Who Has Diabetes? Best Estimates of Diabetes Prevalence in the Department of Veterans Affairs Based on Computerized Patient Data

DONALD R. MILLER, SCD[1,2]
MONIKA M. SAFFORD, MD[3,4]
LEONARD M. POGACH, MD, MBA[3,4]

OBJECTIVE — To optimize methods for identifying patients with diabetes based on computerized records and to obtain best estimates of diabetes prevalence in Department of Veterans Affairs (VA) patients.

RESEARCH DESIGN AND METHODS — The VA Diabetes Epidemiology Cohort (DEpiC) is a linked national database of all VA patients since 1998 with data from VA medical visits, Medicare claims, pharmacy and laboratory records, and patient surveys. Using DEpiC, we examined concordance of diabetes indicators, including ICD-9-CM codes (250.xx), prescription drug treatment, HbA_{1c} tests, and patient self-report. We determined the optimal criterion for identifying diabetes and used it in estimating diabetes prevalence in the VA.

RESULTS — The best criterion was a prescription for a diabetes medication in the current year and/or 2+ diabetes codes from inpatient and/or outpatient visits (VA and Medicare) over a 24-month period. This definition had high sensitivity (93%) and specificity (98%) against patient self-report, and reasonable rates of HbA_{1c} testing (75%). HbA_{1c} testing alone added few additional cases, and a single diagnostic code added many patients, but without confirmation (reduced specificity). However, including codes from Medicare was critical. Applying this definition for 1998–2000, we identified an average of 500,000 VA patients with diabetes per year. We also estimated high and increasing diabetes prevalence rates of 16.7% in FY1998, 18.6% in FY1999, and 19.6% in FY2000 and an incidence estimated to be ~2% per year.

CONCLUSIONS — Development and evaluation of methodology for analyzing computerized patient data can improve the identification of patients with diabetes. The increasing high prevalence of diabetes in VA patients will present challenges for clinicians and health system management.

Diabetes Care **27 (Suppl. 2):B10–B21, 2004**

Diabetes is a common, expensive, and burdensome chronic illness. It afflicts nearly 1 in 10 adults over the age of 45 years and is the seventh leading cause of death (1,2). The total annual costs attributable to diabetes are estimated to be nearly 100 billion dollars, and it accounts for 15% of all health care expenditures in the U.S. (3,4). Given the magnitude of this problem, health care systems need reliable and consistent systems-level data on diabetes in order to monitor its prevalence and incidence, estimate its burden on the system, and evaluate the impacts of care on its prevention and progression (5,6). While historical methods for disease rate estimation in health care systems have been expensive and unreliable, increasing availability of computerized administrative and medical records data represents a promising new resource for this purpose (6,7). Their use may be limited by a number of factors, however, including the dynamic nature of the populations, absence of complete enumeration lists, variable periodicity in patient visits, and uncertainty in the accuracy of diagnostic codes and other data. A few recent studies have contributed to our understanding on how these data may be used appropriately for disease rate estimation (8–16), but much remains to be done in developing and testing methods to analyze and make proper inferences from them.

One example of a health care system with particular need for good diabetes estimates is the Department of Veterans Affairs (VA). This is the largest integrated health care system in the U.S., providing comprehensive services to over 4.5 million veterans through 163 medical centers. VA patients tend to be relatively sick with multiple morbidities, and it is likely that diabetes has a high prevalence in this population (17–19). Fortunately, the VA information system is well developed and includes a rich assortment of patient-level data maintained in national databases (7,20). There have been prior attempts to use these data to estimate diabetes prevalence among VA patients (18,19), but these investigators used simple methods without testing for validity or consistency.

We have addressed these needs and report here our initial results from the national VA Diabetes Epidemiology Cohorts

From [1]Boston University, School of Public Health, Boston, Massachusetts; the [2]Center for Health Quality, Outcomes, and Economic Research, Bedford, Massachusetts; [3]VA New Jersey Health Care System, East Orange, New Jersey; and [4]University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, New Jersey.

Address correspondence and reprint requests to Donald R. Miller, ScD, Center for Health Quality, Outcomes & Economic Research (CHQOER), Edith Nourse Rogers Memorial Hospital (152), Bldg 70, 200 Springs Rd., Bedford, MA 01730. E-mail: drmiller@bu.edu.

Received for publication 1 July 2003 and accepted 25 July 2003.

Funding for this supplement was provided by The Seattle Epidemiologic Research and Information Center and the VA Cooperative Studies Program.

Abbreviations: DEpiC, Diabetes Epidemiology Cohort; VA, Department of Veterans Affairs.

A table elsewhere in this issue shows conventional and Système International (SI) units and conversion factors for many substances.

(DEpiC), which is a registry of virtually all VA patients with diabetes since 1998. It contains patient-level data on medical visits, with diagnoses and procedures for VA and non-VA care (from Medicare claims data), linked with VA pharmacy and laboratory data, and with patient reported data from large national surveys. Using these linked datasets, we have developed and tested methodology for diabetes rate estimation that may have application beyond the VA. Furthermore, we have applied these methods in estimating disease rates and trends in the VA and in forming cohorts of patients for further epidemiologic and health services research studies.

RESEARCH DESIGN AND METHODS

— This study was conducted using DEpiC databases for estimation of diabetes rates over 3 years, from October 1997 through September 2000 (fiscal years, from October through September 1998, 1999, and 2000). These databases consist of linked individual patient-level data obtained from a number of sources.

Data sources

National VA patient medical and administrative data were obtained from the Austin Automation Center in Texas. This center routinely obtains selected data abstracts from each VA facility, including information on patient demographics and eligibility for care, medical diagnoses and procedures, outpatient visits, hospital and long-term care stays, and vital status. Quality control is maintained centrally and is facilitated by system-wide standardization of data collection and processing. Medical diagnoses for conditions under treatment are recorded as ICD-9-CM codes (21) and medical procedures are coded as ICD-9-E or CPT4 codes (22). The accuracy and completeness of demographic data were maximized using custom algorithms that made use of data points across multiple years and from other sources. For this analysis, we used VA patient medical data from October 1996 through September 2002.

