BMJ Open

Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-009952
Article Type:	Research
Date Submitted by the Author:	17-Sep-2015
Complete List of Authors:	Khokhar, Bushra; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health Jette, Nathalie; University of Calgary, O'Brien Institute for Public Health; University of Calgary, Clinical Neurosciences Metcalfe, Amy; University of Calgary, Department of Obstetrics and Gynecology; Alberta Children's Hospital Research Institute, Cunningham, Ceara Tess; University of Calgary, Community Health Sciences Quan, Hude; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health Kaplan, Gilaad; University of Calgary; University of Calgary, O'Brien Institute for Public Health Butalia, Sonia; University of Calgary, Community Health Sciences Rabi, Doreen; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health
Primary Subject Heading:	Public health
Secondary Subject Heading:	Health informatics, Epidemiology, Diabetes and endocrinology, Health services research, Diagnostics
Keywords:	diabetes, validation studies, case definition, administrative data

Title: Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data

Corresponding Author: Bushra Khokhar
Postal Address: 3280 Hospital Drive NW, 3rd floor TRW Calgary, Alberta T2N 4N1, Canada
E-mail: bushra.khokhar@ucalgary.ca
Telephone Number: (403) 210-3807
Author List: 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[1], and Doreen Rabi[1,2]

*Corresponding author
[1] Department of Community Health Sciences, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[2] O'Brien Institute for Public Health, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[3] Department of Clinical Neurosciences & Hotchkiss Brain Institute, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada
[4] Department of Obstetrics and Gynecology, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada
[5] Alberta Children's Hospital Research Institute, 3330 Hospital Drive NW, Calgary, Alberta T2N 4NI, Canada

Keywords: diabetes, validation studies, case definition, and administrative data

Word Count (excluding title page, abstract, references, figures and tables): 2,484 words

ABSTRACT

Objectives: Diabetes surveillance systems provide information about the distribution of diabetes within populations. Administrative health data are frequently used for surveillance, however several different case definitions have been developed. We undertook a systematic review to examine the validity of different case definitions across a variety of data sources.

Methods: Electronic databases (Medline and Embase) were systematically searched for validation studies where an administrative data diabetes case definition (using International Classification of Diseases codes) was validated against a reference and test measures reported.

Results: Search strategy identified 2,895 abstracts among which 18 studies were included. In studies using physician claims data, sensitivity ranged from 26.9 to 97%, specificity ranged from 94.3 to 99.4%, and PPV ranged from 71.4 to 96.2%. In studies using hospital discharge data, sensitivity ranged from 59.1 to 92.6%, specificity ranged from 95.5 to 99%, and PPV ranged from 62.5 to 96%. In studies using both physician claims data and hospital discharge data, the sensitivity ranged from 72 to 95.6%, specificity ranged from 88 to 98.5%, and PPV ranged from 54 to 80%.

Conclusions: This review demonstrates that the more data sources used (physician claims and hospital discharges), the longer the observation period, the better the definition performed. The outcomes with respect to sensitivity, specificity and PPV for each seem to differ due to variations in the definition of primary diagnosis in health data, the use of hospital discharge versus physician claims, by the type of ICD coding system used, and by geographical location. Overall, administrative health databases are useful for undertaking diabetes surveillance but the awareness of variation in sensitivity, specificity, PPV, NPV and kappa being affected by disease case definition is significant.

STRENGTHS AND LIMITATIONS

- 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.
- Most of the studies, 17 out of the 18[13-21, 23-30] 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.
- Lastly, 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 articles whose 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. The accuracy of charts reviews depends principally on physician documentation availability of records, and the accuracy of coding[34]. 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 individuals 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 an individual's level of education can have an effect on the reporting of diabetes[35-37]. Those with poorly controlled diabetes have been found to underreport their disease status[38].
- Lastly, difference between type 1 diabetes mellitus and type 2 diabetes mellitus is not clear in studies using administrative databases. In this review we included only those ≥ 18 years of age that is primarily the type 2 diabetes mellitus population.

BACKGROUND

Diabetes is a chronic disease that has increased substantially during the past 20 years[1]. At present, diabetes is the leading cause of blindness[2], renal failure[3] and non-traumatic lower limb amputations[4] and is a major risk factor for cardiovascular disease[5]. Because of 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 health care 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 analyze routinely collected health-related information from healthcare systems and provide reports on risk factors, care practices, morbidity, 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 therefore the need for health administrative data users to examine the validity of case ascertainment in their data sources before use[8].

Surveillance depends on a consistent case definition of diabetes. 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 their potential effect on prevalence estimation. However, our study will add to the literature, as our objective is to systematically review validated International Classification of Diseases (ICD), 9th edition (ICD-9) and ICD-10 based case definitions for diabetes and to compare the validity of different case definitions across studies and countries and not restrict it to a particular case definition. This is particularly important because many countries do not have outpatient data.

A consistent case definition needs to be validated in order to minimize misclassification bias and to be able to compare studies. The aim of this study was to provide recommendations for researchers on the optimal case definition to use for diabetes case ascertainment in administrative health data.

METHODS

Search Strategy

This systematic review was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[11]. Two citation databases, Medline and Embase, were searched using an OVID platform up from 1980 until September, 2015. The search strategy consisted of the following set of terms (Appendix A): (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) the 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 (e.g. ICD-9-CM).

Study Selection

The articles were evaluated for eligibility in a two-stage procedure, in duplicate - in stage one, all identified titles and abstracts were reviewed and in stage two, a full text review was performed on all of the articles that met the predefined eligibility criteria as well as all articles for which there was uncertainty as to eligibility. If either reviewer defined an article as eligible, in stage one, it was included in the full-text review, in stage two, disagreements were resolved by discussion or consultation with a third reviewer.

Inclusion/Exclusion Criteria

An article was considered included in the systematic review if it met the following criteria: (1) study population included those ≥ 18 years with type 1 diabetes mellitus or type 2 diabetes mellitus (2) statistical estimates [sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) or, kappa] 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 (e.g. self-report from population-based surveys or patient medical chart reviews); and (5) if it reported on original data. Articles that validated diabetes in specialized populations (e.g. cardiovascular disease) were excluded to ensure the case definitions would be generalizable to the general population. Papers that did not employ solely medical encounter data in their definitions (e.g. the inclusion of pharmacy or laboratory data) were also excluded, as the independent validity of the administrative definition could not be calculated. Bibliographies of included articles were manually searched for additional articles, which were then screened and reviewed using the same methods described above.

Data Extraction and Quality Assessment

The primary outcomes were sensitivity, specificity, PPV, NPV and kappa reported for each of the ICD-coded diabetes definitions. Other extracted data included sample size, age, and ICD codes used. If test measures 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 case definitions and reference standards used across studies. Data were tabulated by 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 2,895 abstract were identified with 193 articles reviewed in full text, of which 18 articles met all eligibility criteria (Figure 1) (Table 1). Ten of these studies were conducted in the United States[15, 17, 19, 21, 25-30], seven in Canada[13-14,16, 18, 20, 23-24], and one in Australia[20]. Fourteen studies used ICD-9 codes[13-17, 19, 22-29] and the remaining four studies used both ICD-9 and ICD-10 codes[16, 18, 20-21]. 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 18 studies reviewed, 10 used medical records[13, 15-23] and 8 used either self-reported surveys or telephone surveys to validate the diabetes diagnosis[14, 24-30]. Eight studies used physician claims data[15-17, 25-26, 28-30], three studies used hospital discharge data [18, 22-23] while five studies used a combination of both[15-16, 22, 26, 29]. Two studies used electronic medical records (EMRs) as their health data source[19, 21].

The scores (Table 2) ranged from 9 to 13, out of a maximum of 14. Regardless of quality assessment scores, all 18 studies are discussed in this systematic review.

The sample size varied from 95 to ~3 million people. Sensitivity and specificity values were available from

all 18 studies, PPV in 16 studies, NPV in 12 studies, and kappa in six. In studies using physician claims data, 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 kappa ranged from 0.8 to 0.9. In studies using 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 kappa ranged from 0.6 to 0.9. In studies using both physician claims data and hospital discharge data, the sensitivity ranged from 72 to 95.6%, specificity ranged from 88 to 98.5%, PPV ranged from 54 to 80%, NPV ranged from 98 to 99.6% and kappa ranged from 0.7 to 0.8. In the two studies using EMRs as their health data source, sensitivity ranged from 71 to 100%, specificity ranged from 98 to 100%, and PPV ranged from 21 to 100%.

A secondary tabulation of data was performed by the type of ICD coding system used. 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 kappa ranged from 0.6 to 0.9 whereas, in the studies using ICD-10 codes, sensitivity ranged from 59.1 to 89.6%, specificity ranged from 95.5 to 99%, PPV ranged from 63.1 to 96%, NPV ranged 90.8 to 98.9%, and kappa ranged from 0.6 to 0.9.

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, by the type of ICD coding system used, and by the geographical location.

DISCUSSION

The validity of administrative case definitions for diabetes varies significantly across studies, but we identified definition features that were associated with better performance. The combinations of more than one physician claim and/or hospital discharge encounter along with a longer observation period consistently performed better. Certain definitions, such as the definition used by the National Diabetes Surveillance System (NDSS) to identify Canadians with diabetes mellitus[31] used a combination of data sources (physician claims and hospital discharge data) and has been shown to have high validity in this study and other validation studies that were not eligible for this review. In a previous examination of administrative database definitions for diabetes, a meta-analysis[10] demonstrated that this commonly-used administrative database definition for diabetes (two physician outpatient billings and/or one hospitalization with a diabetes record on the discharge abstract summary within a two-year period) has a pooled sensitivity of 82.3% (95% CI 75.8, 87.4) and specificity of 97.9% (95% CI 96.5, 98.8%).

Approaches used in developing case definitions for diabetes can be simple and practical and result in high sensitivity, specificity and PPV. This systematic review, which reviewed the performance of a number of ICD-9 and ICD-10 based case definitions for diabetes, provides new knowledge on factors that are associated with enhanced definition performance. It also demonstrated a wide variation in definition performance that we speculate may be related to the type of administrative data source (physician claims, hospital discharge data, and a combination of the two) and the study purpose. In addition, method of data collection, purpose of collection, availability of the type of data and clinical detail of data on hand are other factors that introduced variability across studies.

Studies included in this systematic review used a variety of case definitions to identify patients with diabetes. These definitions include hospital discharge data or physician claims or some form of a combination of the two. It is important to understand the difference in accuracy between these definitions. 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 standardized manner. Billing practices do vary by practitioner, which may in turn be influenced by the nature of physician reimbursement (salary versus fee for service). Further, individuals 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 [32-33]. In contrast, hospital discharge data are limited to clinical information that is relevant to an individual hospitalization, capturing diagnostic and treatment information usually for a brief window of time. However, the advantage of hospital discharge data for surveillance is that discharge diagnostic and medical procedure information are recorded by medical coders with standardized training with a detailed review of medical charts.

What are considered 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 PPV and sensitivity 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 hospitalization, within a certain period of time is appropriate.

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. Most of the studies, 17 out of the 18[13-21, 23-30] 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. Lastly, 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 articles whose 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. The accuracy of charts reviews depends principally on physician documentation availability of records, and the accuracy of coding[34]. 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 individuals 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 an individual's level of education can have an effect on the reporting of diabetes[35-37]. Those with poorly controlled diabetes have been found to underreport their disease status[38].

