

BMJ Open Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations

Bushra Khokhar,^{1,2} Nathalie Jette,^{1,2,3} Amy Metcalfe,^{2,4,5}
Ceara Tess Cunningham,¹ Hude Quan,^{1,2} Gilaad G Kaplan,^{1,2} Sonia Butalia,^{2,6}
Doreen Rabi^{1,2,6}

To cite: Khokhar B, Jette N, Metcalfe A, *et al.* Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open* 2016;**6**:e009952. doi:10.1136/bmjopen-2015-009952

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-009952>).

Received 17 September 2015
Revised 29 June 2016
Accepted 30 June 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to

Bushra Khokhar;
bushra.khokhar@ucalgary.ca

ABSTRACT

Objectives: With steady increases in 'big data' and data analytics over the past two decades, administrative health databases have become more accessible and are now used regularly for diabetes surveillance. The objective of this study is to systematically review validated International Classification of Diseases (ICD)-based case definitions for diabetes in the adult population.

Setting, participants and outcome measures:

Electronic databases, MEDLINE and Embase, were searched for validation studies where an administrative case definition (using ICD codes) for diabetes in adults was validated against a reference and statistical measures of the performance reported.

Results: The search yielded 2895 abstracts, and of the 193 potentially relevant studies, 16 met criteria. Diabetes definition for adults varied by data source, including physician claims (sensitivity ranged from 26.9% to 97%, specificity ranged from 94.3% to 99.4%, positive predictive value (PPV) ranged from 71.4% to 96.2%, negative predictive value (NPV) ranged from 95% to 99.6% and κ ranged from 0.8 to 0.9), hospital discharge data (sensitivity ranged from 59.1% to 92.6%, specificity ranged from 95.5% to 99%, PPV ranged from 62.5% to 96%, NPV ranged from 90.8% to 99% and κ ranged from 0.6 to 0.9) and a combination of both (sensitivity ranged from 57% to 95.6%, specificity ranged from 88% to 98.5%, PPV ranged from 54% to 80%, NPV ranged from 98% to 99.6% and κ ranged from 0.7 to 0.8).

Conclusions: Overall, administrative health databases are useful for undertaking diabetes surveillance, but an awareness of the variation in performance being affected by case definition is essential. The performance characteristics of these case definitions depend on the variations in the definition of primary diagnosis in ICD-coded discharge data and/or the methodology adopted by the healthcare facility to extract information from patient records.

BACKGROUND

Diabetes is a chronic disease that has increased substantially during the past

Strengths and limitations of this study

- Our systematic review was comprehensive as it had a broad search strategy that bore no language or time restriction.
- All included studies captured patient information at the population level with clear case definitions encompassing a broad spectrum of patients.
- There is the potential for a language bias as studies where full texts were not available in English were not considered.
- There are potential limitations for all reference standards used to validate administrative definitions for diabetes.

20 years.¹ At present, diabetes is the leading cause of blindness,² renal failure³ and non-traumatic lower limb amputations⁴ and is a major risk factor for cardiovascular disease.⁵ Owing to its chronic nature, the severity of its complications and the means required to control it, diabetes is a costly disease. The healthcare costs associated with this condition are substantial and can account for up to 15% of national healthcare budgets.⁶

Understanding the distribution of diabetes and its complications in a population is important to understand disease burden and to plan for effective disease management. Diabetes surveillance systems using administrative data can efficiently and readily analyse routinely collected health-related information from healthcare systems, provide reports on risk factors, care practices, morbidity and mortality and estimate incidence and prevalence at a population level.⁷ With steady increases in 'big data' and data analytics over the past two decades, administrative health databases have become more accessible to health services researchers and are now used regularly to study the processes and

outcomes of healthcare. However, administrative health data are not collected primarily for research or surveillance. There is a need for health administrative data users to examine the validity of case ascertainment in their data sources before use.⁸

By definition, surveillance depends on a valid case definition that is applied constantly over time. A case definition is set of uniform criteria used to define a disease for surveillance.⁹ However, a variety of diabetes case definitions exist, resulting in variation in reported diabetes prevalence estimates. A systematic review and meta-analysis of validation studies on diabetes case definitions from administrative records has been performed.¹⁰ This review aimed to determine the sensitivity and specificity of a commonly used diabetes case definition, “two physician claims or one hospital discharge abstract record within a two-year period” and its potential effect on diabetes prevalence estimation. Our study extends this body of work by systematically reviewing validated International Classification of Diseases (ICD), 9th edition (ICD-9)-based and ICD-10-based case definitions for diabetes and comparing the validity of different case definitions across studies and countries.

METHODS

Search strategy

This systematic review was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines¹¹ (see online supplementary appendix A). Two citation databases, MEDLINE and Embase, were searched using an OVID platform from 1980 until September 2015. The search strategy consisted of the following set of terms (see online supplementary appendix B): (1) (health services research or administrative data or hospital discharge data or ICD-9 or ICD-10 or medical record or health information or surveillance or physician claims or claims or hospital discharge or coding or codes) AND (2) (validity or validation or case definition or algorithm or agreement or accuracy or sensitivity or specificity or positive predictive value or negative predictive value) AND (3) medical subject heading terms for diabetes. Searches were limited to human studies published in English. The broad nature of the search strategy allowed for the detection of modifications of ICD codes, such as international clinical modification (eg, ICD-9-CM).