Many VA patients also obtain some of their care outside of the VA system with the majority being covered by Medicare (23). For this reason we obtained Medicare claims data from the Centers for Medicare and Medicaid Services for all VA patients with any indication of diabetes in VA data, supplemented by any additional VA patients with any codes for diabetes in the claims data alone. A verification process identified dependents and others receiving care and claimed under each beneficiary's number to assure that we applied claims correctly to the VA patients. We used Medicare files for institutional inpatient and outpatient care (Part A), physician care (Part B), skilled nursing facilities, home health care, hospice, and durable medical equipment. Data elements include vital status, demographics, encounter dates, and diagnostic and procedure codes (ICD-9-CM and CPT4). These files are known to be relatively complete for care covered by Medicare that is provided on a fee-for-service basis (24). For this analysis, we used Medicare data from October 1996 through December 2000.

VA pharmacy data were obtained from two sources. The VA Health Care Analysis Information Group used a software package in place at the majority of VA facilities to routinely collect data on diabetes medications prescribed through VA pharmacies. These data were available for prescriptions dispensed from October 1997 through September 2000, although, in the first year, data were missing for a number of sites representing 5.8% of the patient population. The second source of VA pharmacy data is the Pharmacy Benefits Management Strategic Health Group, which has maintained a national VA prescription database for all VA patients since October 1998. They extract prescription data from every VA facility on a monthly basis, verify its accuracy, and process it centrally into a national database. For this analysis, we used VA prescription data from October 1997 through September 2000, with the first year (FY1998) coming from the Health Care Analysis Information Group, and the last 2 years (FY1999 and FY2000) coming from the Pharmacy Benefits Management Group.

The VA Health Care Analysis Information Group also obtained and processed selected laboratory values, including HbA_{1c}, from a majority of VA facilities. For each of the 3 years of the study, these data were not available from facilities representing 26.5, 9.7, and 7.1% of the VA patient populations, although we found no apparent bias in rates obtained from other measures for this subset of the population. Given the variability in median values among facilities, we explored standardization of HbA_{1c} values to the normal range provided by each facility's laboratory, but—since this resulted in only small differences—we used unadjusted HbA_{1c} in this study. For this analysis, we used HbA_{1c} data from October 1997 through September 2000.

Vital status of all VA patients was determined from three sources. The Beneficiary Identification and Record Location files maintained at the Austin Automation Center is a national database of veterans who applied for death benefits. It has been reported that this file is 95% complete for the VA patient population (25,26), and recently the VA has conducted routine cross checks with Social Security Administration files to identify any missing deaths among veterans. Using an algorithm developed and tested for this purpose, we supplemented this database with indicators of mortality from VA inpatient records and Medicare files to improve completeness and accuracy of death records through September 2000.

The 1999 National Health Survey of VA Enrollees collected information directly from a large sample of VA patients, including self-report measures of diabetes and other diseases and use of VA and non-VA services (18). The diabetes question asked "Has a doctor ever told you that you had diabetes or high blood glucose?" For this survey, a stratified random sample of over 1.4 million enrollees (41% of all enrollees as of March 1999) were mailed surveys in July of 1999, and a total of 877,775 returned completed surveys (63.1% response rate). Medical and administrative records for all enrollees in the sampling frame were analyzed and used to derive weights to adjust for response and sampling bias. These data were linked with other DEpiC data for the evaluation.

Data analysis

Fixed-period prevalence of diabetes was estimated for each of the 3 fiscal years. All patients alive at the beginning of the fiscal year and receiving any VA health care during that year were included in the denominator. Nonveteran patients were further excluded since they were mostly relatives and VA employees who received sporadic, emergency care. Different definitions of diabetes using the various data elements were evaluated in those patients in the denominator to estimate 1-year period prevalence.

Definitions of diabetes under consid-

eration included combinations of ICD-9 codes for diabetes (250, 357.2, 362.0, 366.41), dispensing of medications for glycemic control (insulin, sulfonyureas, biguanides, thiazolidinediones, other hypoglycemic medications) or supplies for glucose monitoring, and orders for or high values from HbA_{1c} testing. Other factors considered in the definition included the use of various data sources for these indicators, the period of time used for surveillance of indicators of diabetes, and the source of the diagnostic codes. We recognized that prescriptions for diabetes medication would be a valid and specific indicator of diabetes, given that these medications are used very rarely for other indications. Other indicators must be used, however, since a large number of patients with diabetes are not treated with medications. Diagnostic codes require that the patient have a medical encounter involving evaluation or treatment of the condition and that it be coded accurately and entered into the information system. For these reasons, we assumed that a long period of surveillance would be needed to identify diabetes codes, beginning with 24 months, and that there may be errors in coding, with variation according to the source of the code. Codes from inpatient and outpatient visits were evaluated separately, and outpatient codes were restricted to those visits involving a face-to-face encounter with a clinician. Thus, we excluded diagnostic codes from laboratory, screening, and telephone encounters in VA outpatient files and in claims from the Medicare physician care files. Diabetes also may be identified through HbA_{1c} testing, but there are concerns over its validity as a sole indicator because a number of other conditions may raise HbA_{1c} levels, and such testing is used increasingly for screening purposes.

Various definitions of diabetes were evaluated in a number of ways. In the subset of 737,340 patients who completed a survey and were in our denominator for FY1999 from facilities with laboratory data, we used response to the diabetes question as an indicator of true diabetes status and calculated positive predictive value, specificity, and sensitivity for the various definitions as test criteria. Overall diabetes prevalence based on these definitions were calculated and compared to estimate the incremental effects of various factors. Cumulative prevalence was calculated on a monthly basis, and trends were examined as another means to evaluate effects of varying definitions on the estimates of prevalence. As a type of face validity, the proportions of patients receiving pharmacologic treatment and HbA_{1c} testing were examined and compared in those classified as having diabetes based on different definitions. Finally, VA patients in FY1998, the first year of the study, who were classified as having diabetes were followed forward in time to evaluate persistence of their assignment based on data from subsequent years, taking into account mortality and attrition from the health care system.

The working definition of diabetes that emerged from this evaluation process was used in estimating 1-year period prevalence of diabetes for each of the three years of the study. Rates are presented overall and in population subsets defined by sex, age, and race. We also present diabetes prevalence based on affirmative response to the diabetes question in the survey, with estimates adjusted for survey weights to account for sampling and response bias (18). All estimates are presented as crude rates and as rates directly standardized to the sex, age, and race distribution of the general U.S. adult population in 2000 (27).

In addition to prevalence, we estimated diabetes incidence over a 2-year period. VA patients in FY1998, the first year of the study, who were classified as not having diabetes were followed forward in time, and those who subsequently met the working definition of diabetes were considered as incident cases. We estimated incidence by both cumulative monthly enumeration of new cases and total number with the definition in subsequent years, taking into account mortality and attrition from the health care system.

