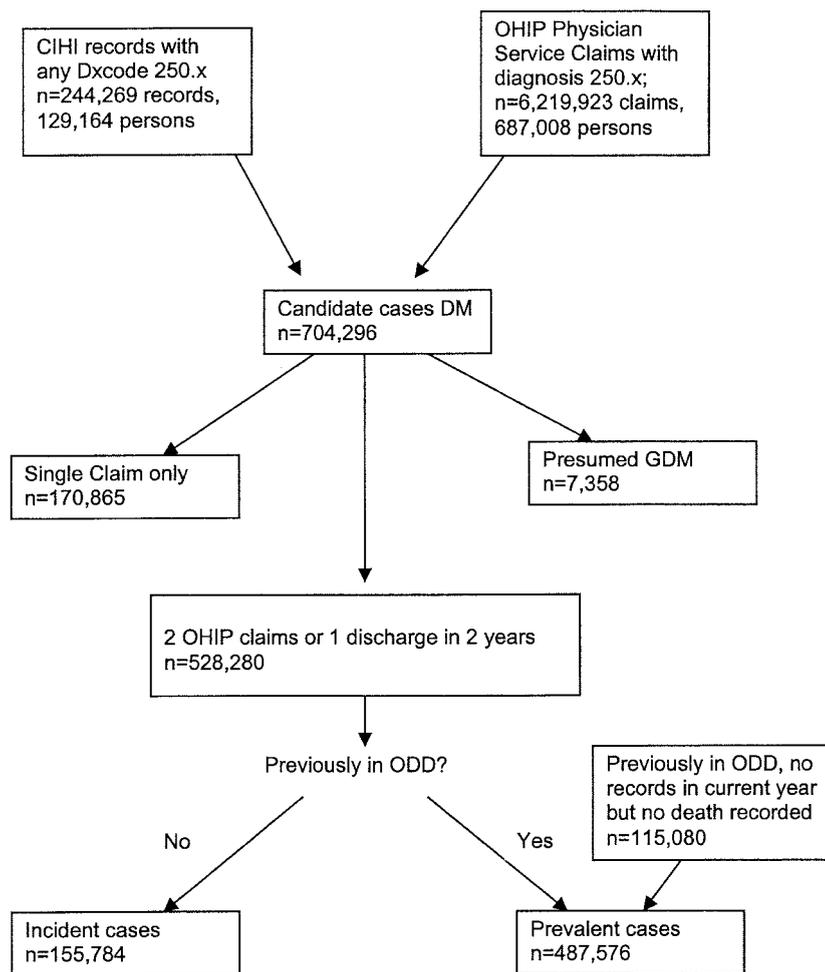
Figure 1—*Development of the Ontario Diabetes Database for 1998.*

confirmed in chart review (85 apparent false positives) are described in Table 2. Because there are no barriers to patients seeing multiple primary care providers, it is possible that some of these individuals had diabetes diagnosed by a nonstudy physician and accordingly may be true rather than false positives. Such a circumstance would be more likely if a patient saw multiple providers, as in our apparent false-positive cases who had seen a median of 5 (range 1–36) different physicians in the last 5 years and had a median of 5 claims (range 0–67) with a diagnosis of diabetes. We did not obtain consent from participating physicians to link provider data; therefore, we were not able to confirm whether the diabetes claims for these individuals were submitted by a study physician or by one or more other physicians. Individuals labeled as having diabetes by administrative data but not confirmed by chart review are also more likely to truly have diabetes if they are receiving antidiabetic drugs, if they have multiple office visits for diabetes, or if their diabetes was diagnosed in hospital, where accuracy of diagnostic information in administrative data are greater. If individuals meeting one or more of these criteria (Table 2) are considered true positives, the PPV of the 2-claim algorithm increases to 98%.

Determination of rates of diabetes
Rates of diabetes prevalence and incidence for the province determined using the 2-claim algorithm are shown in Fig. 2. There is a substantial and steady increase in the prevalence of diabetes over the observation period.

Prevalence rates by county ranged

parent false positives suggested that many of these were in fact true positives and that diabetes is under-reported in this survey.