Study selection

Studies were evaluated in duplicate for eligibility in a two-stage procedure. In stage 1, all identified titles and abstracts were reviewed and in stage 2, a full text review was performed on all studies that met the predefined eligibility criteria. If either reviewer defined a study as eligible in stage 1, it was included in the full text review in stage 2. Disagreements were resolved by discussion or consultation with a third reviewer.

Inclusion/exclusion criteria

A study was included in the systematic review if it met the following criteria: (1) study population included those ≥ 18 years of age with type 1 diabetes mellitus or type 2 diabetes mellitus; (2) statistical estimates (sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) or κ) were reported or could be calculated; (3) an ICD-9 or ICD-10 case definition for diabetes was reported and validated; (4) a satisfactory reference standard (eg, self-report from population-based surveys or patient medical chart reviews) and (5) if it reported on original data. Studies validating diabetes in specialised populations (eg, cardiovascular disease) were excluded to ensure that the diabetes case definitions would be generalisable. Studies not employing a sole medical encounter data in their diabetes case definition (eg, inclusion of pharmacy or laboratory data) were also excluded, as the independent validity of such definitions could not be calculated. Bibliographies of included studies were manually searched for additional studies, which were then screened and reviewed using the same methods described above.

Data extraction and quality assessment

Primary outcomes were sensitivity, specificity, PPV, NPV and κ reported for each of the ICD-coded diabetes case definition. Other extracted data included sample size and ICD codes used. If statistical estimates were not reported in the original paper, estimates were calculated from data available.

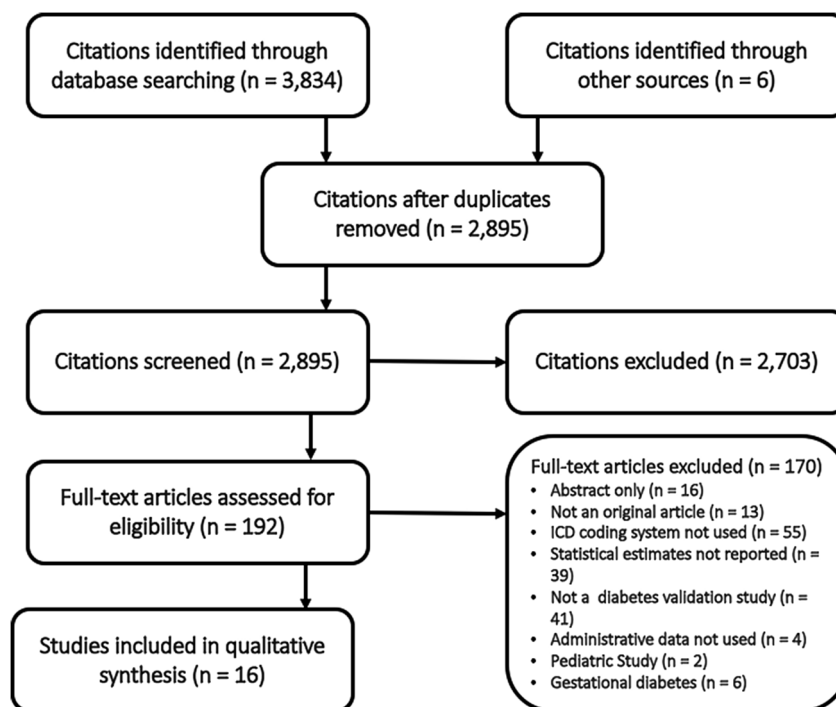
Calculating a pooled estimate of surveillance performance measures using meta-analytic techniques was deemed inappropriate given the heterogeneity of diabetes case definitions and reference standards used across studies. Data were tabulated by the type of administrative health data used. Study quality was evaluated using the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS) criteria.¹²

RESULTS

Identification and description of studies

A total of 2895 abstracts were identified with 193 studies reviewed in full text, of which 16 studies met all eligibility criteria (figure 1). Eight of these studies were conducted in the USA,^{13–20} seven in Canada^{21–27} and one in Australia.²⁸ Thirteen studies used ICD-9 codes,^{13–19 21–23 26–28} and the remaining three studies used ICD-9 and ICD-10 codes.^{23–25} None of the studies differentiated or commented as to whether a particular code of interest was in the primary or in one of the secondary diagnostic positions. Of the 16 studies reviewed, 8 used medical records,^{13 14 21 23–26 28} and 8 used either self-reported surveys or telephone surveys to validate the diabetes diagnosis.^{15–20 22–27} Eight studies used physician claims data,^{13–16 18–20 23} four studies used hospital discharge data,^{22 24 26 28} and four studies used a combination of both.^{17 21 25 27} Two studies used electronic

Figure 1 Study flow chart. ICD, International Classification of Diseases.



medical records (EMRs) as their health data source,^{29 30} but these were removed from the review since EMRs were not a part of our search strategy.

The QUADAS Scores (table 1) ranged from 9 to 13 of a maximum of 14. Five questions were selected from QUADAS to constitute the ‘bias assessment’. Regardless of quality assessment scores, all 16 studies are discussed in this systematic review.

The sample size varied from 93 to ~3 million people. Sensitivity and specificity values were available from all 18 studies, PPV in 16 studies, NPV in 12 studies and κ in 6 studies. All 16 studies were categorised by the type of administrative health data source being used.

Physician claims data

Table 2 lists the eight studies^{13–16 18–20 23} using physician claims data. In these studies, the sensitivity ranged from 26.9% to 97%, specificity ranged from 94.3% to 99.4%, PPV ranged from 71.4% to 96.2%, NPV ranged from 95% to 99.6% and κ ranged from 0.8 to 0.9. Four of the eight studies using physician claims data had a least one diabetes case definition where sensitivity and specificity exceed 80%.