RESULTS — There were a total of ~3 million veteran VA patients in each of the 3 years of the study (Table 1). Over this period, the population increased by 6–8% per year. Overall, most (95.3%) VA patients are men. The age distribution is irregular, reflecting the different theaters of war, with ~50% of patients aged 65 years or older. The racial and ethnic distribution of VA patients is similar to the overall adult U.S. population, except for underrepresentation of Hispanics, Native Americans, and Asian/Pacific Islanders. After excluding the patients with unknown racial/ethnic status, most VA patients are white (75.9%) or African American (17.9%).

Evaluation

Table 2 presents results from our evaluation of diabetes definitions based on diagnostic codes. In this evaluation, we used codes identified over a 24-month period from both VA and Medicare data with selected exclusions of outpatient codes. Varying combinations of numbers of codes and whether they come from inpatient or outpatient sources were considered and compared with the survey response to the diabetes question. The top part of the table presents the percent of those classified as having diabetes based on the various definitions that affirm diabetes in the survey. If the survey is considered as the standard, these are the predictive value positives for each group. Among patients with no diabetes diagnostic codes, 6.8% indicated diabetes in the survey. The percentage increases substantially with the presence of codes, to 37.4% for a single outpatient code, 50.8% for a single inpatient code, and >70% for all combinations of two or more codes (indicated in bold type). Based on these findings, we considered two or more diagnostic codes to be part of our working definition for diabetes.

The bottom of Table 2 shows a number of parameters for varying definitions based on codes, with our choice of two or more codes in bold. Prevalence varies predictably from a low of 6.9% for any inpatient codes to a high of 22% for any codes. As the number of codes required increases, the prevalence and sensitivity decrease. Specificity varies less across the definitions but is generally higher with more codes required. The other columns in this table show the percent of patients who were on diabetes medications or had HbA_{1c} testing with various test results over the prior 9-month period. The percentages tend to increase with increasing numbers of codes required.

Our choice of two or more codes represents a reasonable compromise. Including those with a single inpatient code, as has been recommended by Hebert et al. (15) based on analyses of Medicare data, would make little difference in prevalence and sensitivity but would include patients with a positive predictive value of only 50% and a medication prescription rate of only 38% (data not shown). Using our

Table 1—*VA patients 1998–2000: demographics and diabetes prevalence*

	FY1998—10/97–9/98*					FY1999—10/98–9/99					FY2000—10/99–9/00					1999 survey¶	
	All patients†		Diabetes‡			All patients		Diabetes			All patients		Diabetes				
				Prevalence (%)‖					Prevalence (%)					Prevalence (%)		Diabetes prevalence	
	No.	%§	No.	Crude	SR	No.	%	No.	Crude	SR	No.	%	No.	Crude	SR	Crude	SR
Overall	2,924,148		489,596	16.7	9.1	3,094,710		576,985	18.6	10.2	3,332,566		654,677	19.6	11.4	20.0	14.2
Sex																	
Male	2,786,407	95.3	479,874	17.2	10.3	2,948,302	95.3	565,435	19.2	11.7	3,177,093	95.3	641,598	20.2	13.1	20.7	15.9
Female	137,741	4.7	9,722	7.1	7.9	146,407	4.7	11,550	7.9	8.9	155,471	4.7	13,079	8.4	9.7	10.3	12.6
Age (years)																	
18–44	371,179	12.7	10,084	2.7	8.0	387,692	12.5	13,323	3.4	3.7	402,214	12.1	16,039	4.0	4.4	7.0	6.7
45–54	598,752	20.5	53,881	9.0	8.7	629,615	20.3	68,367	10.9	10.4	661,935	19.9	81,036	12.2	11.8	16.7	15.9
55–64	504,522	17.3	81,001	16.1	17.4	544,542	17.6	100,882	18.5	19.3	597,212	17.9	118,860	19.9	21.0	24.4	27.2
65–74	646,225	22.1	154,154	23.9	24.1	703,886	22.7	184,089	26.2	26.3	791,411	23.7	213,991	27.0	28.6	27.0	32.1
75–84	683,000	23.4	167,269	24.5	23.9	715,977	23.1	182,921	25.5	25.5	772,972	23.2	198,950	25.7	26.9	24.7	28.5
85–108	120,470	4.1	23,207	19.3	22.2	112,998	3.7	22,576	20.0	23.0	106,822	3.2	21,149	19.8	23.4	18.9	21.8
Race																	
White	1,780,199	60.9	354,970	19.9	8.4	1,885,611	60.9	418,339	22.2	9.4	2,000,738	60.0	476,426	23.8	10.5	19.0	12.8
African American	415,119	14.2	83,827	20.2	11.3	420,849	13.6	93,545	22.2	12.7	420,156	12.6	98,236	23.4	13.9	22.1	19.3
Hispanic	119,821	4.1	26,816	22.4	10.5	124,965	4.0	31,117	24.9	12.0	131,940	4.0	35,057	26.6	13.3	24.2	19.5
Native American	9,127	0.3	1,989	21.8	12.0	9,079	0.3	2,185	24.1	13.8	8,978	0.3	2,268	25.3	14.9	24.3	28.9
Asian/Pacific Islander	21,060	0.7	5,641	26.8	13.1	21,565	0.7	6,440	29.9	14.9	22,102	0.7	6,883	31.1	16.1	20.9	18.9
Unknown	578,822	19.8	16,353	—	—	632,641	20.4	25,359	—	—	748,652	22.5	35,807	—	—		—

*Fiscal years from October 1997 to September 2000; †veteran with any VA medical care use in year; ‡among all patients, those with a VA diabetes medication prescription in the year and/or two or more diabetes diagnostic codes in the prior 24 months in the combined health care data from the VA and Medicare; §column %—distribution of all patients by demographics; ‖row %—proportion of all patients in row with diabetes; ¶National Health Survey (n = 877,775) conducted July 1999; SR, standardized rate—directly standardized to the sex, age, and race distribution of the general U.S. adult population in 2000.