Lastly, difference between type 1 diabetes mellitus and type 2 diabetes mellitus is not clear in studies using administrative databases. In this review we included only those ≥ 18 years of age that is primarily the type 2 diabetes mellitus population.

Generalizability

As previously mentioned, 90% of included studies were conducted in North America and therefore these validation studies are highly comparable. 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.

CONCLUSION

This review demonstrates that the more data sources used (physician claims and hospital discharges), the longer the observation period, the better the definition performed. A conclusive recommendation of an optimal definition cannot be made because the definition depends on the purpose of use and the availability of the type of data available on hand. 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[39-40] but the awareness of the variation in sensitivity, specificity, PPV, NPV and kappa being affected by disease case

definition is significant.

COMPETING INTERESTS

The authors declare that they have no competing interest.

AUTHOR'S CONTRIBUTIONS

Dr. Nathalie Jette wrote the protocol. Ms. Bushra Khokhar, Dr. Amy Metcalfe, and Ms. Ceara Tess Cunningham carried out the systematic review. Bushra Khokhar wrote the manuscript. Dr. Nathalie Jette, Dr. Hude Quan, Dr. Gilaad G. Kaplan, Dr. Sonia Butalia, and Dr. Doreen Rabi provided final approval of the version to be published. All authors read and approved the final manuscript.

FUNDING SOURCES

Ms. Bushra Khokhar was supported by the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD) and The Western Regional Training Centre for Health Services Research (WRTC). Dr. Nathalie Jette 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. Ms. Ceara Tess Cunningham is funded by a Canadian Institute of Health Research doctoral research scholarship. Dr. Kaplan is a Population Health Investigator supported by Alberta Innovates - Health Solutions. Dr. Doreen Rabi is a Population Health Investigator supported by Alberta Innovates - Health Solutions.

DATA SHARING STATEMENT

There was no additional unpublished data used from any of the studies included in this systematic review.

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, and Dillon J. Early Diagnosis of Diabetes Through The Eye. *Photochem Photobiol* Published Online First 27 August 2015. doi: 10.1111/php.12524.
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 Sep 7;100:1134-46.
6) World Health Organization. Diabetes: The Cost of Diabetes. [http://www.who.int/mediacentre/factsheets/fs236/en/] Accessed August 27, 2014.
7) Jutte DR, Roos LL, and 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: 1792-99.
9) Centers for Disease Control and Prevention: National Notifiable Diseases Surveillance System (NNDSS). [http://wwwn.cdc.gov/nndss/case-definitions.html]. Accessed September 1, 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. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Ann. Intern. Med. 2009;339:b2700.
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) 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;25:512-6.
14) Robinson JR, Young TK, Roos LL, et al. Estimating the burden of disease. Comparing administrative data and self-reports. *Medical Care*. 1997;35:932-47.
15) 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.
16) Wilchesky M, Tamblyn RM, and Huang A. Validation of Diagnostic Codes within Medical Services Claims. *J Clin epidemiol.* 2004;57:131-41.
17) 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.
18) So L, Evans D, and Quan H. ICD-10 coding algorithms for defining comorbidities of acute myocardial infarction. *BMC Health Serv Res*. 2006;6.
19) Zgibor JC, Orchard TJ, Saul M, et al. Developing and Validating a Diabetes Database in a Large Health System. *Diabetes Res Clin Pract.* 2007;75:313-9.
20) 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.
21) Kandula S, Zeng-Treitler Q, Chen L, et al. A Bootstrapping Algorithm to Improve Cohort Identification Using Structured Data. *J Biomed Inform*. 2011;44:S63-68.
22) 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.
23) 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;4: 1424-41.
24) Young TK, Roos NP, and Hammerstrand KM. Estimated burden of diabetes mellitus in Manitoba according to health-insurance claims: a pilot study. *CMAJ*. 1991;144:318-24.
25) Hebert PL, Geiss LS, Tierney EF, et al. Identifying Persons with Diabetes Using Medicare Claims Data. *Am J Med Qual.* 1999;14:270-7.
26) 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.
27) 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.
28) Miller DR, Safford MM, and 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:B10-21.
29) Singh JA. Accuracy of Veterans Affairs Databases for Diagnoses of Chronic Diseases. *Preventative Chronic Disease*. 2009;6:A126.
30) O'Connor PJ, Rush W A, 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.
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

August 28, 2014.
32) Carral F, Olveira G, Aguilar M, et al. Hospital discharge records under-report the prevalence of diabetes in inpatients. *Diabetes Res Clin Pract*. 2003;59:145-51.
33) Horner RD, Paris JA, Purvis JR, et al. Accuracy of patient encounter and billing information in ambulatory care. *J Fam Pract*. 1991;33:593–598.
34) O'Malley KJ, Cook KF, Price MD, et al. Measuring Diagnoses: ICD Code Accuracy. *Health Serv Res*. 2005;40:1620-39.
35) 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.
36) 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. *J Clin Epidemiol*. 1996;49:1407- 17.
37) Mackenbach JP, Looman CW, and van deer 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.
38) Garay-Sevilla ME, Malacara JM, Gutierrez-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.
39) King NA, Caudwell P, Hopkins M, et al. Metabolic and Behavioral Compensatory Responses to Exercise Interventions: Barriers to Weight Loss. *Obesity*. 2007;15:1373-83.
40) Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. [http://www.cdc.gov/MMWR/preview/mmwrhtml/00020852.htm] Accessed August 28, 2014.

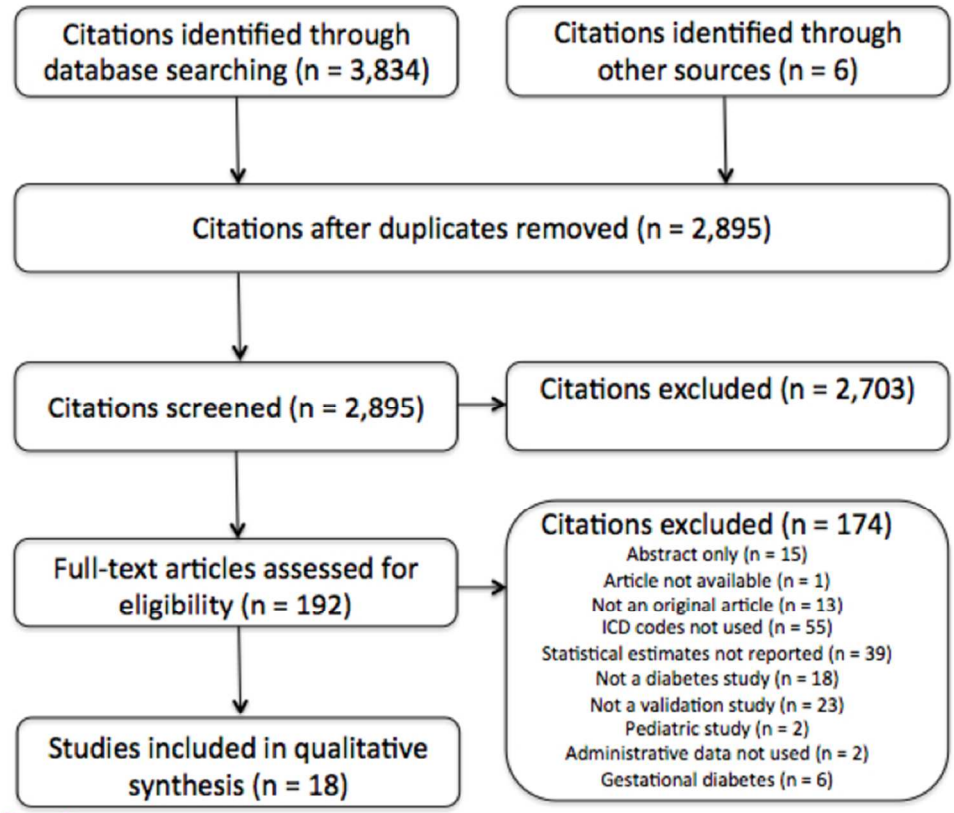


Figure 1. Study Flow

269x219mm (72 x 72 DPI)

noses recorded in the claims of study physicians	2,752	51.78 (49.9, 53.6)	98.41 (98.2, 98.6)				ICD-9 250.0 - 250.9
odes recorded on claims made o provided medical services to r prior to the start of the study		64.43 (62.6, 66.2)	96.82 (96.5, 97.1)				
one clinician-coded diagnoses	1,441	93 (86, 100)	99 (99, 100)	91 (83, 99)			ICD 9 250.0, 250.1, 250.2, 250.3
osis in National Department of A) database, Out-Patient Clinic file over one year	1,176	97	96			0.92	ICD 9 250.x
osis in National Department of A) database, Out-Patient Clinic file over one year						0.91	
osis in National Department of A) database, Out-Patient Clinic file over two years						0.89	
osis in National Department of A) database, Out-Patient Clinic file over two years						0.93	
gnoses of diabetes in any claim file over 1-year period		71.6	96.6	79			ICD 9-CM 250.00 - 250.93, 357.2, 362.0 - 362.02, 366.41
gnoses of diabetes in any claim file over 2-year period		79.1	94.3	71.4			
or more ICD-9 diagnostic codes	1,976	92.22*	98.62*	76.15*	99.63*		ICD 9 250.x
		76	98	91	95	0.79	

diagnosis code						(0.76 - 0.85)	
Any diagnostic code	2,924,148	78.3	95.7	85.3			ICD 9 250, 357.2, 362.0, 366.41
ny out-patient diagnostic code		77.5	95.9	85.8			
Any in-patient diagnostic code		26.9	99.4	93.7			
t codes OR ≥ 1 in-patient code		73.7	98.1	92.8			
≥ 2 any diagnostic code		73.1	98.3	93.4			
≥ 2 out-patient codes		72.2	98.4	93.7			
≥ 3 any diagnostic code		69	98.4	95.2			
≥ 3 out-patient codes		68	98.9	95.4			
≥ 4 any diagnostic code		65	99.1	96			
≥ 4 out-patient codes		63.8	99.2	96.2			
Diabetes with Complications	93	80 (51.91, 95.67)	98.3 (95.15, 99.65)	80 (51.91, 95.67)	98.3 (95.15, 99.65)		ICD-9 250.1 -250.9
		66.7 (38.38, 88.18)	98.9 (96.00, 99.86)	83.3 (51.59, 97.91)	97.2 (93.67, 99.10)		ICD-10 E10.0 -E10.8, E11.0 - E11.8, E12.0 -E12.8, E13.0 -E13.8, E14.0 -E14.8
tes with Chronic Complications	4,008	63.6	98.9	62.5	99	0.62	ICD 9 250.4 -250.7
		59.1	99	63.1	98.9	0.6	ICD 10 E10.2 - E10.5, E10.7, E11.2 - E11.5, E11.7, E12.2 - E12.5, E12.7, E13.2 - E13.5, E13.7, E14.2 - E14.5, E14.7,
without Chronic Complications		77.7	98.4	86.5	97	0.8	ICD 9 250.0 -250.3, 250.8, 250.9
		75.8	98.7	88.5	96.8	0.79	E10.0, E10.1, E10.6, E10.8, E10.9, E110, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9,