Studies comparing physician claims-based case definitions over multiple years^{13 15 16} consistently show increases in sensitivity values and a slight decrease in specificity and PPV overtime. This relationship is consistent with the study¹⁸ looking at changes in the statistical estimates with increasing the number of appearance of diagnostic codes in the case definition—the sensitivity was the highest when any diagnostic code (inpatient or outpatient) was used, whereas the specificity and PPV

were the highest when most number of outpatient diagnostic codes were used.

Hospital discharge data

Table 3 lists the four studies^{22 24 26 28} using only hospital discharge data. In these studies, the sensitivity ranged from 59.1% to 92.6%, specificity ranged from 95.5% to 99%, PPV ranged from 62.5% to 96%, NPV ranged from 90.8% to 99% and κ ranged from 0.6 to 0.9. Two of the four studies using hospital discharge data had a least one diabetes case definition where sensitivity and specificity exceed 80%. In contrast to the physician claims-based case definitions, the sensitivity seemed to improve when a longer duration was used in the case definition, however the specificity and the PPV behaved inversely.

Combination of physician claims and hospital discharge data

Table 4 lists out the four studies^{17 21 25 27} using a combination of physician claims and hospital discharge data. In these studies, the sensitivity ranged from 57% to 95.6%, specificity ranged from 88% to 98.5%, PPV ranged from 54% to 80%, NPV ranged from 98% to 99.6% and κ ranged from 0.7 to 0.8. Using a combination of two or more data sources increases the minimum value of the range for sensitivity compared to using either physician claims or hospital discharge data-based definitions individually. All four of the studies using a combination of physician claims and hospital discharge data had a least one case definition where sensitivity and specificity exceed 80%.

Another factor affecting the statistical estimates is the number of claims being used in the definition. Rector

Table 1 Study quality characteristics using QUADAS tool

QUADAS tool item	Hux <i>et al</i> ²¹	Robinson <i>et al</i> ²²	Borzecki <i>et al</i> ¹³	Wilchesky <i>et al</i> ²³	Crane <i>et al</i> ¹⁴	So <i>et al</i> ²⁴	Chen <i>et al</i> ²⁵	Nedkoff <i>et al</i> ²⁸	Quan <i>et al</i> ²⁶	Young <i>et al</i> ²⁷	Hebert <i>et al</i> ¹⁵	Ngo <i>et al</i> ¹⁶	Rector <i>et al</i> ¹⁷	Miller ¹⁸	Singh ¹⁹	O'Connor <i>et al</i> ²⁰
Was the spectrum of patients representative of the patients who will receive the test in practice?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were selection criteria clearly described?	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Is the reference standard likely to correctly classify the target condition?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Did patients receive the same reference standard regardless of the index test result?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was the reference standard independent of the index test (ie, the index test did not form part of the reference standard)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was the execution of the index test described in sufficient detail to permit replication of the test?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was the execution of the reference standard described in sufficient detail to permit its replication?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Continued

Table 1 Continued

QUADAS tool item	Hux <i>et al</i> ²¹	Robinson <i>et al</i> ²²	Borzecki <i>et al</i> ¹³	Wilchesky <i>et al</i> ²³	Crane <i>et al</i> ¹⁴	So <i>et al</i> ²⁴	Chen <i>et al</i> ²⁵	Nedkoff <i>et al</i> ²⁸	Quan <i>et al</i> ²⁶	Young <i>et al</i> ²⁷	Hebert <i>et al</i> ¹⁵	Ngo <i>et al</i> ¹⁶	Rector <i>et al</i> ¹⁷	Miller ¹⁸	Singh ¹⁹	O'Connor <i>et al</i> ²⁰
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	Unclear	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Unclear
Were uninterpretable/intermediate test results reported?	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	Yes
Were withdrawals from the study explained?	Unclear	Unclear	No	No	No	No	Unclear	Yes	Unclear	Unclear	Unclear	No	No	No	Unclear	Unclear
Score (maximum 14)	11	11	10	12	10	9	11	13	12	12	11	11	12	12	12	12
Bias assessment (maximum 5)	5	5	5	5	4	4	5	5	5	5	5	5	5	5	5	5

QUADAS tool is extracted from table 2 of Whiting *et al*.¹²

QUADAS, Quality Assessment Tool for Diagnostic Accuracy Studies.

Table 2 Study characteristics and test measures of studies for physician claims data