Table 2—*Agreement of diabetes diagnostic codes with patient survey, medication prescriptions, and HbA_{1c} testing among 737,340 VA patients who completed a 1999 survey*

		% indicating diabetes in survey by diabetes diagnostic codes in prior 24 months*						
			Number of codes in inpatient records					
			0	1	2	3	4+	
			19.1%	92.0%	95.0%	96.1%	96.3%	Total number of codes
	0	7.0%	6.8%	50.8%	70.2%	82.9%	71.6%	0 6.8%
# of codes in	1	40.0%	37.4%	74.9%	80.5%	76.5%	77.9%	1 38.6%
outpatient	2	72.5%	70.3%	85.5%	88.1%	82.5%	89.0%	2 70.8%
records	3	84.3%	82.8%	91.0%	91.8%	97.7%	86.9%	3 83.0%
	4+	96.2%	95.8%	97.0%	96.6%	97.1%	97.0%	4+ 96.0%

					Prior treatment and testing†			
						HbA_{1c} testing		
Diagnostic codes	Prevalence rate	Positive predictive value‡	Specificity§	Sensitivity‖	Diabetes medication	Any	value of 7+	value of 8+
Any	22.0%	85.3%	95.7%	78.3%	68.9%	67.1%	36.1%	23.5%
Any OP	21.7%	85.8%	95.9%	77.5%	70.2%	68.7%	37.4%	24.4%
Any IP	6.9%	93.7%	99.4%	26.9%	74.7%	67.5%	37.6%	25.0%
2+OP or 1+IP	18.9%	92.8%	98.1%	73.7%	77.6%	72.8%	41.0%	26.8%
2+ any	**18.7%**	**93.4%**	**98.3%**	**73.1%**	**80.9%**	**75.4%**	**42.1%**	**27.6%**
2+ OP	18.4%	93.7%	98.4%	72.2%	82.1%	76.4%	43.0%	28.2%
3+ any	17.3%	95.2%	98.9%	69.0%	84.4%	77.8%	44.5%	29.3%
3+ OP	17.0%	95.4%	98.9%	68.0%	85.3%	78.6%	45.2%	29.7%
4+ any	16.1%	96.0%	99.1%	65.0%	86.4%	79.1%	45.9%	30.3%
4+ OP	15.8%	96.2%	99.2%	63.8%	87.1%	79.8%	46.6%	30.7%

*Diagnostic codes from VA and Medicare records from 7/97 to 6/99; †diabetes medication prescriptions and HbA_{1c} testing from 10/98 to 6/99; ‡those indicating diabetes in survey as a percent of those with specified diabetes diagnostic codes in row; §those without the specified diabetes diagnostic codes in row as a percent of those not indicating diabetes in survey; ‖those with specified diabetes diagnostic codes in row as a percent of those indicating diabetes in survey. OP, outpatient; IP, inpatient.

definition, the diabetes prevalence is 18.7%; specificity and sensitivity are 98.3 and 73.1%, respectively; and the medication treatment and HbA_{1c} testing proportions seem reasonable, at 80.9 and 75.4%, respectively.

Figure 1*A* and *B* show results of cumulative prevalence calculations based on diagnostic codes for the 3-year period from October 1996 through September 1999. Figure 1*A* shows how rates vary over time with variation in the number of codes required. As the number of codes required in the definition increases, the length of surveillance required to obtain nearly stable rates increases. For the definition using two or more codes, 24 months are needed to attain the maximum, stable rate, although >90% of that rate is attained at 12 months. These rates are unadjusted since they accumulate both numerator (diabetes cases) and denominator (VA patients) incrementally among veteran patients without accounting for death or attrition or restricting to patients alive and using services in a given year. Figure 1*B* shows cumulative prevalence for the same period using two or more codes as the definition but with adjustment incrementally for death and attrition; attrition is defined as no use of services for 12 months following the last visit or stay.

Based on these analyses and on the knowledge that diabetes medication use is a specific indicator of the disease, our working definition for the identification of patients with diabetes is as follows: prescription for diabetes medication in the current year and/or two or more ICD-9-CM codes for diabetes in current and past years (24 months) from inpatient stays and/or outpatient visits on separate days, excluding codes from lab tests and other nonclinician visits.

We applied this definition in estimating diabetes prevalence for the 3 years of the study; the estimates, along with additional estimates of how these rates vary with modifications of the definition, are presented in Table 3. The prevalence estimates for the 3 years are 16.7, 18.6, and 19.6%. Diagnostic codes from outpatient visits are the most important source of indicators of the disease, since missing medication or inpatient data results in decrements in the prevalence estimates of no more than 0.8%. Missing Medicare data would be a more critical element since its inclusion accounts for ~2% in the estimates or nearly 60,000 cases in each year. Furthermore, given the larger decrement in rates observed for lack of medication data in the absence of Medicare data, Medicare data seem to identify many patients with VA prescriptions but minimal use of VA services. Attempts to improve the accuracy of outpatient diagnostic codes by either requiring 7 days between days with codes or restricting to codes from primary care visits made small differences in the rates, and our decision to exclude codes from laboratory tests and other nonclinician visits also had a very small effect on the rates. Classifying patients with a single inpatient code as diabetic (15) increases prevalence by no more than 0.3%, but considering those with a single outpatient code as diabetic increases rates substantially; if both are included in the definition, rates increase by ~3%. We have already shown in our evaluation how this does not seem to be justified. Classifying patients with a sole indicator of either multiple HbA_{1c} tests or high values from these tests adds few additional patients and does not seem justified.

Prevalence estimation

Prevalence of diabetes in the VA patient population is summarized in Table 1. We present both crude and standardized rates, overall and within subsets defined by sex, age, and race. Given that the VA patient population is older and predominantly male, it is not surprising that standardization resulted in lower rates. The crude prevalence estimates for the 3 years are 16.7, 18.6, and 19.6%, indicating rates that are both high and trending upward. Even after standardization, the rates are 9.1, 10.2, and 11.4%, indicating a 25% increase in prevalence over the 3-year period. Diabetes prevalence is higher in men and increases with age up to 65–74 years, with modest declines in rates at older ages. As expected, prevalence is higher in Hispanic and racial minority patients than in whites, with the highest standardized rates in patients who are Native American or Asian/Pacific islanders.

Diabetes prevalence estimates based on survey responses to the diabetes question are presented for comparison purposes. Since the survey was completed in the last quarter of FY1999, it should be compared with FY2000 estimates from the medical and administrative data. Survey estimates of prevalence are similar and somewhat higher, particularly after standardization. Standardization made less of a difference in the survey estimates because of the much lower proportion of unknown race in the survey sample (4.1%) as compared with the total population (19.8%), although unknown race from administrative data was <5% among patients with diabetes.