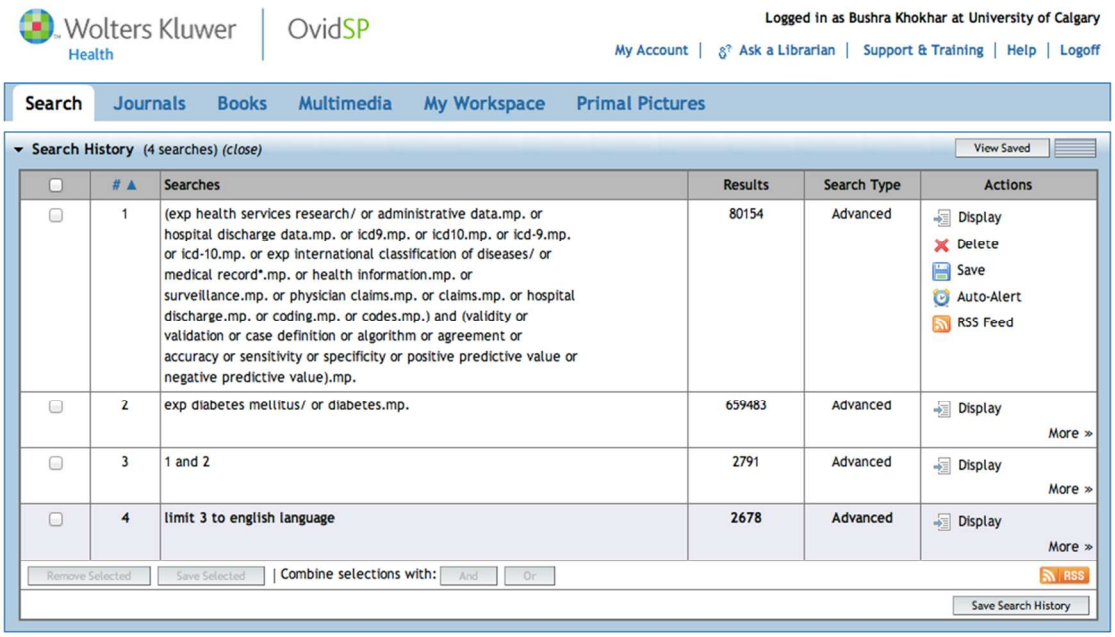
2-years		87.3	96.7	93.5	93.4	0.854	
5-years		89.3	95.9 †	92.2	94.4	0.859	
10-years		89.6	95.6 †	91.6	94.5	0.856	
15-years		89.6	95.5 †	91.5	94.5	0.855	
claim or 1 hospitalization over 3 years	2,651	72	98	76	98	0.72 (0.67 - 0.77)	ICD 9 CM
Physician Service Claims or One tion with diagnosis of diabetes	3,317	91	92*	61	99*		ICD-9 250.x
Physician Service Claims or One tion with diagnosis of diabetes		86	97*	80	98*		
Years Observation Period Data	3,362	95.6 (92.5–97.7)	92.8 (91.9–93.7)	54 (49.6–58.5)	99.6 (99.4–99.8)	0.65 (0.61–0.69)	ICD 9 250.xx ICD 10 E10.x–E14.x
Years Observation Period Data		86.4 (82.4–90.5)	97.1 (96.5–97.7)	72.4 (67.5–77.3)	98.8 (98.4–99.2)	0.77 (0.73–0.81)	
Years Observation Period Data		91.2 (87.9–94.6)	97.6 (97.1–98.1)	72.1 (67.5–76.9)	99.2 (98.9–99.5)	0.82 (0.78–0.85)	
Years Observation Period Data		76.6 (71.5–81.6)	99.3 (99.0–99.6)	90.9 (87.2–94.6)	98 (97.5–98.4)	0.82 (78.0–85.5)	
One or more ICD - 9 in the onic Health Record database	200	71	100	100	74		ICD 9 250.x
e for diabetes in the outpatient record	254	79	98	94			ICD 9 250

							366.41
Two 1999 claims with dx		85	96				
One 1999 or 2000 claim with dx		95	88				
Two 1999 or 2000 claims with dx		93	93				
s of provincial residents claims mitted to the Manitoba Health mission (MHSC)] AND [Hospital f provincial residents claims for mitted to the Manitoba Health ion (MHSC) AND Claims by the o the Manitoba Health Services mmission (MHSC) or payment]	1,000	82.7	96.3				ICD 9-CM
s of provincial residents claims mitted to the Manitoba Health ion (MHSC) AND Claims by the o the Manitoba Health Services C) or payment] AND [Claims by o the Manitoba Health Services mmission (MHSC) or payment]		82.1	98.5				
s of provincial residents claims mitted to the Manitoba Health mission (MHSC)] AND [Hospital f provincial residents claims for mitted to the Manitoba Health ion (MHSC) AND Claims by the o the Manitoba Health Services C) or payment] AND [Claims by o the Manitoba Health Services mmission (MHSC) or payment]		83.9	95.8				

Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unclear	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Unclear
No	No	No	No	No	No	No	No	No	No	No	No	No	Yes
No	No	No	Unclear	No	Yes	Unclear	Unclear	Unclear	No	No	No	Unclear	Unclear
10	9	11	11	11	13	12	12	11	11	12	12	12	12

Appendix A: Search Strategies

Embase Search Criteria



Medline Search Criteria

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE: Systematic Review of Validated Case Definition for Diabetes in ICD-9 and ICD-10 Coded Data			
Title	1	Title identifies study as a systematic review.	1
ABSTRACT			
Structured summary	2	An abstract is provided including, background, methods, results, and conclusion.	2
INTRODUCTION			
Rationale	3	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 therefore, the need for health administrative data users to examine the validity of disease case ascertainment in their data sources before use. Surveillance depends on a consistent case definition of diabetes. However, a variety of diabetes case definition exists, resulting in variation in reported diabetes prevalence estimates.	
Objectives	4	The purpose of the present study is to perform a systematic review of validated ICD-9 and ICD-10 based case definitions for diabetes and to compare the validity of different case definitions across studies and countries and not restrict it to a particular case definition.	
METHODS			
Protocol and registration	5		
Eligibility criteria	6	An article was considered included in the systematic review if it met the following criteria: (1) study population included those ≥ 18 years with type 1 diabetes mellitus or type 2 diabetes mellitus (2) statistical estimates [sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) or, kappa] 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 (e.g. self-report from population-based surveys or patient medical chart reviews); and (5) if it reported on original data. Articles that validated diabetes in specialized populations (e.g. cardiovascular disease) were excluded to ensure the case definitions would be generalizable to the general population. Papers that did not employ solely medical encounter data in their definitions (e.g. the inclusion of pharmacy or laboratory data) were also excluded, as the independent validity of the administrative definition could not be calculated. Bibliographies of included articles were manually searched for additional articles, which were then screened and reviewed using the same methods described above.	
Information sources	7	This systematic review was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[11]. Two citation databases, Medline and Embase, were searched using an OVID platform up from 1980 until September, 2015. The search strategy consisted of the following set of terms: (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 or case definition or algorithm	

PRISMA 2009 Checklist

		or agreement or accuracy or sensitivity or specificity or positive predictive value or negative predictive value] AND (3) the 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 (e.g. ICD-9-CM).	
Search	8	Exact search strategy used in Medline and Embase shown in Appendix A.	
Study selection	9	The articles were evaluated for eligibility in a two-stage procedure, in duplicate - in stage one, all identified titles and abstracts were reviewed and in stage two, a full text review was performed on all of the articles that met the predefined eligibility criteria as well as all articles for which there was uncertainty as to eligibility. If either reviewer defined an article as eligible, in stage one, it was included in the full-text review, in stage two, disagreements were resolved by discussion or consultation with a third reviewer.	
Data collection process	10	Reviewers, independently, extracted data from all studies that satisfied the inclusion criteria. Any disagreement in data extraction and/or study inclusion was resolved through discussion between reviewers.	
Data items	11	The primary outcomes were sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and kappa scores reported for each of the ICD-coded diabetes definitions. Other extracted data included sample size, ICD codes used, and geographic location.	
Risk of bias in individual studies	12	Study quality was evaluated using the Quality Assessment tool for Diagnostic Accuracy Studies (QUADAS) criteria.	
Summary measures	13	Not applicable because calculating a pooled estimate of surveillance performance measures using meta-analytic techniques was deemed inappropriate given the heterogeneity of case definitions and references standards used across studies.	
Synthesis of results	14	Not applicable.	

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Not applicable.	
Additional analyses	16	Not applicable.	
RESULTS			
Study selection	17	See Flow Diagram (Figure 1)	
Study characteristics	18	See Study Characteristics and Test Measures of Studies (Table 1)	

PRISMA 2009 Checklist

Risk of bias within studies	19	Not applicable.	
Results of individual studies	20	Not applicable.	
Synthesis of results	21	Not applicable.	
Risk of bias across studies	22	Not applicable.	
Additional analysis	23	Not applicable.	
DISCUSSION			
Summary of evidence	24	The validity of administrative case definitions for diabetes varies significantly across studies, but we identified definition features that were associated with better performance. The combinations of more than one physician claim and/or hospital discharge encounter along with a longer observation period consistently performed better.	
Limitations	25	There is the potential for a language bias as articles whose 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. Difference between type 1 diabetes mellitus and type 2 diabetes mellitus is not clear in studies using administrative databases. In this review we included only those ≥ 18 years of age that is primarily the type 2 diabetes mellitus population.	
Conclusions	26	This review demonstrates that the more data sources used (physician claims and hospital discharges), the longer the observation period, the better the definition performed. A conclusive recommendation of an optimal definition cannot be made because the definition depends on the purpose of use and the availability of the type of data available on hand. 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[39-40] but the awareness of the variation in sensitivity, specificity, PPV, NPV and kappa being affected by disease case definition is significant.	
FUNDING			
Funding	27	Ms. Bushra Khokhar was supported by the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD) and The Western Regional Training Centre for Health Services Research (WRTC). Dr. Nathalie Jette 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. Ms. Ceara Tess Cunningham is funded by a Canadian Institute of Health Research doctoral research scholarship. Dr. Kaplan is a Population Health Investigator supported by Alberta Innovates - Health Solutions. Dr. Doreen Rabi is a Population Health Investigator supported by Alberta Innovates - Health Solutions.	

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

BMJ Open

Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-009952.R1
Article Type:	Research
Date Submitted by the Author:	04-Jan-2016
Complete List of Authors:	Khokhar, Bushra; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health Jette, Nathalie; University of Calgary, O'Brien Institute for Public Health; University of Calgary, Clinical Neurosciences Metcalfe, Amy; University of Calgary, Department of Obstetrics and Gynecology; Alberta Children's Hospital Research Institute, Cunningham, Ceara Tess; University of Calgary, Community Health Sciences Quan, Hude; University of Calgary, O'Brien Institute for Public Health; University of Calgary, Community Health Sciences Kaplan, Gilaad; University of Calgary; University of Calgary, O'Brien Institute for Public Health Butalia, Sonia; University of Calgary, Community Health Sciences Rabi, Doreen; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health informatics, Epidemiology, Diabetes and endocrinology, Health services research, Diagnostics
Keywords:	diabetes, validation studies, case definition, administrative data

Title: Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data

Corresponding Author: Bushra Khokhar
Postal Address: 3280 Hospital Drive NW, 3rd floor TRW Calgary, Alberta T2N 4N1, Canada
E-mail: bushra.khokhar@ucalgary.ca
Telephone Number: (403) 210-3807
Author List: 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], and Doreen Rabi[1,2,6]

*Corresponding author
[1]Department of Community Health Sciences, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[2]O'Brien Institute for Public Health, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[3]Department of Clinical Neurosciences & Hotchkiss Brain Institute, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada
[4]Department of Obstetrics and Gynecology, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada
[5]Alberta Children's Hospital Research Institute, 3330 Hospital Drive NW, Calgary, Alberta T2N 4NI, Canada
[6] Division of Endocrinology, Department of Medicine, 1820 Richmond Road SW, Calgary, Alberta TCC CCC

Keywords: diabetes, validation studies, case definition, and administrative data

Word Count (excluding title page, abstract, references, figures and tables): 3,114 words

ABSTRACT

Objectives: Diabetes surveillance systems provide information about the distribution of diabetes within populations. Administrative health data are frequently used for surveillance; however several different case definitions have been developed. We undertook a systematic review to examine the validity of different case definitions across a variety of data sources.