Country	Study years	Author ^(reference)	Reference	Type of administrative data	Diabetes case definition	ICD codes used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	κ
Canada	1995–1996	Wilchesky <i>et al</i> ²³	Medical chart	Physician claims	Using only diagnoses recorded in the claims of study physicians	ICD-9 250.0-.9	2752	51.78 (49.9 to 53.6)	98.41 (98.2 to 98.6)			
					Using diagnostic codes recorded on claims made by all physicians who provided medical services to patients in the year prior to the start of the study	ICD-9 250.0-.9		64.43 (62.6 to 66.2)	96.82 (96.5 to 97.1)			
USA	1997–2001	Crane <i>et al</i> ¹⁴	Clinician documentation in EMR progress notes	Physician claims	At least one clinician-coded diagnoses	ICD 9 250.0, .1, .2, .3	1441	93 (86 to 100)	99 (99 to 100)	91 (83 to 99)		
USA	1998–1999	Borzecki <i>et al</i> ¹³	Medical charts	Physician claims	At least one diagnosis in National Department of Veterans Affairs (VA) database, Outpatient Clinic file over 1 year	ICD 9 250.x	1176	97	96			0.92
					At least two diagnoses in National Department of Veterans Affairs (VA) database, Outpatient Clinic file over 1 year	ICD 9 250.x						0.91
					At least one diagnosis in National Department of Veterans Affairs (VA) database, Outpatient Clinic file over 2 years	ICD 9 250.x						0.89
					At least two diagnoses in National Department of Veterans Affairs (VA) database, Outpatient Clinic file over 2 years	ICD 9 250.x						0.93
USA	1992–1995	Hebert <i>et al</i> ¹⁵	Self-reported survey	Physician claims	One or more diagnoses of diabetes in any claim file over 1-year period	ICD 9-CM 250.00-.93, 357.2, 362.0-362.02, 366.41		71.6	96.6	79		
					One or more diagnoses of diabetes in any claim file over 2-year period	ICD 9-CM 250.00-.93, 357.2, 362.0–0.02, 366.41		79.1	94.3	71.4		
USA	1993–1994	O'Connor <i>et al</i> ²⁰	Telephone survey	Physician claims	Two or more ICD-9 diagnostic codes	ICD 9 250.x	1976	92.22*	98.62*	76.15*	99.63*	

Continued

Table 2 Continued

Country	Study years	Author ^(reference)	Reference	Type of administrative data	Diabetes case definition	ICD codes used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	κ
USA	1996–1998	Singh ¹⁹	Self-reported survey	Physician claims	Veterans Affairs databases	ICD 9 250		76 (75 to 76)	98 (98 to 98)	91 (91 to 91)	95 (94 to 95)	0.79 (0.79 to 0.80)
USA	1997	Ngo <i>et al</i> ¹⁶	Self-reported survey	Physician claims	Oregon Medicaid Claims Data, any claim ≤24 months before interview with a diabetes diagnosis code	ICD 9 250, 357.2, 362, 366.41	21 564	83.9	97.9	81.9	98.2	0.81 (0.77 to 0.85)
					Oregon Medicaid Claims Data, any claim ≤12 months before interview with a diabetes diagnosis code	ICD 9 250, 357.2, 362, 366.41		88.7	97.4	76.4	98.9	0.8 (0.76 to 0.85)
USA	1997–2000	Miller <i>et al</i> ¹⁸	Self-reported survey	Physician claims (Medicare)	Any diagnostic code	ICD 9 250, 357.2, 362.0, 366.41	2 924 148	78.3	95.7	85.3		
					Any outpatient diagnostic code	ICD 9 250, 357.2, 362.0, 366.41		77.5	95.9	85.8		
					≥2 any diagnostic code	ICD 9 250, 357.2, 362.0, 366.41		73.1	98.3	93.4		
					≥2 outpatient codes	ICD 9 250, 357.2, 362.0, 366.41		72.2	98.4	93.7		
					≥3 any diagnostic code	ICD 9 250, 357.2, 362.0, 366.41		69	98.4	95.2		
					≥3 outpatient codes	ICD 9 250, 357.2, 362.0, 366.41		68	98.9	95.4		
					≥4 any diagnostic code	ICD 9 250, 357.2, 362.0, 366.41		65	99.1	96		
					≥4 outpatient codes	ICD 9 250, 357.2, 362.0, 366.41		63.8	99.2	96.2		

Superior performance characteristics within studies have been highlighted in bold.

*Sensitivity, specificity, PPV and NPV are all hand-calculated:

sensitivity identifies the proportion of patients who truly do have the disease/condition;

specificity identifies the proportion of patients who truly do not have the disease/condition;

PPV is the probability that participants with a positive screening test truly have the disease/condition;

NPV is the probability that participants with a negative screening test truly do not have the disease/condition;

κ is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

EMR, electronic medical record; ICD 10-AM, International Classification of Diseases, Tenth Revision, Australian Modification; ICD, International Classification of Diseases; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; NPV, negative predictive value; PPV, positive predictive value.



Table 3 Study characteristics and test measures of studies for hospital discharge data