Mortality, persistence, and incidence estimation

VA patients from FY 1998, stratified according to diabetes status, were followed forward in time to estimate persistence of diabetes status in medical and administrative data, mortality among diabetic patients, and incidence of new diabetes among those without it at baseline. This is summarized in Table 4 and cumulative incidence is presented in Fig. 1*C* and *D*. Mortality among patients with diabetes averaged 5.0% per year over the 2-year period of follow-up as compared with 2.6% among patients without diabetes. A small percent (1.2%) of diabetic patients with no registration of death had no record of health care from VA or Medicare records. These patients may be receiving health care with other sources of coverage, and many of them are likely to be in long-term care, probably with Medicaid coverage. Another 4.7% of diabetic patients had no VA service use in that period but had Medicare claims only. Of the remaining 411,492 patients with diabetes in FY1998 and subsequent use of VA services, most (90.8%) met our definition for diabetes in the subsequent 2-year period, 3.1% had a single diabetes diagnostic code, and 5.9% had no indication of diabetes. Persistence was higher (96.4%) among surviving and VA service–using patients taking diabetes medication.

Diabetes incidence was estimated in the 1,946,413 VA patients who did not meet our definition for diabetes in FY1998 but survived and continued using VA services. Of these patients, 3.9% subsequently met the diabetes definition, providing an estimate of annual incidence of almost 2%. Another 2.7% of these patients subsequently had a single diabetes diagnostic code. Those patients who we considered to be free of diabetes in FY1998 included nearly 4% with a single diabetes diagnostic code in the baseline year. While a large percentage of these patients received the code in error or as

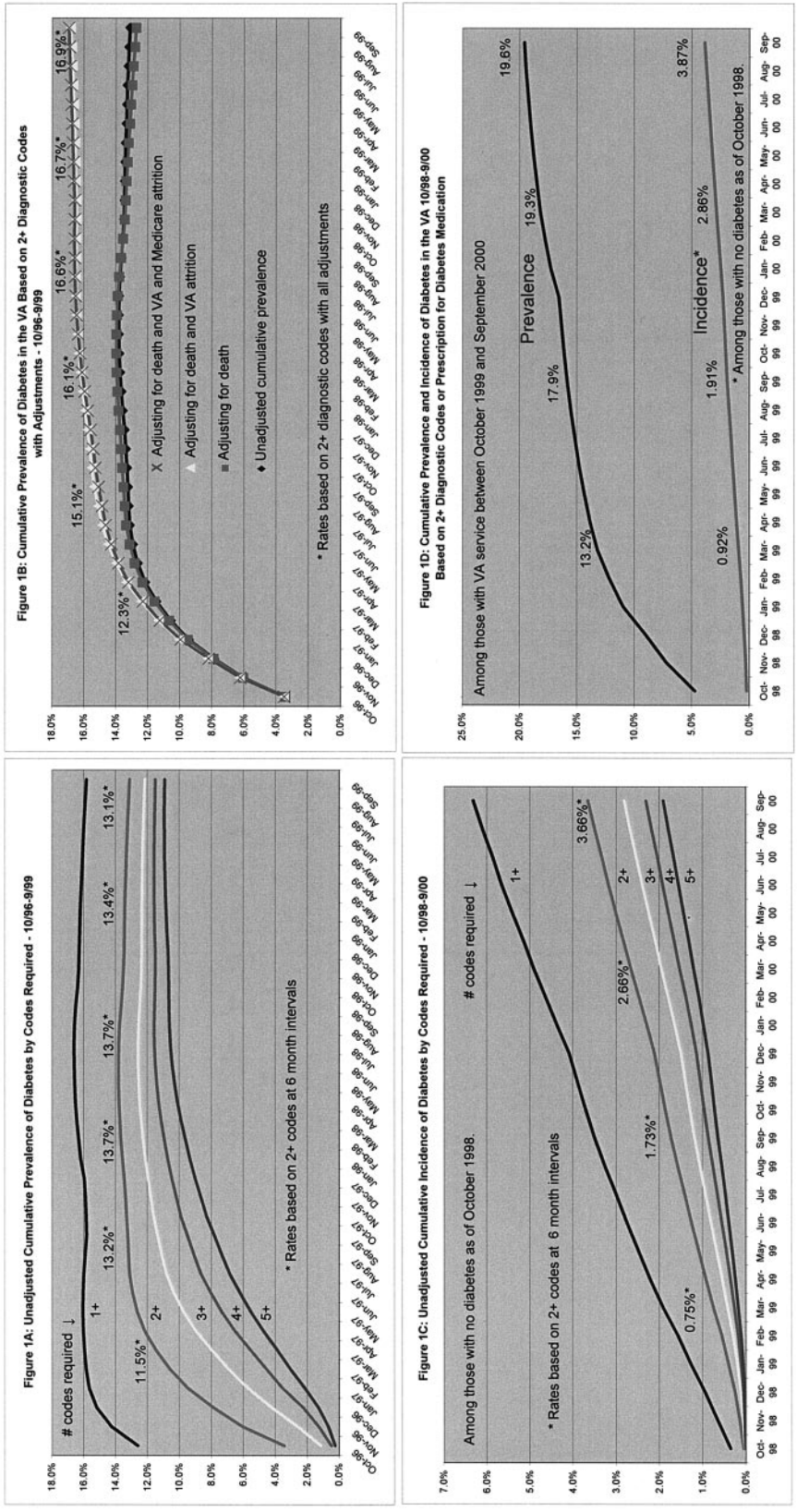

Figure 1—*Cumulative incidence and prevalence of diabetes.*

Table 3—*Diabetes prevalence in VA patients 1998–2000*—with variation in diabetes definition*

	1998 (2,924,148)		1999 (3,094,710)		2000 (3,332,566)	
	No.	Prevalence†	No.	Prevalence	No.	Prevalence
Working definition:						
VA and/or Medicare, diabetes medication in current year, and/or 2+ ICD-9-CM codes for diabetes over current and past year (24 months) from inpatient stays and/or outpatient visits on separate days excluding codes from lab tests and other nonclinician visits	489,596	16.7%	576,985	18.6%	654,677	19.6%
Same as working definition except missing data sources:						
No medication data, 2+ codes	485,551	16.6%	556,240	18.0%	637,417	19.1%
No inpatient data, medication and/or 2+ codes from outpatient visits only	474,886	16.2%	552,323	17.8%	643,442	19.3%
No outpatient visit data, medication and/or 2+ codes from inpatient stays only	373,247	12.8%	464,609	15.0%	476,084	14.3%
Same as working definition except no Medicare data:						
VA data only medication and/or 2+ codes	438,110	15.0%	510,926	16.5%	588,105	17.6%
VA data only, no medication data, 2+ codes	410,845	14.1%	474,546	15.3%	559,504	16.8%
VA data only, no diagnostic code data, diabetes medication	337,074	11.5%	439,283	14.2%	434,507	13.0%
Same as working definition, with all data, varying code definitions:						
Medication and/or 2+ codes from inpatient or outpatient visits requiring at least 7 days between days with codes	482,553	16.5%	561,044	18.1%	638,788	19.2%
Medication and/or 2+ codes from inpatient or primary care outpatient visits only	471,160	16.1%	549,807	17.8%	634,022	19.0%
Medication and/or 2+ codes from inpatient or outpatient visits including those from lab tests[2]	491,197	16.8%	578,464	18.7%	656,622	19.7%
Medication and/or 1 code from inpatient stays and/or 2+ codes from outpatient visits	494,181	16.9%	581,805	18.8%	663,181	19.9%
Medication and/or any (1+) codes from inpatient or outpatient visits	573,063	19.6%	676,854	21.9%	775,800	23.3%
Same as working definition, adding cases based on HbA_{1c} testing:						
Medication and/or 2+ codes and/or 2+ HbA_{1c} tests in current year	489,910	16.8%	577,922	18.7%	658,642	19.8%
Medication and/or 2+ codes and/or HbA_{1c} value of seven or higher in current year	490,235	16.8%	583,633	18.9%	662,130	19.9%