Methods: Electronic databases (Medline and Embase) were systematically searched for validation studies where an administrative data diabetes case definition (using International Classification of Diseases codes) was validated against a reference and test measures reported.

Results: Search strategy identified 2,895 abstracts, among which 18 studies were included. In studies using physician claims data, sensitivity ranged from 26.9 to 97%, specificity ranged from 94.3 to 99.4%, and PPV ranged from 71.4 to 96.2%. In studies using hospital discharge data, sensitivity ranged from 59.1 to 92.6%, specificity ranged from 95.5 to 99%, and PPV ranged from 62.5 to 96%. In studies using both physician claims data and hospital discharge data, the sensitivity ranged from 72 to 95.6%, specificity ranged from 88 to 98.5%, and PPV ranged from 54 to 80%.

Conclusions: Overall, administrative health databases are useful for undertaking diabetes surveillance through combining physician claims and hospital discharge data, but the awareness of variation in sensitivity, specificity, PPV, NPV and kappa being affected by disease case definition is significant. This review demonstrates that sensitivity and specificity will vary with data sources used (physician claims or hospital discharges) and duration of observation period. A conclusive recommendation of an optimal definition cannot be made because the definition depends on the purpose of surveillance and the availability of the type of data available on hand. Approaches used in developing case definitions for diabetes can be simple and practical and result in high sensitivity, specificity and PPV.

STRENGTHS AND LIMITATIONS

- 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.
- Most of the studies, 15 out 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.
- Lastly, 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 articles whose 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. The accuracy of charts reviews depends principally on physician documentation availability of records, and the accuracy of coding[34]. 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 individuals 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 an individual's level of education can have an effect on the reporting of diabetes[35-37]. Those with poorly controlled diabetes have been found to underreport their disease status[38].

- Lastly, difference between type 1 diabetes mellitus and type 2 diabetes mellitus is not clear in studies using administrative databases. In this review we included only those ≥ 18 years of age that is primarily the type 2 diabetes mellitus population.

BACKGROUND

Diabetes is a chronic disease that has increased substantially during the past 20 years[1]. At present, diabetes is the leading cause of blindness[2], renal failure[3] and non-traumatic lower limb amputations[4] and is a major risk factor for cardiovascular disease[5]. Because of 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 health care 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 analyze routinely collected health-related information from healthcare systems and provide reports on risk factors, care practices, morbidity, 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 therefore the need for health administrative data users to examine the validity of case ascertainment in their data sources before use[8].

Surveillance depends on a consistent case definition of diabetes. 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 their potential effect on prevalence estimation. However, our study will add to the literature, as our objective is to systematically review validated International Classification of Diseases (ICD), 9th edition (ICD-9) and ICD-10 based case definitions for diabetes and to compare the validity of different case definitions across studies and countries and not restrict it to a particular case definition. This is particularly important because many countries do not have outpatient data.

A consistent case definition needs to be validated in order to minimize misclassification bias and to be able to compare studies. The aim of this study was to provide recommendations for researchers on the optimal case definition to use for diabetes case ascertainment in administrative health data.

METHODS

Search Strategy

This systematic review was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[11][Appendix A]. Two citation databases, Medline and Embase, were searched using an OVID platform up from 1980 until September, 2015. The search strategy consisted of the following set of terms [Appendix]: (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) the 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 (e.g. ICD-9-CM).

Study Selection

The articles were evaluated for eligibility in a two-stage procedure, in duplicate - in stage one, all identified titles and abstracts were reviewed and in stage two, a full text review was performed on all of the articles that met the predefined eligibility criteria as well as all articles for which there was uncertainty as to eligibility. If either reviewer defined an article as eligible, in stage one, it was included in the full-text review, in stage two, disagreements were resolved by discussion or consultation with a third reviewer.

Inclusion/Exclusion Criteria

An article was considered included in the systematic review if it met the following criteria: (1) study population included those ≥ 18 years with type 1 diabetes mellitus or type 2 diabetes mellitus (2) statistical estimates [sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) or, kappa] 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 (e.g. self-report from population-based surveys or patient medical chart reviews); and (5) if it reported on original data. Articles that validated diabetes in specialized populations (e.g. cardiovascular disease) were excluded to ensure the case definitions would be generalizable to the general population. Papers that did not employ solely medical encounter data in their definitions (e.g. the inclusion of pharmacy or laboratory data) were also excluded, as the independent validity of the administrative definition could not be calculated. Bibliographies of included articles were manually searched for additional articles, which were then screened and reviewed using the same methods described above.

Data Extraction and Quality Assessment

The primary outcomes were sensitivity, specificity, PPV, NPV, and kappa reported for each of the ICD-coded diabetes definitions. Other extracted data included sample size, age, and ICD codes used. If test measures 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 case definitions and reference standards used across studies. Data were tabulated by 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 2,895 abstract were identified with 193 articles reviewed in full text, of which 16 articles met all eligibility criteria (Figure 1). Eight of these studies were conducted in the United States[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 both 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] while four studies used a combination of both[17, 21, 25, 27]. Two studies used electronic 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, out of a maximum of 14. Five questions were selected from QUAS 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 kappa in six. All 16 studies were categorized by the type of 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 kappa ranged from 0.8 to 0.9. Four out of the eight studies using physician claims data had a least one case definition where both sensitivity and specificity exceed 80%.

Studies comparing physician claims data definition over a multiple years time period[13, 15-16] consistently show increases in sensitivity values and a slight decrease in specificity and PPV with increase in time duration. This relationship is consistent with the study[18] looking at changes in the statistical estimates with increasing the number of appearance of diagnostic codes in the definition – the sensitivity was the highest when any diagnostic code was used, the specificity was the highest when any diagnostic code was used, the PPV was the highest when ≥ 4 out-patients codes were used.

Hospital Discharge Data

Table 3 lists the four studies[22, 24, 26, 28] using hospital discharge data alone. 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 kappa ranged from 0.6 to 0.9. Two out of the four studies using hospital discharge data had a least one case definition where both sensitivity and specificity exceed 80%. In contrast to the physician claims based definitions, the sensitivity seemed to improve when a longer duration was used the 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 kappa 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 admissions based definitions alone. All four of the studies using a combination of physician claims and hospital discharge data had a least one case definition where both sensitivity and specificity exceed 80%.

Studies comparing hospital discharge data definition over a multiple years[21, 25] consistently show similar increase in sensitivity and decrease in specificity and PPV with increase in time period similar to the definitions using physician claims, as shown above. Another factor affecting the statistical estimates is use of number of claims in the definition. Rector et al.'s study[17] shows consistent results where the sensitivity is higher when at least one claims data is used in the definition, but the specificity is higher when at least two are used. Lastly, Young et al.'s study[27] demonstrates the highest sensitivity when 2 physician claims and 2 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. 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 kappa 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 kappa ranged from 0.6 to 0.9. Eight studies using ICD-9 coding systems are from the United States and four studies from Canada. Four studies use both ICD-9 and ICD-10 coding systems – three of these are from Canada and one from Western Australia.

In this systematic review, case definitions appear to perform more reliably 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, by the type of ICD coding system used, and by the geographical location.

DISCUSSION

The validity of administrative case definitions for diabetes varies significantly across studies, but

we identified definition features that were associated with more reliable performance. The combinations of more than one data source, physician claim and /or hospital discharge encounter along with an observation period of more than one 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 (NDSS) to identify Canadians with diabetes mellitus[31] as it uses a combination of data sources (physician claims and hospital discharge data) and has been shown to have high validity in this study and other validation studies that were not eligible for this review. In a previous examination of administrative database definitions for diabetes, a meta-analysis[10] demonstrated that this commonly-used administrative database definition for diabetes (two physician outpatient billings and/or one hospitalization with a diabetes record on the discharge abstract summary within a two-year period) has a pooled sensitivity of 82.3% (95% CI 75.8, 87.4) and specificity of 97.9% (95% CI 96.5, 98.8%).

This systematic review, which reviewed the performance of a number of ICD-9 and ICD-10 based case definitions for diabetes, 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, coding system 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 important information on how given 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 standardized manner. Billing practices do vary by practitioner, which may in turn be influenced by the nature of physician reimbursement (salary versus fee for service). Further, individuals 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[32-33]. In contrast, hospital discharge data are limited to clinical information that is relevant to an individual hospitalization, 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 standardized training with a detailed review of medical charts. However, the standard method of discharge coding does vary regionally and thus one will still see variation around validity estimates based on these differences in coding practices.

What are considered 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 PPV and sensitivity 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 hospitalization, within a certain period of time is appropriate. It is also important to recognize 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. Most of the studies, 15 out 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. Lastly, 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 articles whose 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. The accuracy of charts reviews depends principally on physician documentation availability of records, and the accuracy of coding[34]. 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 individuals 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 an individual's level of education can have an effect on the reporting of diabetes[35-37]. Those with poorly controlled diabetes have been found to underreport their disease status[38].

Lastly, 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.

Generalizability

As previously mentioned, 90% of included studies were conducted in North America and therefore these validation studies are highly comparable. 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.

CONCLUSION

Most studies use similar case definitions that require one or more diagnoses of diabetes using ICD-9 (or ICD-9-CM) or ICD-10. The overall quality of most of the case definitions included in this review are good. The performance characteristics of these case definitions depends on the version of ICD coding system used, variations in the definition of primary diagnosis in ICD-coded health data, the use of fee-for service payment model or salary based model for physicians and primary care providers, and/or the methodology adopted by the health care facility to extract information from patient records.

This review demonstrates that sensitivity and specificity will vary with data sources used (physician claims or hospital discharges) and duration of observation period. A conclusive

recommendation of an optimal definition cannot be made because the definition depends on the purpose of use and the availability of the type of data available on hand. 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[39-40] but the awareness of the variation in performance characteristics being affected by disease case definition is significant.

COMPETING INTERESTS

The authors declare that they have no competing interest.

DATA SHARING

No additional data available.

AUTHOR'S CONTRIBUTIONS

Dr. Nathalie Jette wrote the protocol. Ms. Bushra Khokhar, Dr. Amy Metcalfe, and Ms. Ceara Tess Cunningham carried out the systematic review. Bushra Khokhar wrote the manuscript. Dr. Nathalie Jette, Dr. Hude Quan, Dr. Gilaad G. Kaplan, Dr. Sonia Butalia, and Dr. Doreen Rabi provided final approval of the version to be published. All authors read and approved the final manuscript.