Country	Study Years	Author ^(Reference)	Reference	Type of administrative data	Diabetes case definition	ICD codes used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	κ	
Canada	1995–2000	So <i>et al</i> ^{P4}	Medical chart	Hospital discharge data	Diabetes with complications	ICD-9 250.1–.9	93	80 (51.91 to 95.67)	98.3 95.15 to 99.65)	80 (51.91 to 95.67)	98.3 (95.15 to 99.65)		
	Diabetes with complications				ICD-10 E10.0–.8, E11.0–.8, E12.0–.8, E13.0–.8, E14.0–.8		66.7 (38.38 to 88.18)	98.9 (96.00 to 99.86)	83.3 (51.59 to 97.91)	97.2 (93.67 to 99.10)			
Canada	2003	Quan <i>et al</i> ^{P6}	Medical chart	Hospital discharge data	Diabetes with chronic complications	ICD 9 250.4–.7	4008	63.6	98.9	62.5	99	0.62	
					Diabetes with chronic complications	ICD 10 E10.2–.5, E10.7, E11.2–.5, E11.7, E12.2–.5, E12.7, E13.2–.5, E13.7, E14.2–.5, E14.7		59.1	99	63.1	98.9	0.6	
					Diabetes without chronic complications	ICD 9 250.0–.3, 250.8, .9		77.7	98.4	86.5	97	0.8	
					Diabetes without chronic complications	E10.0, .1, .6, .8, .9, E11.0, .1, .6, E11.8, .9, E12.0, .1, .6, .8, .9, E13.0, .1, .6, .8, .9, E14.0, .1, .6, .8, .9		75.8	98.7	88.5	96.8	0.79	
Western Australia	1998	Nedkoff <i>et al</i> ^{P8}	Medical chart	Hospital discharge data	Look back period: Index admission	ICD 9/ICD-9 CM 250	1685	91.1	98.7	93.3	97.4	0.912	
					1 year			91.6	98.1	92.8	97.6	0.902	
					2 years			92.1	97.9	92.1	97.8	0.903	
					5 years			92.4	97.7	91.9	97.8	0.9	
					10 years			92.6	97.6	91.4	97.8	0.9	
					15 years			92.6	97.5		97.8	0.897	
	2002–2004				Look back period: Index admission	ICD 10-AM E10-E14	2258	81.5	98.2	96	90.8	0.825	
					1 year			86.3	97.3	94.4	93	0.853	
					2 years			87.3	96.7	93.5	93.4	0.854	
					5 years			89.3	95.9	92.2	94.4	0.859	
Canada	1989–1990	Robinson <i>et al</i> ^{P2}	Self-reported survey	Hospital discharge data and physician claims	1, 2 or 3 physician claim or 1 hospitalisation over 3 years	ICD 9 CM	2651	72	98	76	98	0.72 (0.67–0.77)	
								89.6	95.5	91.5	94.5	0.855	

Superior performance characteristics within studies have been highlighted in bold.

Sensitivity identifies the proportion of patients who truly do have the disease/condition;

specificity identifies the proportion of patients who truly do not have the disease/condition;

PPV is the probability that participants with a positive screening test truly have the disease/condition;

NPV is the probability that participants with a negative screening test truly do not have the disease/condition;

κ is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

ICD 10-AM, International Classification of Diseases, Tenth Revision, Australian Modification; ICD, International Classification of Diseases; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; NPV, negative predictive value; PPV, positive predictive value.

Table 4 Study characteristics and test measures of studies for physician claims data and hospital discharge data

Country	Study years	Author ^(reference)	Reference	Type of administrative data	Diabetes case definition	ICD codes used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	κ
Canada	1992–1999	Hux <i>et al</i> ²¹	Medical chart	Physician claims and hospital discharge data	One physician service claims or one hospitalisation with diagnosis of diabetes	ICD-9 250.x	3317	91	92*	61	99*	
					Two physician service claims or one hospitalisation with diagnosis of diabetes	ICD-9 250.x		86	97*	80	98*	
Canada	2000–2002	Chen <i>et al</i> ²⁵	Medical chart	Physician claims and hospital discharge data	3 years observation period data	ICD 9 250.xx, ICD 10 E10.x-14.x	3362	95.6 (92.5 to 97.7)	92.8 (91.9 to 93.7)	54 (49.6 to 58.5)	99.6 (99.4 to 99.8)	0.65 (0.61 to 0.69)
					2 years observation period data	ICD 9 250.xx, ICD 10 E10.x-14.x		86.4 (82.4 to 90.5)	97.1 (96.5 to 97.7)	72.4 (67.5 to 77.3)	98.8 (98.4 to 99.2)	0.77 (0.73 to 0.81)
				Physician claims	3 years observation period data	ICD 9 250.xx, ICD 10 E10.x-14.x		91.2 (87.9 to 94.6)	97.6 (97.1 to 98.1)	72.1 (67.5 to 76.9)	99.2 (98.9 to 99.5)	0.82 (0.78 to 0.85)
					2 years observation period data	ICD 9 250.xx, ICD 10 E10.x-14.x		76.6 (71.5 to 81.6)	99.3 (99.0 to 99.6)	90.9 (87.2 to 94.6)	98 (97.5 to 98.4)	0.82 (0.78 to 0.85)
USA	1999	Rector <i>et al</i> ¹⁷	Telephone surveys	Hospital discharge data and physician claims	One 1999 claim with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	3633	90	93			
					One 1999 face-to-face encounter claim with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41		82	96			
					One 1999 face-to-face encounter claim with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41		72	98			
					Two 1999 claims with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41		85	96			
					Two 1999 face-to-face encounter claims with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41		70	98			

Continued

Table 4 Continued

Country	Study years	Author ^(reference)	Reference	Type of administrative data	Diabetes case definition	ICD codes used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	κ
	1999–2000				Two 1999 face-to-face encounter claims with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	57		99			
					One 1999 or 2000 claim with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41)	95		88			
					One 1999 or 2000 face-to-face encounter claim with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	94		92			
					One 1999 or 2000 face-to-face encounter claim with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	87		96			
					Two 1999 or 2000 claims with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	93		93			
					Two 1999 or 2000 face-to-face encounter claims with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	91		95			
					Two 1999 or 2000 face-to-face encounter claims with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	77		98			

Continued

Table 4 Continued

Country	Study years	Author ^(reference)	Reference	Type of administrative data	Diabetes case definition	ICD codes used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	κ
Canada	1980–1984	Young <i>et al</i> ²⁷	Self-reported survey	Hospital admission and physician claims	(Hospital admissions of provincial residents claims for which are submitted to the MHSC) AND (Hospital admissions of provincial residents claims for which are submitted to the MHSC AND Claims by the physician to the MHSC or payment)	ICD 9-CM	1000	82.7	96.3			
					(Hospital admissions of provincial residents claims for which are submitted to the MHSC AND Claims by the physician to the MHSC or payment) AND (Claims by the physician to the MHSC or payment)	ICD 9-CM		82.1	98.5			
					(Hospital admissions of provincial residents claims for which are submitted to the MHSC) AND (Hospital admissions of provincial residents claims for which are submitted to the MHSC AND Claims by the physician to the MHSC or payment) AND (Claims by the physician to the MHSC or payment)	ICD 9-CM		83.9	95.8			

Superior performance characteristics within studies have been highlighted in bold.