*Fiscal years from October 1997 to September 2000; †proportion of all patients in year with specified definition of diabetes; ‡and other nonclinician visits.

part of screening for diabetes, many of them were likely to have been diabetic or at least pre-diabetic. Therefore, it is not surprising that a substantial fraction of these patients met our definition for diabetes in the subsequent 2 years (26.7%) or had a single diabetes diagnostic code in that period (9%). It is evident from the cumulative incidence curves (Fig. 1*C*) that the number of patients enumerated with a single diabetes diagnostic code increases more steeply than the number with two or more or with use of diabetes medication (Fig. 1*D*). This suggests a combination of both increasing incidence and increasing screening for diabetes with assignment of initial diagnostic codes for diabetes.

CONCLUSIONS

Developing methodology for use of medical administrative data

This study has demonstrated the value of using computerized medical and administrative data for disease-rate estimation in a patient population, but also shows the importance of considering a number of potential limitations in using and interpreting these data. Our best definition of diabetes using diagnostic codes and prescription records showed good performance in a number of ways. Validity was evident from the evaluation using survey self-report of diabetes as the standard, in that our diabetes diagnostic code definition was found to be a good predictor (93%) of diabetes report in the survey, a highly specific (98%) indicator of diabetes with few false positives, and reasonably sensitive (73%) for identifying all patients with diabetes. Our best definition also had good face validity in that prevalence rates showed expected variation according to sex, age, and race, and patients identified as having diabetes had near-expected levels of medication treatment (81%), HbA_{1c} diagnostic testing (75%), and mortality (5% per year). The definition also appeared to have reasonably good reliability in that nearly all (91%) patients classified with diabetes who survived and continued using VA services had their status confirmed in subsequent years.

Table 4—*Survival, health care use, and diabetes persistence and incidence in follow-up of 1998 VA patients*

			Diabetes present in FY1998						Diabetes absent in FY1998					
	All patients		Either		Medication prescription		Two or more diagnoses		Single diagnosis, no medication		No diagnoses, no medication		Either	
1998 overall*	2,924,148		489,596	16.7%	337,074	11.5%	485,551	16.6%	83,467	2.9%	2,351,085	80.4%	2,434,552	83.3%
Among those with VA use 10/97–9/98:														
Subsequent vital status and service use														
Died 10/97–9/98	63,055	2.2%	21,226	4.3%	12,274	3.6%	21,269	4.4%	1,125	1.3%	40,704	1.7%	41,829	2.2%
Died 10/98–9/99	90,445	3.1%	27,822	5.7%	18,110	5.4%	27,875	5.7%	3,747	4.5%	58,876	2.5%	62,623	3.1%
Alive 9/99, no VA use or Medicare claims 10/99–9/00	340,386	11.6%	5,992	1.2%	2,483	0.7%	5,544	1.1%	188	0.2%	331,727	14.1%	331,915	13.6%
Alive 9/99, no VA use, Medicare claims only 10/99–9/00	72,357	2.5%	23,065	4.7%	6,278	1.9%	23,072	4.8%	2,345	2.8%	49,426	2.1%	51,771	2.1%
Alive 9/99, VA use 10/99–9/00	2,357,905	80.6%	411,492	84.0%	297,929	88.4%	407,791	84.0%	76,061	91.1%	1,870,352	79.6%	1,946,413	79.9%
Among those with VA use 10/97–9/98 and 10/99–9/00:														
Subsequent diabetes status														
2+ diabetes codes 10/98–9/00 or diabetes medication 10/99–9/00	476,696	20.2%	373,634	90.8%	287,308	96.4%	371,285	91.0%	20,289	26.7%	55,037	2.9%	75,326	3.9%
1 diabetes code only 10/98–9/00, no diabetes medication 10/99–9/00	73,361	3.1%	12,660	3.1%	5,238	1.8%	12,257	3.0%	6,861	9.0%	45,887	2.5%	52,748	2.7%
No diabetes codes 10/98–9/00, no diabetes medication 10/99–9/00	1,843,537	78.2%	25,197	6.1%	5,383	1.8%	24,249	5.9%	48,912	64.3%	1,769,427	94.6%	1,818,339	93.4%
1999 overall†	3,094,710		576,985	18.6%	439,283	14.2%	556,240	18.0%	99,869	3.2%	2,417,856	78.1%	2,517,725	81.4%
2000 overall‡	3,332,566		654,677	19.6%	434,507	13.0%	637,417	19.1%	121,123	3.6%	2,556,766	76.7%	2,677,889	80.4%

*Veteran patients with VA use 10/97–9/98 with diabetes defined as a diabetes medication prescription 10/97–9/98 and/or 2+ ICD-9 codes in VA or Medicare records for diabetes 10/96–9/98; †veteran patients with VA use 10/98–9/99 with diabetes defined as a diabetes medication prescription 10/98–9/99 and/or 2+ ICD-9 codes in VA or Medicare records for diabetes 10/97–9/99; ‡veteran patients with VA use 10/99–9/00 with diabetes defined as a diabetes medication prescription 10/99–9/00 and/or 2+ ICD-9 codes in VA or Medicare records for diabetes 10/98–9/00.