FUNDING SOURCES

Ms. Bushra Khokhar was supported by the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD) and The Western Regional Training Centre for Health Services Research (WRTC). Dr. Nathalie Jette 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. Ms. Ceara Tess Cunningham is funded by a Canadian Institute of Health Research doctoral research scholarship. Dr. Kaplan is a Population Health Investigator supported by Alberta Innovates - Health Solutions. Dr. Doreen Rabi is a Population Health Investigator supported by Alberta Innovates - Health Solutions.

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, and Dillon J. Early Diagnosis of Diabetes Through The Eye. *Photochem Photobiol* Published Online First 27 August 2015. doi: 10.1111/php.12524.
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 Sep 7;100:1134-46.
6) World Health Organization. Diabetes: The Cost of Diabetes. [http://www.who.int/mediacentre/factsheets/fs236/en/] Accessed August 27, 2014.
7) Jutte DR, Roos LL, and 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: 1792-99.
9) Centers for Disease Control and Prevention: National Notifiable Diseases Surveillance System (NNDSS). [http://wwwn.cdc.gov/nndss/case-definitions.html]. Accessed September 1, 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. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Ann. Intern. Med. 2009;339:b2700.
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, and 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:B10-21.
19) Singh JA. Accuracy of Veterans Affairs Databases for Diagnoses of Chronic Diseases. *Preventative Chronic Disease*. 2009;6:A126.
20) O'Connor PJ, Rush W A, 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;25:512-6.
22) Robinson JR, Young TK, Roos LL, et al. Estimating the burden of disease. Comparing

administrative data and self-reports. *Medical Care*. 1997;35:932-47.
23) Wilchesky M, Tamblyn RM, and Huang A. Validation of Diagnostic Codes within Medical Services Claims. *J Clin epidemiol.* 2004;57:131-41.
24) So L, Evans D, and Quan H. ICD-10 coding algorithms for defining comorbidities of acute myocardial infarction. *BMC Health Serv Res*. 2006;6.
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;4: 1424-41.
27) Young TK, Roos NP, and 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;75:313-9.
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: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 August 28, 2014.
32) Carral F, Olveira G, Aguilar M, et al. Hospital discharge records under-report the prevalence of diabetes in inpatients. *Diabetes Res Clin Pract*. 2003;59:145-51.
33) Horner RD, Paris JA, Purvis JR, et al. Accuracy of patient encounter and billing information in ambulatory care. *J Fam Pract*. 1991;33:593–598.
34) O'Malley KJ, Cook KF, Price MD, et al. Measuring Diagnoses: ICD Code Accuracy. *Health Serv Res*. 2005;40:1620-39.
35) 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.
36) 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. *J Clin Epidemiol*. 1996;49:1407- 17.
37) Mackenbach JP, Looman CW, and van deer 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.
38) Garay-Sevilla ME, Malacara JM, Gutierrez-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.
39) King NA, Caudwell P, Hopkins M, et al. Metabolic and Behavioral Compensatory Responses to Exercise Interventions: Barriers to Weight Loss. *Obesity*. 2007;15:1373-83.
40) Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. [http://www.cdc.gov/MMWR/preview/mmwrhtml/00020852.htm] Accessed August 28, 2014.

Table 1. Study Quality Characteristics using QUADAS Tool

QUADAS Tool Item	Hux[21]	Robinson[2]	Borzecki[1]	Wilchesky[2]	Crane[1]	So[24]	Chen[25]	Nedkoff[2]	Quan[26]	Young[2]	Hebert[1]	Ngo[1]	Rector[1]	Miller[1]	Singh[19]	O'Connor[2]
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 (i.e. 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

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 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[12].

Table 2: Study Characteristics and Test Measures of Studies for Physician Claims Data
(Superior performance characteristics within studies have been highlighted in bold.)

Country	Study Year	Author[Reference]	Reference	Type of Administrative Data	Definition/ICD Codes Used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Kappa
Canada	1995 - 1996	Wilcheky[23]	Medical Chart	Physician Claims	Using only diagnoses recorded in the claims of study physicians *[ICD-9 250.0 - 250.9]*	2,752	51.78 (49.9, 53.6)	**98.41 (98.2, 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 - 250.9]*		**64.43 (62.6, 66.2)**	96.82 (96.5, 97.1)			

USA	1997 - 2001	Crane[14]	Clinician documentation in Electronic Medical Record progress notes	Physician Claims	At least one clinician-coded diagnoses *[ICD 9 250.0, 250.1, 250.2, 250.3]*	1,441	93 (86, 100)	99 (99, 100)	91 (83, 99)		
USA	1998 - 1999	Borzecki[13]	Medical Charts	Physician Claims	At least one diagnosis in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over one year *[ICD 9 250.x]*	1,176	97	96			0.92
					At least two diagnoses in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over one year *[ICD 9 250.x]*						0.91
					At least one diagnosis in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over two years *[ICD 9 250.x]*						0.89
					At least two diagnoses in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over two years *[ICD 9 250.x]*						**0.93**
USA	1992 - 1995	Hebert[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 - 250.93, 357.2, 362.0 -*		71.6	**96.6**	**79**		

					362.02, 366.41]						
					One or more diagnoses of diabetes in any claim file over 2-year period *[ICD 9-CM 250.00 - 250.93, 357.2, 362.0 - 362.02, 366.41]*		**79.1**	94.3	71.4		
USA	1993 - 1994	O'Connor[20]	Telephone Survey	Physician Claims	Two or more ICD-9 diagnostic codes *[ICD 9 250.x]*	1,976	92.22*	98.62*	76.15*	99.63*	
USA	1996 - 1998	Singh[19]	Self-reported Survey	Physician Claims	Veterans Affairs databases *[ICD 9 250]*		76 (75 - 76)	98 (98 - 98)	91 (91 - 91)	95 (94 - 95)	0.79 (0.79 - 0.80)
USA	1997	Ngo[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 - 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 - 0.85)
USA	1997 - 2000	Miller[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 out-patient 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 out-patient 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 out-patient 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 out-patient codes *[ICD 9 250, 357.2, 362.0, 366.41]*		63.8	99.2	**96.2**		

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification

Note: *Sensitivity, Specificity, PPV and NPV are all hand-calculated.

Sensitivity identifies the proportion of individuals who truly do have the disease/condition.
Specificity identifies the proportion of individuals who truly do not have the disease/condition.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

Table 3: Study Characteristics and Test Measures of Studies for Hospital Discharge Data
(Superior performance characteristics within studies have been highlighted in bold.)

Country	Study Year	Author[Reference]	Reference	Type of Administrative Data	Definition/ICD Codes Used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Kappa
Canada	1995 - 2000	So[24]	Medical Chart	Hospital Discharge Data	Diabetes with Complications *[ICD-9 250.1 -250.9]*	93	**80** (51.91, 95.67)	98.3 (95.15, 99.65)	80 (51.91, 95.67)	**98.3 (95.15, 99.65)**	
	2001 - 2004				Diabetes with Complications *[ICD-10 E10.0 -E10.8, E11.0 - E11.8, E12.0 -E12.8, E13.0 - E13.8, E14.0 -E14.8]*		66.7 (38.38, 88.18)	**98.9 (96.00, 99.86)**	**83.3** (51.59, 97.91)	97.2 (93.67, 99.10)	
Canada	2003	Quan[26]	Medical Chart	Hospital Discharge Data	Diabetes with Chronic Complications *[ICD 9 250.4 -250.7]*	4,008	**63.6**	98.9	62.5	**99**	**0.62**

					Diabetes with Chronic Complications *[ICD 10 E10.2 - E10.5, E10.7, E11.2 - E11.5, E11.7, E12.2 - E12.5, E12.7, E13.2 - E13.5, E13.7, E14.2 - E14.5, E14.7]*		59.1	**99**	**63.1**	98.9	0.6
					Diabetes without Chronic Complications *[ICD 9 250.0 -250.3, 250.8, 250.9]*		**77.7**	98.4	86.5	**97**	**0.8**
					Diabetes without Chronic Complications *[E10.0, E10.1, E10.6, E10.8, E10.9, E110, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9]*		75.8	**98.7**	**88.5**	96.8	0.79
Western Australia	1998	Nedkoff[28]	Medical Chart	Hospital Discharge Data	Look back period: Index admission *[ICD 9/ICD-9 CM 250]*	1,685	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
	2002–04				Look back period: Index admission *[ICD 10-AM E10-E14]*	2,258	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**
					10-years		89.6	95.6 †	91.6	**94.5**	0.856

					15-years		**89.6**	95.5 †	91.5	**94.5**	0.855
Canada	1989 - 1990	Robinson[22]	Self-reported Survey	Hospital Discharge Data and Physician Claims	1, 2 or 3 physician claim or 1 hospitalization over 3 years *[ICD 9 CM]*	2,651	72	98	76	98	0.72 (0.67 - 0.77)

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification

Note: *Sensitivity, Specificity, PPV and NPV are all hand-calculated.

Sensitivity identifies the proportion of individuals who truly do have the disease/condition.
Specificity identifies the proportion of individuals who truly do not have the disease/condition.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

Table 4: Study Characteristics and Test Measures of Studies for both Physician Claims Data and Hospital Discharge Data
(Superior performance characteristics within studies have been highlighted in bold.)

Country	Study Year	Author[Reference]	Reference	Type of Administrative Data	Definition/ICD Codes Used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Kappa
Canada	1992 - 1999	Hux[21]	Medical Chart	Physician Claims and Hospital Discharge Data	One Physician Service Claims or One Hospitalization with diagnosis of diabetes *[ICD-9 250.x]*	3,317	**91**	92*	61	**99***	
					Two Physician Service Claims or One Hospitalization with diagnosis of diabetes *[ICD-9 250.x]*		86	**97***	**80**	98*	
Canada	2000 - 2002	Chen[25]	Medical Chart	Physician Claims and Hospital Discharge Data	3 Years Observation Period Data *[ICD 9 250.xx, ICD 10 E10.x–E14.x]*	3,362	**95.6 (92.5–97.7)**	92.8 (91.9–93.7)	54 (49.6–58.5)	**99.6 (99.4–99.8)**	0.65 (0.61–0.69)
					2 Years Observation Period Data *[ICD 9 250.xx, ICD 10 E10.x–E14.x]*		86.4 (82.4–90.5)	**97.1 (96.5–97.7)**	**72.4 (67.5–77.3)**	98.8 (98.4–99.2)	**0.77 (0.73–0.81)**

				Physician Claims	3 Years Observation Period Data *[ICD 9 250.xx, ICD 10 E10.x–E14.x]*		**91.2 (87.9–94.6)**	97.6 (97.1–98.1)	72.1 (67.5–76.9)	**99.2 (98.9–99.5)**	**0.82 (0.78–0.85)**
					2 Years Observation Period Data *[ICD 9 250.xx, ICD 10 E10.x–E14.x]*		76.6 (71.5–81.6)	**99.3 (99.0–99.6)**	**90.9 (87.2–94.6)**	98 (97.5–98.4)	**0.82 (78.0–85.5)**
USA	1999	Rector[17]	Telephone surveys	Hospital Discharge Data and Physician Claims	One 1999 claim with dx *[ICD 9 250.xx, 357.2x, 362.0x, 366.41]*	3,633	**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			

					Two 1999 face-to-face encounter claims with primary dx		57	**99**			
	1999 - 2000				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**			

Canada	1980 - 1984	Young[27]	Self-reported Survey	Hospital Admission and Physician Claims	[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC)] AND [Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment] *[ICD 9-CM]*	1,000	82.7	96.3			
					[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment] AND [Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment] *[ICD 9-CM]*		82.1	**98.5**			
					[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC)] AND [Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment] AND [Claims by the physician to the Manitoba Health Services Commission		**83.9**	95.8			

					(MHSC) or payment] *[ICD 9-CM]*						

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification

Note: *Sensitivity, Specificity, PPV and NPV are all hand-calculated.