*Sensitivity, specificity, PPV and NPV are all hand-calculated:

sensitivity identifies the proportion of patients who truly do have the disease/condition;

specificity identifies the proportion of patients who truly do not have the disease/condition;

PPV is the probability that participants with a positive screening test truly have the disease/condition;

NPV is the probability that participants with a negative screening test truly do not have the disease/condition;

κ is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

ICD 10-AM, International Classification of Diseases, Tenth Revision, Australian Modification; ICD, International Classification of Diseases; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; MHSC, Manitoba Health Services Commission; NPV, negative predictive value; PPV, positive predictive value.

et al's study¹⁷ shows consistent results where the sensitivity is higher when at least one claims data are used in the definition, but the specificity is higher when at least two are used. Finally, Young *et al's* study²⁷ demonstrates the highest sensitivity when two physician claims and two hospital discharge data are used in the definition and the highest specificity when one physician claim and two hospital claims are used in the definition.

A secondary tabulation of data was performed by the type of ICD coding system used. Eight studies using ICD-9 coding systems are from the USA and four studies from Canada. Four studies use ICD-9 and ICD-10 coding systems—three of these are from Canada and one from Western Australia. In studies using ICD-9 codes, sensitivity ranged from 26.9% to 100%, specificity ranged from 88% to 100%, PPV ranged from 21% to 100%, NPV ranged from 74% to 99.6% and κ ranged from 0.6 to 0.9; whereas, in the studies using ICD-10 codes, the ranges for sensitivity (59.1% to 89.6%) and specificity (95.5% to 99%) narrowed significantly, and PPV ranged from 63.1% to 96%, NPV ranged 90.8% to 98.9% and κ ranged from 0.6 to 0.9.

DISCUSSION

In this systematic review, case definitions appear to perform better when more data sources are used over a longer observation period. The outcomes with respect to sensitivity, specificity and PPV for each of these studies seem to differ due to variations in the definition of primary diagnosis in ICD-coded health data, the use of hospital discharge versus physician billing claims and by the geographical location.

The validity of diabetes case definitions varies significantly across studies, but we identified definition features that were associated with better performance. The combinations of more than one data source, physician claim and/or hospital discharge encounter along with an observation period of more than 1 year consistently demonstrated higher sensitivity with only a modest decline in specificity. These definition characteristics are present in the definition used by the National Diabetes Surveillance System to identify Canadians with diabetes mellitus.³¹ The performance of this particular definition has been widely studied, and a meta-analysis pooling the results of these studies demonstrates a pooled sensitivity of 82.3% (95% CI 75.8% to 87.4%) and a specificity of 97.9% (95% CI 96.5% to 98.8%).¹⁰

This systematic review provides new knowledge on factors that are associated with enhanced definition performance and outlines the trade-offs one encounters with respect to sensitivity and specificity (and secondarily PPV and NPV) related to data source and years of follow-up. The development of an administrative case definition of diabetes is often related to pragmatic considerations (type of data on hand); however, this systematic review provides health services researchers with important information on how case definitions may perform given definition characteristics.

There was considerable 'within-data definition' variation in measures of validity. This variation likely reflects that neither physician claims nor hospital discharge data are primarily collected for surveillance; hence, the accuracy of diagnoses coded in these data sources remains suspect. Physician claims, while potentially rich in clinical information, are not recorded in a standardised manner. Billing practices do vary by practitioner, which may in turn be influenced by the nature of physician reimbursement (salary vs fee for service).^{23 32 33} Furthermore, patients with diabetes commonly carry multiple comorbidities, so while patients may have diabetes and be seen by a physician, providers will file billing claims for conditions other than diabetes.^{34 35} In contrast, hospital discharge data are limited to clinical information that is relevant to an individual hospitalisation, capturing diagnostic and treatment information usually for a brief window of time. The advantage of hospital discharge data for surveillance is that discharge diagnostic and medical procedure information are recorded by medical coders with standardised training with a detailed review of medical charts. However, the standard method of discharge coding does vary regionally, and thus variation around validity estimates based on these differences in coding practices will be observed.

Ideal performance parameters will vary based on the clinical condition of interest, the nature of surveillance and the type of data being used for surveillance. When studying diabetes trends and incidence rate, a case definition that has high but balanced measures of sensitivity and PPV is preferred. This will ensure maximal capture of potential patients and that patients captured likely have diabetes. This systematic review suggests that the commonly used two physician outpatient billings and/or one hospitalisation within a certain period of time is appropriate. It is also important to recognise that the data source used may also affect the type of patient identified with administrative data definitions. Hospital discharge data (when used in isolation) will potentially identify patients with more advanced disease or more complications and therefore may not be fully representative of the entire diabetes population. Similarly, physician claims data may identify a comparatively well, ambulatory population that has access to physician care in the community.