Our best definition was chosen only after careful consideration of the trade-offs of various modifications. For example, any modifications to our definition in terms of number of codes required or source of codes resulted in less favorable trade-offs in terms of sensitivity, specificity, and resulting prevalence. One variation that we did consider carefully was the inclusion of patients with a single inpatient code since this definition has been used by others based on an analysis of Medicare data (15). It is interesting that, in our evaluation, sensitivity (74%) and specificity (98%) for this definition were identical to those reported from the Medicare analysis, while positive predictive value was higher (93 vs. 84%), reflecting the higher prevalence of diabetes in the VA. Nevertheless, we judged this definition to be less optimal then ours since diabetes was confirmed in only 50% of the added patients and relatively few of them received diabetes treatment or HbA_{1c} testing. It is possible that the Hebert definition (15) performs better in patients covered by Medicare as opposed to the VA, but there is no evidence for this. While there may be some overcoding of diagnoses in VA records (12), coding of diagnoses in Medicare records may be susceptible to gaming or "diagnosis-related group or DRG creep" (10,28–30), resulting in more error and less reason to accept a single code in the definition.

Another modification that we considered that has been reported by others was the broader inclusion of all patients with any diagnostic code for diabetes (18,31). Clearly many of those patients with a single code in the 24-month period did have diabetes, as suggested by the high proportion (36%) of these patients with such codes and/or diabetes medication prescriptions in subsequent years. Not including them in the definition results in some underestimation of the true prevalence by perhaps 1%. Including them would introduce significant error, however, in falsely classifying patients as diabetic who may be pre-diabetic or not have the disease at all. In our evaluation, specificity only goes down to 95.7% for this definition, but it is likely that a proportion of patients affirming diabetes in the survey did not actually receive a diagnosis but were only screened for the disease or discussed it with their doctor.

A noteworthy finding from our analysis is the importance of including Medicare data in estimating disease rates in the VA. Most VA patients over the age of 65 years and a substantial proportion of those who are younger are eligible for Medicare benefits. It is known from both administrative records and patient reports that the majority of VA patients use non-VA services for part of their health care (23,32). In our analysis, without Medicare data, we would have missed identification of nearly 60,000 cases per year, and prevalence rates would be lower by ~10%. Without Medicare and VA medication data, the rates would be 15% lower. This finding highlights the importance of considering dual coverage in VA patients and the broader problem of complexity in health care coverage and provision for many in the U.S. in using medical and administrative records for disease estimation.

Other modifications to our definition made less difference. This includes the period of surveillance used to capture indicators of disease, various steps that may be taken in attempting to improve accuracy of diagnostic codes, and the use of laboratory data as a sole indicator of disease. Our analysis informed our decisions in how to best estimate diabetes prevalence in the VA, and the findings of others may vary depending on differences in the systems of care studied and the data used in the estimation process. Our analysis and findings should be instructive, however, to those who plan to estimate disease rates using medical and administrative records.

Diabetes in the VA

Our findings confirm the relatively high prevalence of diabetes in the VA patient population. Based on our findings, nearly one in five patients in the VA have diabetes, and nearly one in four patients who are a racial or ethnic minority have the disease. Furthermore, the prevalence appears to have increased substantially from 1998 to 2000.

Our estimate for diabetes prevalence in the VA differs from those of previous reports. Pogach et al. (19) reported that, in 1994, 11.8% of patients at 62 VA medical centers had diabetes based on prescriptions for diabetes medication or glucose monitoring strips. Kazis et al. (18) found that 16.9% of VA enrollees who were in the sampling frame for the 1999 survey had any diagnostic codes for diabetes in VA patient files over the past 2 years. These earlier and lower estimates are not in disagreement with our findings, but were calculated using incomplete data sources (e.g., no Medicare data) and relatively simple, untested methods. In this sense, the methodology developed as part of DEpiC and the prevalence estimates reported here represent a state-of-the-art approach to disease-rate estimation using computerized medical and administrative data.

The prevalence of diabetes among VA patients is substantially higher than that of the general population and most other patient populations. It has been reported that 7.2% of the adult population in the U.S. had diabetes in 2000 (33). This compares with our FY2000 standardized estimate of 11.4% based on medical and administrative data or, perhaps more appropriately, with our standardized estimate from the 1999 survey of 14.2%, since the general population estimate comes from a very similar survey question in the Behavioral Risk Factor Surveillance System. Recently published estimates of diabetes prevalence based on insurance claims or medical data tend to be lower, including estimates of 4.4% in Hawaii (34), 4.5% in Ontario (31), 6.7% in Manitoba (35), and 3.6% in the Northern California Kaiser Permanente Medical Care Program (36). Hebert et al. estimated a prevalence rate of 13.6% among Medicare beneficiaries >66 years of age in 1993 (15), as compared with our standardized estimate of 24.0% in VA patients in 1998.

The higher prevalence of diabetes in the VA may be due in part to more intense screening for the disease by VA clinicians. This may be particularly true for the more recent years following the change in diagnostic criteria for diabetes by the American Diabetes Association (37) and the subsequent increase in use of HbA_{1c} diagnostic testing. Based on analysis of data from the Third National Health and Nutrition Examination Survey, it has been reported that over a third of diabetes is undiagnosed and that there are a substantial number of additional people with impaired glucose tolerance (fasting plasma levels of between 110 and 126 mg/dl) (38). Intense screening may have resulted in higher proportions of patients with diabetes receiving a diagnosis in the VA, but, given that rates are nearly double those in other populations, it is still likely that the disease is more common in VA patients.

The heavy burden of diabetes among VA patients and the trend of increasing prevalence of this condition present challenges for VA clinicians and health system management in the provision of health care. Continued monitoring of prevalence, incidence, mortality, and measures of disease severity and progression, using the best available methodologies, will continue to be of great value in this process.