Sensitivity identifies the proportion of individuals who truly do have the disease/condition.
Specificity identifies the proportion of individuals who truly do not have the disease/condition.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

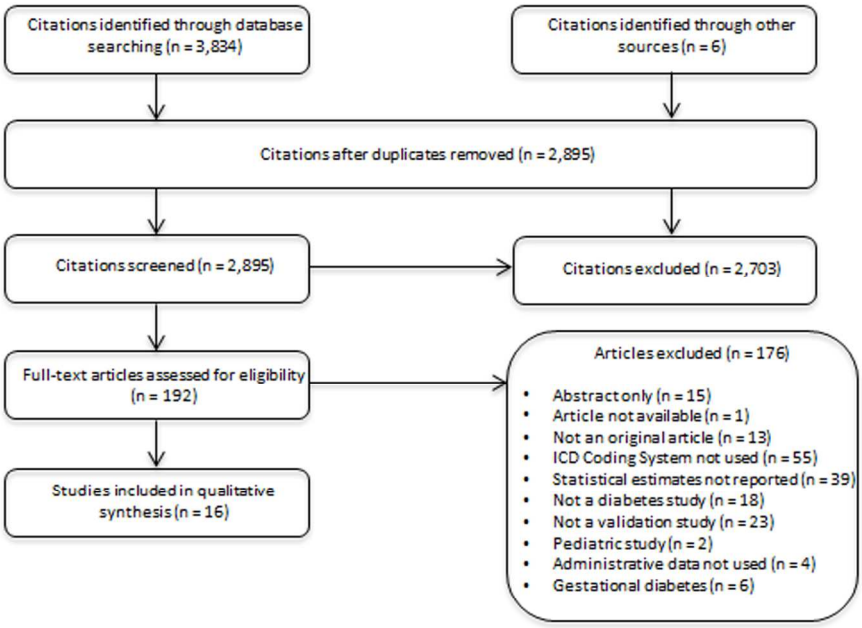


Figure 1. Study Flow

186x139mm (300 x 300 DPI)

Appendix A: Search Strategies

Embase Search Criteria

Wolters Kluwer Health | OvidSP

Logged in as Bushra Khokhar at University of Calgary

My Account | Ask a Librarian | Support & Training | Help | Logoff

Search | Journals | Books | Multimedia | My Workspace | Primal Pictures

Search History (4 searches) (close) — View Saved

	# ▲	Searches	Results	Search Type	Actions
☐	1	(exp health services research/ or administrative data.mp. or hospital discharge data.mp. or icd9.mp. or icd10.mp. or icd-9.mp. or icd-10.mp. or exp international classification of diseases/ or medical record*.mp. or health information.mp. or surveillance.mp. or physician claims.mp. or claims.mp. or hospital discharge.mp. or coding.mp. or codes.mp.) and (validity or validation or case definition or algorithm or agreement or accuracy or sensitivity or specificity or positive predictive value or negative predictive value).mp.	80154	Advanced	Display, Delete, Save, Auto-Alert, RSS Feed
☐	2	exp diabetes mellitus/ or diabetes.mp.	659483	Advanced	Display, More »
☐	3	1 and 2	2791	Advanced	Display, More »
☐	4	limit 3 to english language	2678	Advanced	Display, More »

Remove Selected | Save Selected | Combine selections with: And Or | RSS

Save Search History

Medline Search Criteria

Wolters Kluwer Health | OvidSP

My Account | Ask a Librarian | Support & Training | Help | Logoff

Search | Journals | Books | Multimedia | My Workspace | Primal Pictures

Search History (4 searches) (close) — View Saved

	# ▲	Searches	Results	Search Type	Actions
☐	1	(exp health services research/ or administrative data.mp. or hospital discharge data.mp. or icd9.mp. or icd10.mp. or icd-9.mp. or icd-10.mp. or exp international classification of diseases/ or medical record*.mp. or health information.mp. or surveillance.mp. or physician claims.mp. or claims.mp. or hospital discharge.mp. or coding.mp. or codes.mp.) and (validity or validation or case definition or algorithm or agreement or accuracy or sensitivity or specificity or positive predictive value or negative predictive value).mp.	58165	Advanced	Display, Delete, Save, Auto-Alert, RSS Feed
☐	2	exp diabetes mellitus/ or diabetes.mp.	420560	Advanced	Display, More »
☐	3	1 and 2	1219	Advanced	Display, More »
☐	4	limit 3 to english language	1156	Advanced	Display, More »

Remove Selected | Save Selected | Combine selections with: And Or | RSS

Save Search History

Appendix A: PRISMA Checklist 2009

Section	#	Checklist Item	Reported on Page Number
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 3
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix A
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Page 4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g. I^2) for each meta-analysis.	N/A
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Tables 1 - 4
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Pages 6,7

Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Page 7
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Page 7
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Page 8

PRISMA 2009 Checklist was extracted from Liberati A. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Ann. Intern. Med. 2009;339:b2700[11].

BMJ Open

Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-009952.R2
Article Type:	Research
Date Submitted by the Author:	31-Mar-2016
Complete List of Authors:	Khokhar, Bushra; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health Jette, Nathalie; University of Calgary, O'Brien Institute for Public Health; University of Calgary, Clinical Neurosciences Metcalfe, Amy; University of Calgary, Department of Obstetrics and Gynecology; Alberta Children's Hospital Research Institute, Cunningham, Ceara Tess; University of Calgary, Community Health Sciences Quan, Hude; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health Kaplan, Gilaad; University of Calgary; University of Calgary, O'Brien Institute for Public Health Butalia, Sonia; University of Calgary, Community Health Sciences Rabi, Doreen; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health informatics, Epidemiology, Diabetes and endocrinology, Health services research, Diagnostics
Keywords:	diabetes, validation studies, case definition, administrative data

Title: Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data

Corresponding Author: Bushra Khokhar
Postal Address: 3280 Hospital Drive NW, 3rd floor TRW Calgary, Alberta T2N 4N1, Canada
E-mail: bushra.khokhar@ucalgary.ca
Telephone Number: (403) 210-3807
Author List: 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], and Doreen Rabi[1,2,6]

*Corresponding author
[1]Department of Community Health Sciences, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[2]O'Brien Institute for Public Health, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[3]Department of Clinical Neurosciences & Hotchkiss Brain Institute, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada
[4]Department of Obstetrics and Gynecology, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada
[5]Alberta Children's Hospital Research Institute, 3330 Hospital Drive NW, Calgary, Alberta T2N 4NI, Canada
[6] Division of Endocrinology, Department of Medicine, 1820 Richmond Road SW, Calgary, Alberta TCC CCC

Keywords: diabetes, validation studies, case definition, and administrative data

Word Count (excluding title page, abstract, references, figures and tables): 3,052words

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.

Methods: Electronic databases were searched for validation studies where an administrative case definition was validated.

Results: The search yielded 2,895 abstracts and of the 193 potentially relevant studies, 16 met criteria. 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 kappa ranged from 0.8 to 0.9 in studies using physician claims 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 kappa ranged from 0.6 to 0.9 in studies using hospital discharge data. 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 kappa ranged from 0.7 to 0.8 in studies using a combination of both.

Conclusion: Performance characteristics of case definitions depends on the variations in the definition of primary diagnosis in ICD-coded data, the use of fee-for service payment model or salary based model for physicians, and/or the methodology adopted by the health care facility to extract information from patient records. The purpose of surveillance and type of data being used should command the performance parameters of a case definition. Approaches used in developing case definitions can be pragmatic and result in high sensitivity, specificity and PPV. Overall, administrative databases are useful for undertaking diabetes surveillance but an awareness of the variation in performance being affected by case definition is essential.

STRENGTHS AND LIMITATIONS

- The greatest strength of this systematic review is its inclusiveness - the search strategy was not restricted by region, time or any particular diabetes case definition of diabetes.
- Most of the studies, 15 out 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.
- Lastly, 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 definitions for diabetes. The accuracy of charts reviews depends principally on physician documentation, availability of records, and the accuracy of coding[34]. 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 patient's level of education can have an effect on the reporting of diabetes[35-37]. Those with poorly controlled diabetes have been found to underreport their disease status[38].

- Lastly, difference between type 1 diabetes mellitus and type 2 diabetes mellitus is not clear in studies using administrative databases. In this review we included only those ≥ 18 years of age that is primarily the type 2 diabetes mellitus population.

BACKGROUND

Diabetes is a chronic disease that has increased substantially during the past 20 years[1]. At present, diabetes is the leading cause of blindness[2], renal failure[3], non-traumatic lower limb amputations[4], and is a major risk factor for cardiovascular disease[5]. Because of 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 health care 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 analyze routinely collected health-related information from healthcare systems and provide reports on risk factors, care practices, morbidity, 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) 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) guidelines[11][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 [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 (e.g. ICD-9-CM).

Study Selection

Studies were evaluated in duplicate for eligibility in a two-stage procedure. In stage one, all identified titles and abstracts were reviewed and in stage two, 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 one, it was included in the full-text review in stage two. 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, kappa] 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 (e.g. self-report from population-based surveys or patient medical chart reviews); and (5) if it reported on original data. Studies validating diabetes in specialized populations (e.g. cardiovascular disease) were excluded to ensure the diabetes case definitions would be generalizable. Studies not employing a sole medical encounter data in their diabetes case definition (e.g. 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 kappa 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 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 2,895 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 United States[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 both 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, eight used medical records[13-14, 21, 23-26, 28], and eight 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 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, out 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 kappa in six. All 16 studies were categorized 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 kappa ranged from 0.8 to 0.9. Four out of the eight studies using physician claims data had a least one diabetes case definition where both 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 over time. This relationship is consistent with the study[18] 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 (in-patient or out-patient) was used, while the specificity and PPV were the highest when most number of out-patient 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 kappa ranged from 0.6 to 0.9. Two out of the four studies using hospital discharge data had a least one diabetes case definition where both 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 kappa 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 both sensitivity and specificity exceed 80%.

Another factor affecting the statistical estimates is the number of claims being used in the definition. Rector et al.'s study[17] shows consistent results where the sensitivity is higher when at least one claims data is used in the definition, but the specificity is higher when at least two are used. Lastly, Young et al.'s study[27] demonstrates the highest sensitivity when 2 physician claims and 2 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 United States and four studies from Canada. Four studies use both 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 kappa 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 kappa ranged from 0.6 to 0.9.

DISCUSSION

In this systematic review, case definitions appear to preform ~~to perform more reliably~~ 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 one 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 (NDSS) 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, 87.4) and specificity of 97.9% (95% CI 96.5, 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 standardized manner. Billing practices do vary by practitioner, which may in turn be influenced by the nature of physician reimbursement (salary versus 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 hospitalization, 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 standardized 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 hospitalization, within a certain period of time is appropriate. It is also important to recognize 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. Most of the studies, 15 out 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. Lastly, 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 charts 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 done.