Because the self-reported NPHS data appeared inadequate as a gold standard, primary care chart abstractions were undertaken. A total of 520 randomly selected physicians were invited to participate through an initial letter with follow-up to nonresponders. Chart abstraction was performed in the offices of 57 physicians (11%) who agreed to participate. Where provided, the most common reasons for declining participation were disruption of office routine and concerns about patient confidentiality. Then, 3,337 charts were abstracted using a standard data collection instrument, of which 3,317 could be linked to the diabetes databases defined from administrative data. The comparison of the two sources is shown in Tables 1 and 2.

Sources of discordance between administrative and chart data were explored in a descriptive analysis. Individuals who were labeled as having diabetes based on administrative data (2-claim rule) but not

Table 1—*Validation of administrative data algorithms against primary care chart data*

	Office charts		
Administrative data	Diabetes	No diabetes	*N*
One PSC or one hospitalization with diagnosis of diabetes*			
Diabetes	353	223	576
No diabetes	36	2,705	2,741
N	389	2,928	3,317
Two PSCs or one hospitalization with diagnosis of diabetes†			
Diabetes	335	85	420
No diabetes	54	2,843	2,897
N	389	2,928	3,317

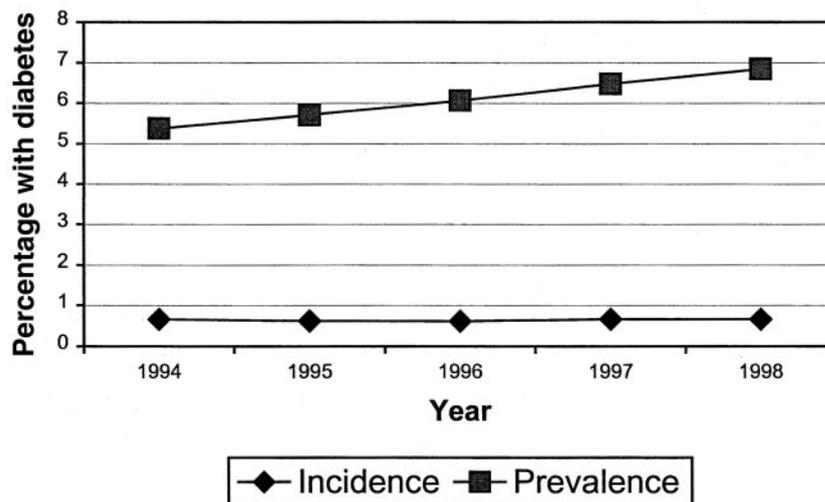Data are *n*. *Sensitivity 0.91, PPV 0.61; †sensitivity 0.86, PPV 0.80.

Table 2—*Description of apparent false positives using 2-claim rule: cases labeled as diabetes based on administrative data but not confirmed in chart reviews (n = 85)*

	n (%)
Individuals using hypoglycemic drugs (of those ≥65 years, n = 18)	3 (16.7)
Individuals with more than three office visits coded with diagnosis of diabetes	55 (64.7)
Individuals with at least one hospitalization with a diagnosis of diabetes	11 (12.9)
Individuals having seen more than three different physicians in last 5 years*	63 (74.1)
Individuals with one or more of the above four criteria	78 (91.8)

*Where a patient routinely sees multiple physicians, it is less likely that a given physician (i.e., the one whose charts were abstracted) would have the patient's full medical history including the diagnosis of diabetes.

Figure 2—*Prevalence and incidence of diabetes in Ontario, 1994–1998.*

from 4.58 to 9.58%. Rates and standard small area variation statistics are shown in Fig. 3. The highest rates were seen in Toronto, the largest and most ethnically diverse urban center in the province, and in counties with large Aboriginal populations. In the latter counties (including the four with the highest rates in Fig. 3), the prevalence was higher among women than men.

CONCLUSIONS — In a single-payer environment, administrative data provide a powerful resource for population-based evaluation of the burden of chronic diseases. The current study demonstrates that they can be used to measure the prevalence and incidence of disease and provide reasonable diagnostic agreement with data abstracted from primary care charts. Administrative data provide an efficient tool for ongoing disease surveillance. Moreover, in contrast to prevalence estimates based on stratified random samples, they define the full population of individuals diagnosed with diabetes, among which process and outcomes of diabetes care can be evaluated. Application of this tool in Ontario has demonstrated a marked increase in disease prevalence over the past 5 years. The trajectory of this growth, which appears unabated over the observation period, has important impli-

cations for those who plan and deliver health care in the province.