The greatest strength of this systematic review is its inclusiveness—the search strategy was not restricted by region, time or any particular case definition of diabetes. However, most of the studies, 15 of the 16, included in the qualitative analysis were conducted in North America with high sensitivity and specificity estimates between the cases identified through the administrative data versus medical records and the administrative data versus population-based surveys across studies, suggesting that public administrative data are a viable substitute for diabetes surveillance. Finally, the study quality across all studies included was generally high as measured by the QUADAS Scale.

There is the potential for a language bias as studies whose full texts were not available in English were not considered. There are potential limitations for all reference standards used to validate administrative case definitions for diabetes. The accuracy of chart reviews depends principally on physician documentation, availability of records and the accuracy of coding.³⁶ Self-reported surveys and telephone surveys are prone to recall bias, social desirability bias, poor understanding of survey questions or incomplete knowledge of their diagnosis. Self-reported surveys can also suffer from participation biases as patients with low diabetes risk may be less willing to participate whereas certain patients with advance diabetes may be too unwell to participate. Age, sex and a patient's level of education can have an effect on the reporting of diabetes.^{37–39} Those with poorly controlled diabetes have been found to underreport their disease status.⁴⁰ The ideal reference standard would be a clinical measure (such as glucose or HbA1c); however, the use of a clinical reference standard is not often performed.

In addition to the limitations of the reference standards used for validation, it should also be noted that even clinical measures as a references standard are imperfect and glucose and HbA1C are surrogates of the underlying disease process. It should also be noted that glucose and HbA1C thresholds for diagnosis have changed (albeit modestly) over the past 20 years. Changes in the clinical definition overtime have significant implications to diabetes surveillance. Understanding changing diagnostic thresholds is critical to interpreting surveillance data. However, the validity of an administrative data case definition is conceptually related but somewhat separate from the clinical definition. If we are to understand the clinical definition as a biological or physiologic definition that denotes the presence or absence of disease, the administrative data definitions are a surrogate of disease and denote the presence or the absence of disease based on care for the disease. The administrative definitions identify patients with a diagnosis of diabetes based on an interaction with the healthcare system in which they received care for diabetes. Therefore, the application of this definition follows the application of the clinical definition. There is a presumption that the clinical definition, whatever it may be at the time of the application, was valid.

Finally, difference between type 1 diabetes mellitus and type 2 diabetes mellitus is not clear in studies using administrative databases. In this systematic review, we included only adult population (≥ 18 years of age), which is primarily the type 2 diabetes population.

Generalisability

Fifteen of the 16 included studies were conducted in North America, and therefore it is not surprising that the validation studies report comparable results. However, even though these studies are nested in the general population, the selected diabetes cohorts used in the validation

studies may not always be truly representative of the general population.

CONCLUSIONS

Most studies included in this review use similar case definitions that require one or more diagnoses of diabetes. The performance characteristics of these case definitions depends on the variations in the definition of primary diagnosis in ICD-coded discharge data and/or the methodology adopted by the healthcare facility to extract information from patient records. Purpose of surveillance and the type of data being used should command the performance parameters of an administrative case definition. Approaches used in developing case definitions for diabetes can be simple and practical and result in high sensitivity, specificity and PPV. Overall, administrative health databases are useful for undertaking diabetes surveillance,^{21 25} but an awareness of the variation in performance being affected by case definition is essential.

Author affiliations

¹Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

²O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, Canada

³Department of Clinical Neurosciences, Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada

⁴Department of Obstetrics and Gynecology, University of Calgary, Calgary, Alberta, Canada

⁵Alberta Children's Hospital Research Institute, Calgary, Alberta, Canada

⁶Division of Endocrinology, Department of Medicine, University of Calgary, Calgary, Alberta, Canada

Contributors NJ wrote the protocol. BK, AM and CTC carried out the systematic review. BK wrote the manuscript. NJ, HQ, GKG, SB and DR provided final approval of the version to be published. All authors read and approved the final manuscript.

Funding BK was supported by the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD) and The Western Regional Training Centre for Health Services Research (WRTC). NJ holds a Canada Research Chair in Neurological Health Services Research and an Alberta Innovates Health Solutions (AI-HS) Population Health Investigator Award and operating funds (not related to this work) from the Canadian Institutes of Health Research, AI-HS, the University of Calgary and the Hotchkiss Brain Institute and Cumming School of Medicine. CTC is funded by a Canadian Institute of Health Research doctoral research scholarship. GKG is a Population Health Investigator supported by Alberta Innovates—Health Solutions. DR is a Population Health Investigator supported by Alberta Innovates—Health Solutions.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Any additional data such as study protocol and data extraction forms are available by emailing the first author at bushra.khokhar@ucalgary.ca