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 over time 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 biologic or physiologic definition that denotes the presence or absence of disease, the administrative data definitions are a surrogate of disease, and denote presence or 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 health care 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.

Lastly, 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.

Generalizability

Fifteen out of the 16 included studies ~~of 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.

CONCLUSION

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, the use of fee-for service payment model or salary based model for physicians and primary care providers, and/or the methodology adopted by the health care 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.

COMPETING INTERESTS

The authors declare that they have no competing interest.

AUTHOR'S CONTRIBUTIONS

Dr. Nathalie Jette wrote the protocol. Ms. Bushra Khokhar, Dr. Amy Metcalfe, and Dr. Ceara Tess Cunningham carried out the systematic review. Bushra Khokhar wrote the manuscript. Dr. Nathalie Jette, Dr. Hude Quan, Dr. Gilaad G. Kaplan, Dr. Sonia Butalia, and Dr. Doreen Rabi provided final approval of the version to be published. All authors read and approved the final manuscript.

FUNDING SOURCES

Ms. Bushra Khokhar was supported by the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD) and The Western Regional Training Centre for Health Services Research (WRTC). Dr. Nathalie Jette 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. Dr. Ceara Tess Cunningham is funded by a Canadian Institute of Health Research doctoral research scholarship. Dr. Kaplan is a Population Health Investigator supported by Alberta Innovates - Health Solutions. Dr. Doreen Rabi is a Population Health Investigator supported by Alberta Innovates - Health Solutions.

FIGURE LEGEND

Figure 1. Study Flow Chart.

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, and Dillon J. Early Diagnosis of Diabetes Through The Eye. *Photochem Photobiol* Published Online First 27 August 2015. doi: 10.1111/php.12524.
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 Sep 7;100:1134-46.
6) World Health Organization. Diabetes: The Cost of Diabetes. [http://www.who.int/mediacentre/factsheets/fs236/en/] Accessed August 27, 2014.
7) Jutte DR, Roos LL, and 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: 1792-99.
9) Centers for Disease Control and Prevention: National Notifiable Diseases Surveillance System (NNDSS). [http://wwwn.cdc.gov/nndss/case-definitions.html]. Accessed September 1, 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. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Ann. Intern. Med. 2009;339:b2700.
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, and 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:B10-21.
19) Singh JA. Accuracy of Veterans Affairs Databases for Diagnoses of Chronic Diseases. *Preventative Chronic Disease*. 2009;6:A126.
20) O'Connor PJ, Rush W A, 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;25:512-6.
22) Robinson JR, Young TK, Roos LL, et al. Estimating the burden of disease. Comparing administrative data and self-reports. *Medical Care*. 1997;35:932-47.
23) Wilchesky M, Tamblyn RM, and Huang A. Validation of Diagnostic Codes within Medical Services Claims. *J Clin epidemiol.* 2004;57:131-41.
24) So L, Evans D, and Quan H. ICD-10 coding algorithms for defining comorbidities of acute myocardial infarction. *BMC Health Serv Res*. 2006;6.
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;4: 1424-41.
27) Young TK, Roos NP, and 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;75:313-9.
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: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 August 28, 2014.
32) Roos LL, Roos NP, Cageorge SM, Nicol P. How good are the data? Reliability of one health care data bank. Med Care 1982;20(3):266–76.
33) Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. J Clin Epidemiol 2000;53(12):1258–67.
34) Carral F, Olveira 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;33:593–598.
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. *J Clin Epidemiol*. 1996;49:1407- 17.
39) Mackenbach JP, Looman CW, and van deer 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, Gutierrez-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.

Table 1. Study Quality Characteristics using QUADAS Tool

QUADAS Tool Item	Hux[21]	Robinson[2]	Borzecki[1]	Wilchesky[2]	Crane[1]	So[24]	Chen[25]	Nedkoff[2]	Quan[26]	Young[2]	Hebert[1]	Ngo[1]	Rector[1]	Miller[1]	Singh[19]	O'Connor[2]
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	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

verification using a reference standard of diagnosis?*																
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 (i.e. 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
Were the index test results interpreted without knowledge of	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

the results of the reference standard?*																
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 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[12].

Table 2: Study Characteristics and Test Measures of Studies for Physician Claims Data
(Superior performance characteristics within studies have been highlighted in bold.)

Country	Study Year	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)	Kappa
Canada	1995 - 1996	Wilcheky[23]	Medical Chart	Physician Claims	Using only diagnoses recorded in the claims of study physicians	ICD-9 250.0 - 250.9	2,752	51.78 (49.9, 53.6)	**98.41 (98.2, 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 - 250.9		**64.43 (62.6, 66.2)**	96.82 (96.5, 97.1)			
USA	1997 - 2001	Crane[14]	Clinician documentation in Electronic Medical Record progress notes	Physician Claims	At least one clinician-coded diagnoses	ICD 9 250.0, 250.1, 250.2, 250.3	1,441	93 (86, 100)	99 (99, 100)	91 (83, 99)		
USA	1998 - 1999	Borzecki[13]	Medical Charts	Physician Claims	At least one diagnosis in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over one year	ICD 9 250.x	1,176	97	96			0.92
					At least two diagnoses in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over one year	ICD 9 250.x						0.91
					At least one diagnosis in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over two years	ICD 9 250.x						0.89

					At least two diagnoses in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over two years	ICD 9 250.x						**0.93**
USA	1992 - 1995	Hebert[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 - 250.93, 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 - 250.93, 357.2, 362.0 - 362.02, 366.41		**79.1**	94.3	71.4		
USA	1993 - 1994	O'Connor[20]	Telephone Survey	Physician Claims	Two or more ICD-9 diagnostic codes	ICD 9 250.x	1,976	92.22*	98.62*	76.15*	99.63*	
USA	1996 - 1998	Singh[19]	Self-reported Survey	Physician Claims	Veterans Affairs databases	ICD 9 250		76 (75 - 76)	98 (98 - 98)	91 (91 - 91)	95 (94 - 95)	0.79 (0.79 - 0.80)
USA	1997	Ngo[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 - 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 - 0.85)
USA	1997 - 2000	Miller[18]	Self-reported Survey	Physician Claims (Medicare)	Any diagnostic code	ICD 9 250, 357.2, 362.0,	2,924,148	**78.3**	95.7	85.3		

						366.41						
					Any out-patient 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 out-patient 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 out-patient 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 out-patient codes	ICD 9 250, 357.2, 362.0, 366.41		63.8	**99.2**	**96.2**		

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
Note: *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.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

Table 3: Study Characteristics and Test Measures of Studies for Hospital Discharge Data

(Superior performance characteristics within studies have been highlighted in bold.)

Country	Study Year	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)	Kappa
Canada	1995 - 2000	So[24]	Medical Chart	Hospital Discharge Data	Diabetes with Complications	ICD-9 250.1 - 250.9	93	**80 (51.91, 95.67)**	98.3 (95.15, 99.65)	80 (51.91, 95.67)	**98.3 (95.15, 99.65)**	
	2001 - 2004				Diabetes with Complications	ICD-10 E10.0 - E10.8, E11.0 - E11.8, E12.0 - E12.8, E13.0 - E13.8, E14.0 - E14.8		66.7 (38.38, 88.18)	**98.9 (96.00, 99.86)**	**83.3 (51.59, 97.91)**	97.2 (93.67, 99.10)	
Canada	2003	Quan[26]	Medical Chart	Hospital Discharge Data	Diabetes with Chronic Complications	ICD 9 250.4 - 250.7	4,008	**63.6**	98.9	62.5	**99**	**0.62**
					Diabetes with Chronic Complications	ICD 10 E10.2 - E10.5, E10.7, E11.2 - E11.5, E11.7, E12.2 - E12.5, E12.7, E13.2 - E13.5, E13.7, E14.2 - E14.5, E14.7		59.1	**99**	**63.1**	98.9	0.6

					Diabetes without Chronic Complications	ICD 9 250.0 - 250.3, 250.8, 250.9		**77.7**	98.4	86.5	**97**	**0.8**
					Diabetes without Chronic Complications	E10.0, E10.1, E10.6, E10.8, E10.9, E110, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9		75.8	**98.7**	**88.5**	96.8	0.79
Western Australia	1998	Nedkoff[28]	Medical Chart	Hospital Discharge Data	Look back period: Index admission	ICD 9/ICD-9 CM 250	1,685	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	2,258	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**
					10-years			89.6	95.6 †	91.6	**94.5**	0.856
					15-years			**89.6**	95.5 †	91.5	**94.5**	0.855
Canada	1989 - 1990	Robinson[22]	Self-reported Survey	Hospital Discharge Data and Physician Claims	1, 2 or 3 physician claim or 1 hospitalization over 3 years	ICD 9 CM	2,651	72	98	76	98	0.72 (0.67 - 0.77)

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
Note: *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.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

Table 4: Study Characteristics and Test Measures of Studies for both Physician Claims Data and Hospital Discharge Data
(Superior performance characteristics within studies have been highlighted in bold.)

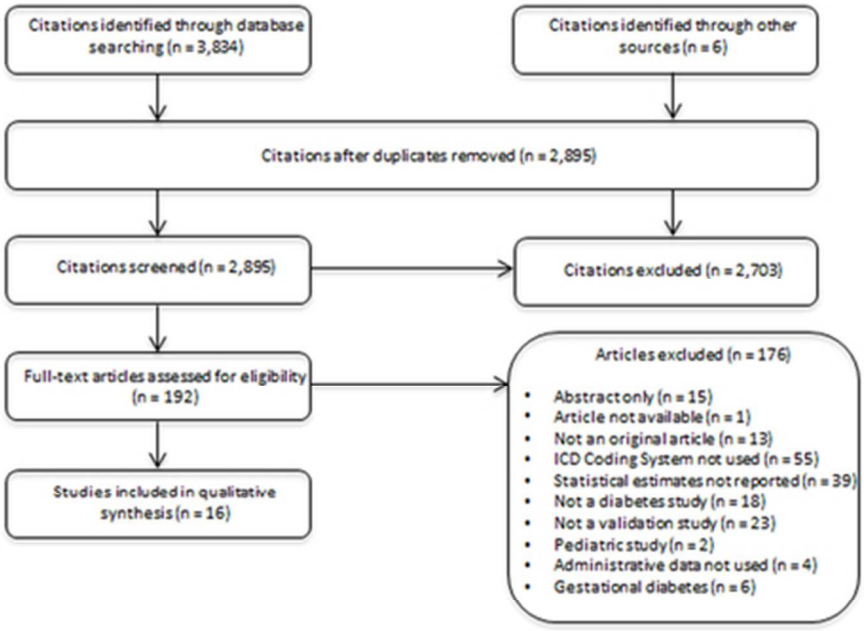
Country	Study Year	Author[Reference]	Reference	Type of Administrative Data	Diabetes Case Definition	ICD Codes Used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Kappa
Canada	1992 - 1999	Hux[21]	Medical Chart	Physician Claims and Hospital Discharge Data	One Physician Service Claims or One Hospitalization with diagnosis of diabetes	ICD-9 250.x	3,317	**91**	92*	61	**99***	
					Two Physician Service Claims or One Hospitalization with diagnosis of diabetes	ICD-9 250.x		86	**97***	**80**	98*	
Canada	2000 - 2002	Chen[25]	Medical Chart	Physician Claims and Hospital Discharge Data	3 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x–E14.x	3,362	**95.6 (92.5–97.7)**	92.8(91.9–93.7)	54 (49.6–58.5)	**99.6 (99.4–99.8)**	0.65 (0.61 – 0.69)
					2 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x–E14.x		86.4 (82.4–90.5)	**97.1 (96.5–97.7)**	**72.4 (67.5–77.3)**	98.8 (98.4–99.2)	**0.77 (0.73 – 0.81)**
				Physician Claims	3 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x–E14.x		**91.2 (87.9–94.6)**	97.6 (97.1–98.1)	72.1 (67.5–76.9)	**99.2 (98.9–99.5)**	**0.82 (0.78 – 0.85)**
					2 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x–E14.x		76.6 (71.5–81.6)	**99.3 (99.0–99.6)**	**90.9 (87.2–94.6)**	98 (97.5–98.4)	**0.82 (78.0 – 85.5)**
USA	1999	Rector[17]	Telephone surveys	Hospital Discharge Data and Physician Claims	One 1999 claim with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	3,633	**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			
					Two 1999 face-to-face encounter claims with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41		57	**99**			
	1999 - 2000				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,		91	95			

						362.0x, 366.41						
					Two 1999 or 2000 face-to-face encounter claims with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41		77	**98**			
Canada	1980 - 1984	Young[27]	Self-reported Survey	Hospital Admission and Physician Claims	[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC)] AND [Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment]	ICD 9-CM	1,000	82.7	96.3			
					[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment] AND [Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment]	ICD 9-CM		82.1	**98.5**			

					[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC)] AND [Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment] AND [Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment]	ICD 9-CM		**83.9**	95.8			

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
Note: *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.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.