Detailed health surveys, such as Canada's National Population Health Survey (19), have been an important source of data for health planners and are attractive because they do not have the bias inherent in administrative data toward individuals who use health services. However, findings in the current study suggested considerable underreporting of diagnosed diabetes in response to that questionnaire.

The use of information abstracted from patients' primary care charts to validate the administrative data findings also proved problematic. Migration between providers and lack of efficient vertical integration of care may contribute to under detection if data are abstracted from the

office chart of a single practitioner because that physician may not represent the patient's regular source of care. The identification of individuals labeled as not having diabetes based on chart review, yet using insulin or oral hypoglycemics, provides strong evidence for the fallibility of such chart reviews. However, it does not precisely quantify the magnitude of the underdetection. We are accordingly left in the unsatisfactory position of having no definitive gold standard against which to validate the administrative data. Biological sampling would clearly be the preferred method for accurately determining the epidemiology of diabetes and would be free of many of the biases described above. Importantly, it would include those individuals who meet the clinical definition of diabetes but have not yet
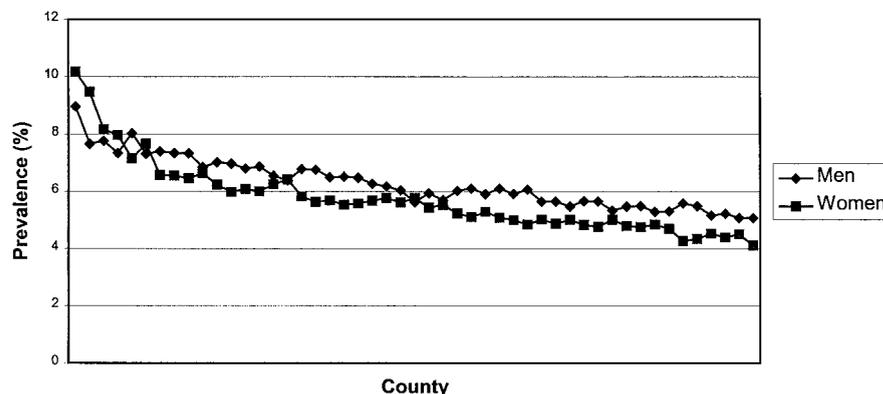
Figure 3—*Diabetes prevalence in Ontario counties. Small area variation statistics support a finding of moderate inter-county variation in prevalence of diabetes. Coefficient of variation 14.3%, extremal quotient 2.02, systematic component of variation 19.3, adjusted $\chi^2 = 1090$, df = 48, P < 0.0001.*

been diagnosed. Nonetheless, a cohort of individuals with diagnosed diabetes, such as the one described here, is of interest to those evaluating the delivery of care because services for the prevention of diabetes complications, for instance, are only appropriately assessed in those who have been diagnosed.

This work sought to determine an optimal administrative data algorithm for detecting diabetes, a task predictably hampered by the trade-off between sensitivity and specificity. Requiring only a single PSC significantly improves sensitivity, but at the cost of unacceptable false positives. These false positives may simply be coding errors or cases where diabetes was clinically suspected but subsequent laboratory tests did not confirm the diagnosis.

The prevalence of diabetes is rising in Ontario despite relatively stable incidence. A moderate degree of regional variation was observed. In particular, districts with a high proportion of Aboriginal residents were found to have both a high prevalence of diabetes and higher rates in women than men, consistent with previous epidemiological studies (20,21). The other area noted for high prevalence was metropolitan Toronto, a finding that may be artifactual, reflecting more effective screening in an urban academic center, but may also be a function of immigration from countries with high rates of diabetes. Further study is required to elucidate the reasons for regional and temporal variation in disease patterns.

This work supports the previously postulated (22,23) feasibility of using administrative data for chronic disease surveillance. One of the principal advantages of such method is that it not only quantifies the burden of disease, but it also defines a population in which process and outcome of disease management may be explored. A population-based cohort of individuals diagnosed with diabetes represents a valuable resource for those seeking to evaluate the delivery and outcomes of care for diabetes.