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Danaei G, Finucane MM, Lu Y, *et al.* National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet* 2011;378:31–40.
2. Karumanchi DK, Gaillard ER, Dillon J. Early diagnosis of diabetes through the eye. *Photochem Photobiol* 2015;91:1497–504.
3. Kiefer MM, Ryan MJ. Primary care of the patient with chronic kidney disease. *Med Clin North Am Online* 2015;99:935–52.
4. Leone S, Pascale R, Vitale M, *et al.* Epidemiology of diabetic foot. *Infez Med* 2012;20(Suppl 1):8–13.
5. Grundy SM, Benjamin IJ, Burke GL, *et al.* Diabetes and cardiovascular disease: a statement for healthcare professionals from the American Heart Association. *Circulation* 1999;100:1134–46.
6. World Health Organization. Diabetes: the cost of diabetes. <http://www.who.int/mediacentre/factsheets/fs236/en/> (accessed 27 Aug 2014).
7. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health* 2011;32:91–108.
8. Molodecky NA, Panaccione R, Ghosh S, *et al.* Challenges associated with identifying the environmental determinants of the inflammatory bowel diseases. *Inflamm Bowel Dis* 2011;17:1792–9.
9. Centers for Disease Control and Prevention: National Notifiable Diseases Surveillance System (NNSS). <http://www.cdc.gov/nndss/case-definitions.html>. (accessed 1 Sep 2015).
10. Leong A, Dasgupta K, Bernatsky S, *et al.* Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. *PLoS One* 2013;8:e75256.
11. Liberati A, Altman DG, Tetzlaff J. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009;151:W65–94.
12. Whiting P, Rutjes AW, Reitsma JB, *et al.* The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
13. Borzecki AM, Wong AT, Hickey EC, *et al.* Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? *Am J Med Qual* 2004;19:201–6.
14. Crane HM, Kadane JB, Crane PK, *et al.* Diabetes case identification methods applied to electronic medical record systems: their use in HIV-infected patients. *Curr HIV Res* 2006;4:97–106.
15. Hebert PL, Geiss LS, Tierney EF, *et al.* Identifying persons with diabetes using Medicare claims data. *Am J Med Qual* 1999;14:270–7.
16. Ngo DL, Marshall LM, Howard RN, *et al.* Agreement between self-reported information and medical claims data on diagnosed diabetes in Oregon's Medicaid population. *J Public Health Manag Pract* 2003;9:542–4.
17. Rector TS, Wickstrom SL, Shah M, *et al.* Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+choice health plans that have chronic medical conditions. *Health Serv Res* 2004;39:1839–57.
18. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the department of veterans affairs based on computerized patient data. *Diabetes Care* 2004;27(Suppl 2):B10–21.
19. Singh JA. Accuracy of veterans affairs databases for diagnoses of chronic diseases. *Prev Chronic Dis* 2009;6:A126.
20. O'Connor PJ, Rush WA, Pronk NP, *et al.* Identifying diabetes mellitus or heart disease among health maintenance organization members: sensitivity, specificity, predictive value, and cost of survey and database methods. *Am J Manag Care* 1998;4:335–42.
21. Hux JE, Ivis F, Flintoft V, *et al.* Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002;25:512–16.
22. Robinson JR, Young TK, Roos LL, *et al.* Estimating the burden of disease. Comparing administrative data and self-reports. *Med Care* 1997;35:932–47.
23. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol* 2004;57:131–41.
24. So L, Evans D, Quan H. ICD-10 coding algorithms for defining comorbidities of acute myocardial infarction. *BMC Health Serv Res* 2006;6:161.
25. Chen G, Khan N, Walker R, *et al.* Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Res Clin Pract* 2010;89:189–95.
26. Quan H, Li B, Saunders LD, *et al.* Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res* 2008;43:1424–41.
27. Young TK, Roos NP, Hammerstrand KM. Estimated burden of diabetes mellitus in Manitoba according to health insurance claims: a pilot study. *CMAJ* 1991;144:318–24.
28. Nedkoff L, Knuiman M, Hung J, *et al.* Concordance between administrative health data and medical records for diabetes status in coronary heart disease patients: a retrospective linked data study. *BMC Med Res Methodol* 2013;13:121.
29. Zgibor JC, Orchard TJ, Saul M, *et al.* Developing and validating a diabetes database in a large health system. *Diabetes Res Clin Pract* 2007;75:313–19.
30. Kandula S, Zeng-Treitler Q, Chen L, *et al.* A Bootstrapping algorithm to improve cohort identification using structured data. *J Biomed Inform* 2011;44(Suppl 1):S63–68.
31. Public Health Agency of Canada. National Diabetes Surveillance System, Public Health Agency of Canada. <http://www.phac-aspc.gc.ca/ccdpc-cpcmc/ndss-snsd/english/index-eng.php> (accessed 20 Aug 2014).
32. Roos LL, Roos NP, Cageorge SM, *et al.* How good are the data? Reliability of one health care data bank. *Med Care* 1982;20:266–76.
33. Klabunde CN, Potosky AL, Legler JM, *et al.* Development of a comorbidity index using physician claims data. *J Clin Epidemiol* 2000;53:1258–67.
34. Carral F, Oliveira G, Aguilar M, *et al.* Hospital discharge records under-report the prevalence of diabetes in inpatients. *Diabetes Res Clin Pract* 2003;59:145–51.
35. Horner RD, Paris JA, Purvis JR, *et al.* Accuracy of patient encounter and billing information in ambulatory care. *J Fam Pract* 1991;33:593–8.
36. O'Malley KJ, Cook KF, Price MD, *et al.* Measuring diagnoses: ICD code accuracy. *Health Serv Res* 2005;40:1620–39.
37. Goldman N, Lin IF, Weinstein M, *et al.* Evaluating the quality of self-reports of hypertension and diabetes. *J Clin Epidemiol* 2003;56:148–54.
38. Kriegsman DM, Penninx BW, van Eijk JT, *et al.* 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* 1996;49:1407–17.
39. 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* 1996;86:706–11.
40. Garay-Sevilla ME, Malacara JM, Gutiérrez-Roa A, *et al.* Denial of disease in type 2 diabetes mellitus: its influence on metabolic control and associated factors. *Diabet Med* 1999;16:238–44.