References

1. Harris MI, Flegal KM, Cowie CC, Eberhardt MS, Goldstein DE, Little RR, Weidmeyer HM, Byrd-Holt, DD: Prevalence of diabetes, impaired fasting glucose tolerance in U.S. adults: the Third National Health and Nutritional Examination Survey, 1988–1994. *Diabetes* 21:518–524, 1998
2. Stewart AL, Greenfield S, Hays RD, Wells K, Rogers WH, Berry SD, McGlynn EA, Ware JE Jr: Functional status and well-being of patients with chronic conditions: results from the Medical Outcomes Study. *JAMA* 262:907–913, 1989
3. Rubin RJ, Altman WM, Mendelson DN: Health care expenditures for people with diabetes mellitus. *J Endocrinol Metab* 78: 809A–809F, 1994
4. American Diabetes Association: Economic consequences of diabetes mellitus in the U.S. in 1997. *Diabetes Care* 21:296–309, 1998
5. Iezzoni LI: *Risk Adjustment for Measuring Health Care Outcomes*. Chicago, Health Administration Press, 1997
6. Ashton CM, Menke TJ, Deykin D, Camberg LC, Charns MP: A state-of-the-art conference on databases pertaining to veterans' health: a resource for research and decision making. *Medical Care* 34 (Suppl. 1):MS1–MS8, 1996
7. Boyko EJ, Koepsell TD, Gaziano JM, Horner RD, Feussner JR: US Department of Veterans Affairs medical care system as a resource to epidemiologists. *Am J Epidemiol* 151:307–314, 2000
8. Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP: Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. *N Engl J Med* 318:352–355, 1988
9. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA: The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health* 14:1016–1025, 1992
10. Jollis JG, Ancukiewicz M, DeLong E, Pryor D, Muhlbaier LH, Mark DB: Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. *Ann Intern Med* 119:844–80, 1993
11. West SL, Strom BL: Validity of pharmacoepidemiology drug and diagnosis data. In *Pharmacoepidemiology*. 2nd ed. Strom BL, Ed. Chichester, England, John Wiley, 1994, p. 549–580
12. Kashner TM: Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. *Medical Care* 36:1324–1336, 1998
13. Benesch C, Witter DM, Wilder AL, Duncan PW, Samsa GP, Matchar DB: Inaccuracy of the ICD-9-CM in identifying the diagnosis of ischemic cerebrovascular disease. *Neurology* 49:660–664, 1997
14. Glynn RJ, Monane M, Gurwitz JH, Choodnovskiy I, Avorn J: Agreement between drug treatment data and a discharge diagnosis of diabetes mellitus in the elderly. *Am J Epidemiol* 149:541–549, 1999
15. 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
16. Woods CR: Impact of different definitions on estimates of accuracy of the diagnosis data in a clinical database. *J Clin Epidemiol* 54:782–788, 2001
17. Kazis LE, Miller DR, Clark J, Skinner K, Lee A, Rogers W, Spiro A, Payne S, Fincke G, Selim A, Linzer M: Health-related quality of life in patients served by the Department of Veterans Affairs: results from the Veterans Health Study. *Arch Intern Med* 158:626–632, 1998
18. Kazis L, Perlin J, Skinner K, Ren XS, Lee A, Rogers W, Spiro A, Selim A, Miller DR: *1999 Large Health Survey of Veteran Enrollees: Health Status and Outcomes of Veterans. Veterans SF-36 Physical and Mental Component Summary Scores, Executive Report*. Washington, DC, Office of Quality and Performance, Veterans Health Administration, May 2000
19. Pogach LM, Hawley G, Weinstock R, Sawin C, Schiebe H, Cutler F, Zieve F, Bates M, Repke D: Diabetes prevalence and hospital and pharmacy utilization in the Veterans Health Administration (1994): use of an ambulatory care pharmacy-derived database. *Diabetes Care* 21: 368–373, 1998
20. Cowper DC, Hynes DM, Kubal JD, Murphy PA: Using administrative databases for outcomes research: select examples from VA Health Services Research and Development. *J Med Syst* 23:249–259, 1999
21. *International Classification of Diseases, 9th Revision, Clinical Modification*. 5th ed. Salt Lake City, UT, 1996
22. *Current Procedural Terminology*. 4th ed. Medicode, Salt Lake City, UT, CPT, 1999
23. Shen Y, Hendricks A, Zhang S, Kazis LE: VA enrollees' health care coverage and use of care. *Med Care Res Rev* 60:253–267, 2003
24. Agency for Health Care Policy and Research: Report to Congress: The feasibility of linking research-related data bases to Federal and non-Federal medical administrative databases. Rockville, MD, Agency for Health Care Policy and Research, 1991 (publ. no. 91-003)
25. Page WF, Braun MM, Caporaso NE: Ascertainment of mortality in the U.S. veteran population: World War II veteran twins. *Mil Med* 160:351–355, 1995
26. Fisher SG, Weber L, Goldberg J, Davis F: Mortality ascertainment in the veteran population: alternatives to the National Death Index. *Am J Epidemiol* 141:242–256, 1995
27. U. S. Census Bureau, Population Division, Population Distribution Branch: U. S. Census national population estimates by age group and sex: selected years from 1990 to 2000 [article online]. Available from www.census.gov. Accessed 1 March 2003
28. Assaf A, Lapane KL, McKenney JL, Carleton RA: Possible influence of the Prospective Payment System on the assignment of discharge diagnosis for coronary heart disease. *N Engl J Med* 329:931–935, 1993
29. Katz JN, Barrett J, Liang MH, Bacon AM, Kaplan H, Kieval RI, Lindsay SM, Roberts WN, Sheff DM, Spencer RT, Weaver AL, Baron JA: Sensitivity and positive predictive value of Medicare Part B physician claims for rheumatologic diagnoses and procedures. *Arthritis Rheum* 40:1594–1600, 1997
30. Taylor DH, Fillenbaum GG, Ezell ME: The accuracy of Medicare claims data in identifying Alzheimer's disease. *J Clin Epidemiol* 55:929–937, 2002
31. Hux JE, Ivis F, Flintoft V, Bica A: Diabetes in Ontario: Determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 25: 512–516, 2002
32. Wright SM, Daley J, Peterson ED, Thibault GE: Outcomes of acute myocardial infarction in the Department of Veterans Affairs: does regionalization of health care work? *Medical Care* 35:128–135, 1997
33. Mokdad AH, Bowman BA, Ford ES, Vinicor F, Marks JS, Kaplan JP: The continuing epidemics of obesity and diabetes in the United States. *JAMA* 286:1195–1200, 2001
34. Maskarinec G: Diabetes in Hawaii: estimating prevalence from insurance claims data. *Am J Public Health* 87:1717–1720, 1997
35. Blanchard JF, Ludwig S, Wajda A, Dean H, Anderson K, Kendall O, Depew N: In-

cidence and prevalence of diabetes in Manitoba. *Diabetes Care* 19:807–811, 1996
36. Selby J, Ray GT, Zhang D, Colby CJ: Excess costs of medical care for patients with diabetes in a managed care population. *Diabetes Care* 20:1396–1402, 1997
37. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 20:1183–1197, 1997
38. Harris MI, Flegal KM, Cowie CC, Eberhardt MS, Goldstein DE, Little RR, Weidmeyer HM, Byrd-Holt DD: Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults: the Third National Health and Nutritional Examination Survey, 1988-1994. *Diabetes Care* 21:518–524, 1998