References

1. U.S. Renal Data System: Ch3. Incidence and causes of treated ESRD. In *US RDS 1995 Annual Data Report*. Bethesda, MD, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, p. 25–37, 1995
2. Kahn HA, Moorhead HB: *Statistics on Blindness in the Model Reporting Area, 1969–70: National Eye Institute*. Washington, DC, U.S. Govt. Printing Office, 1973 (NIH publ. no. 73-427)
3. Stamler J, Vaccaro O, Neaton JD, Wentworth D, the MRFIT Research Group: Diabetes, other risk factors, and 12-yr cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial. *Diabetes Care* 16:434–444, 1993
4. The Diabetes Control and Complications Trial Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 329:977–986, 1993
5. UK Prospective Diabetes Study Group: Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 352:837–853, 1998
6. UK Prospective Diabetes Study Group: Effect of intensive blood glucose control policy with metformin on complications in type 2 diabetes patients (UKPDS 34). *Lancet* 352:854–865, 1998
7. Early Treatment of Diabetic Retinopathy Study Research Group: Early photocoagulation for diabetic retinopathy, ETDRS report number 9. *Ophthalmology* 98:766–785, 1991
8. Ravid M, Lang R, Rachmani R, Lishner M: Long-term renoprotective effect of angiotensin-converting enzyme inhibition in non-insulin dependent diabetes mellitus: a 7 year follow-up study. *Arch Intern Med* 156:286–289, 1996
9. UK Prospective Diabetes Study Group: Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. *BMJ* 317:703–713, 1998
10. Tan M-H, MacLean DR: Epidemiology of diabetes mellitus in Canada. *Clin Invest Med* 18:240–246, 1995
11. Harris MI, Robbins DC: Prevalence of adult-onset IDDM in the U.S. population. *Diabetes Care* 17:1337–1340, 1994
12. National Center for Health Statistics: Current estimates from the National Health Interview Survey, 1992. In *Vital and Health Statistics*. Series 10, no. 190, 1994
13. LaPorte RE, Tajima N, Akerblom HK, Berlin N, Brosseau J, Christy M, Drash AL, Fishbein H, Green A, Hamman R: Geographic differences in the risk of insulin-dependent diabetes mellitus: the importance of registries. *Diabetes Care* 8 (Suppl. 1):101–107, 1985
14. Leibson CL, O'Brien PC, Atkinson E, Palumbo PJ, Melton LJ III: Relative contributions of incidence and survival to increasing prevalence of adult-onset diabetes mellitus: a population-based study. *Am J Epidemiol* 146:12–22, 1997
15. Mackenbach JP, Looman CW, van der Meer JB: Differences in the misreporting of chronic conditions, by level of education: the effect on inequalities in prevalence rates. *Am J Public Health* 86:706–711, 1996
16. Kriegsman DM, Penninx BW, van Eijk JT, Boeke AJ, Deeg DJ: Self-reports and general practitioner information on the presence of chronic diseases in community-dwelling elderly: a study on the accuracy of patients' self-reports and on determinants of inaccuracy. *J Clin Epidemiol* 49:1407–1417, 1996
17. Martin LM, Leff M, Calogne N, Garrett C, Nelson DE: Validation of self-reported chronic conditions and health services in a managed care population. *Am J Prev Med* 18:215–218, 2000
18. Blanchard JF, Ludwig S, Wajda A, Dean H, Anderson K, Kendall O, Depew N: Incidence and prevalence of diabetes in Manitoba, 1986–1991. *Diabetes Care* 19:807–811, 1996
19. James R, Young TK, Mustard CA, Blanchard J: The health of Canadians with diabetes. *Health Rep* 9:47–52, 1997
20. Young TK, Reading J, Elias B, O'Neil JD: Type 2 diabetes mellitus in Canada's first nations: status of an epidemic in progress. *CMAJ* 163:561–566, 2000
21. Burrows NR, Geiss LS, Engelgau MM, Acton KJ: Prevalence of diabetes among Native Americans and Alaska Natives, 1990–1997: an increasing burden. *Diabetes Care* 23:1786–1790, 2000
22. Hebert PL, Geiss LS, Tierney EF, Engelgau MM, Yawn BP, McBean AM: Identifying persons with diabetes using Medicare claims data. *Am J Med Qual* 14:270–277, 1999
23. Maskarinec G: Diabetes in Hawaii: estimating prevalence from insurance claims data. *Am J Public Health* 87:1717–1720, 1997