186x139mm (72 x 72 DPI)

Appendix A: PRISMA Checklist 2009

Section	#	Checklist Item	Reported on Page Number
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 3
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix A
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Page 4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g. I^2) for each meta-analysis.	N/A
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Tables 1 - 4
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Pages 6,7

Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Page 7
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Page 7
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Page 8

PRISMA 2009 Checklist was extracted from Liberati A. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventio Explanation and Elaboration. Ann. Intern. Med. 2009;339:b2700[11].

Appendix B: Search Strategies

Embase Search Criteria

Medline Search Criteria

BMJ Open

Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data in Adult Population.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-009952.R3
Article Type:	Research
Date Submitted by the Author:	25-May-2016
Complete List of Authors:	Khokhar, Bushra; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health Jette, Nathalie; University of Calgary, O'Brien Institute for Public Health; University of Calgary, Clinical Neurosciences Metcalfe, Amy; University of Calgary, Department of Obstetrics and Gynecology; Alberta Children's Hospital Research Institute, Cunningham, Ceara Tess; University of Calgary, Community Health Sciences Quan, Hude; University of Calgary, O'Brien Institute for Public Health; University of Calgary, Community Health Sciences Kaplan, Gilaad; University of Calgary; University of Calgary, O'Brien Institute for Public Health Butalia, Sonia; University of Calgary, Community Health Sciences Rabi, Doreen; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health informatics, Epidemiology, Diabetes and endocrinology, Health services research, Diagnostics
Keywords:	diabetes, validation studies, case definition, administrative data

Title: Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data in Adult Population.

Corresponding Author: Bushra Khokhar
Postal Address: 3280 Hospital Drive NW, 3rd floor TRW Calgary, Alberta T2N 4N1, Canada
E-mail: bushra.khokhar@ucalgary.ca
Telephone Number: (403) 210-3807
Author List: 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], and Doreen Rabi[1,2,6]

*Corresponding author
[1]Department of Community Health Sciences, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[2]O'Brien Institute for Public Health, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[3]Department of Clinical Neurosciences & Hotchkiss Brain Institute, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada
[4]Department of Obstetrics and Gynecology, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada
[5]Alberta Children's Hospital Research Institute, 3330 Hospital Drive NW, Calgary, Alberta T2N 4NI, Canada
[6] Division of Endocrinology, Department of Medicine, 1820 Richmond Road SW, Calgary, Alberta TCC CCC

Keywords: diabetes, validation studies, case definition, and administrative data

Word Count (excluding title page, abstract, references, figures and tables): 3,052words

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.

Methods: Electronic databases were searched for validation studies where an administrative case definition was validated.

Results: The search yielded 2,895 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%, PPV ranged from 71.4 to 96.2%, NPV ranged from 95 to 99.6% and kappa 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 kappa 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 kappa ranged from 0.7 to 0.8).

Conclusion: 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 depends on the variations in the definition of primary diagnosis in ICD-coded discharge data, and/or the methodology adopted by the health care facility to extract information from patient records.

STRENGTHS AND LIMITATIONS

- 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.
- Most of the studies, 15 out 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.
- Lastly, 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 definitions for diabetes. The accuracy of charts reviews depends principally on physician documentation, availability of records, and the accuracy of coding[34]. 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 patient's level of education can have an effect on the reporting of diabetes[35-37]. Those with poorly controlled diabetes have been found to underreport

their disease status[38].

- Lastly, difference between type 1 diabetes mellitus and type 2 diabetes mellitus is not clear in studies using administrative databases. In this review we included only those ≥ 18 years of age that is primarily the type 2 diabetes mellitus population.

BACKGROUND

Diabetes is a chronic disease that has increased substantially during the past 20 years[1]. At present, diabetes is the leading cause of blindness[2], renal failure[3], non-traumatic lower limb amputations[4], and is a major risk factor for cardiovascular disease[5]. Because of 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 health care 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 analyze routinely collected health-related information from healthcare systems and provide reports on risk factors, care practices, morbidity, 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) 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) guidelines[11][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 [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 (e.g. ICD-9-CM).

Study Selection

Studies were evaluated in duplicate for eligibility in a two-stage procedure. In stage one, all identified titles and abstracts were reviewed and in stage two, 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 one, it was included in the full-text review in stage two. 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, kappa] 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 (e.g. self-report from population-based surveys or patient medical chart reviews); and (5) if it reported on original data. Studies validating diabetes in specialized populations (e.g. cardiovascular disease) were excluded to ensure the diabetes case definitions would be generalizable. Studies not employing a sole medical encounter data in their diabetes case definition (e.g. 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 kappa 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 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 2,895 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 United States[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 both 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, eight used medical records[13-14, 21, 23-26, 28], and eight 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 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, out 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 kappa in six. All 16 studies were categorized 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 kappa ranged from 0.8 to 0.9. Four out of the eight studies using physician claims data had a least one diabetes case definition where both 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 over time. This relationship is consistent with the study[18] 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 (in-patient or out-patient) was used, while the specificity and PPV were the highest when most number of out-patient 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 kappa ranged from 0.6 to 0.9. Two out of the four studies using hospital discharge data had a least one diabetes case definition where both 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 kappa 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 both sensitivity and specificity exceed 80%.

Another factor affecting the statistical estimates is the number of claims being used in the definition. Rector et al.'s study[17] shows consistent results where the sensitivity is higher when at least one claims data is used in the definition, but the specificity is higher when at least two are used. Lastly, Young et al.'s study[27] demonstrates the highest sensitivity when 2 physician claims and 2 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 United States and four studies from Canada. Four studies use both 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 kappa 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 kappa ranged from 0.6 to 0.9.

DISCUSSION

In this systematic review, case definitions appear to preform ~~to perform more reliably~~ 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 one 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 (NDSS) 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, 87.4) and specificity of 97.9% (95% CI 96.5, 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 standardized manner. Billing practices do vary by practitioner, which may in turn be influenced by the nature of physician reimbursement (salary versus 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 hospitalization, 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 standardized 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 hospitalization, within a certain period of time is appropriate. It is also important to recognize 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 out 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. Lastly, 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 charts 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 done.

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 over time 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 biologic or physiologic definition that denotes the presence or absence of disease, the administrative data definitions are a surrogate of disease, and denote presence or 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 health care 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.

Lastly, 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.

Generalizability

Fifteen out 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.

CONCLUSION

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 health care 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.

COMPETING INTERESTS

The authors declare that they have no competing interest.

AUTHOR'S CONTRIBUTIONS

Dr. Nathalie Jette wrote the protocol. Ms. Bushra Khokhar, Dr. Amy Metcalfe, and Dr. Ceara Tess Cunningham carried out the systematic review. Bushra Khokhar wrote the manuscript. Dr. Nathalie Jette, Dr. Hude Quan, Dr. Gilaad G. Kaplan, Dr. Sonia Butalia, and Dr. Doreen Rabi provided final approval of the version to be published. All authors read and approved the final manuscript.

FUNDING SOURCES

Ms. Bushra Khokhar was supported by the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD) and The Western Regional Training Centre for Health Services Research (WRTC). Dr. Nathalie Jette 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. Dr. Ceara Tess Cunningham is funded by a Canadian Institute of Health Research doctoral research scholarship. Dr. Kaplan is a Population Health Investigator supported by Alberta Innovates - Health Solutions. Dr. Doreen Rabi is a Population Health Investigator supported by Alberta Innovates - Health Solutions.

DATA SHARING

Any additional data, such as study protocol, data extraction forms, etc. are available by emailing the first author at bushra.khokhar@ucalgary.ca

FIGURE LEGEND

Figure 1. Study Flow Chart.

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, and Dillon J. Early Diagnosis of Diabetes Through The Eye. *Photochem Photobiol* Published Online First 27 August 2015. doi: 10.1111/php.12524.
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 Sep 7;100:1134-46.
6) World Health Organization. Diabetes: The Cost of Diabetes. [http://www.who.int/mediacentre/factsheets/fs236/en/] Accessed August 27, 2014.
7) Jutte DR, Roos LL, and 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: 1792-99.
9) Centers for Disease Control and Prevention: National Notifiable Diseases Surveillance System (NNDSS). [http://wwwn.cdc.gov/nndss/case-definitions.html]. Accessed September 1, 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. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Ann. Intern. Med. 2009;339:b2700.
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, and 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:B10-21.
19) Singh JA. Accuracy of Veterans Affairs Databases for Diagnoses of Chronic Diseases. *Preventative Chronic Disease*. 2009;6:A126.
20) O'Connor PJ, Rush W A, 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;25:512-6.
22) Robinson JR, Young TK, Roos LL, et al. Estimating the burden of disease. Comparing administrative data and self-reports. *Medical Care*. 1997;35:932-47.
23) Wilchesky M, Tamblyn RM, and Huang A. Validation of Diagnostic Codes within Medical Services Claims. *J Clin epidemiol.* 2004;57:131-41.
24) So L, Evans D, and Quan H. ICD-10 coding algorithms for defining comorbidities of acute myocardial infarction. *BMC Health Serv Res*. 2006;6.
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;4: 1424-41.
27) Young TK, Roos NP, and 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;75:313-9.
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: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 August 28, 2014.
32) Roos LL, Roos NP, Cageorge SM, Nicol P. How good are the data? Reliability of one health care data bank. Med Care 1982;20(3):266–76.
33) Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. J Clin Epidemiol 2000;53(12):1258–67.
34) Carral F, Olveira 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;33:593–598.
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. *J Clin Epidemiol*. 1996;49:1407- 17.
39) Mackenbach JP, Looman CW, and van deer 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, Gutierrez-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.

Table 1. Study Quality Characteristics using QUADAS Tool

QUADAS Tool Item	Hux[21]	Robinson[2]	Borzecki[1]	Wilchesky[2]	Crane[1]	So[24]	Chen[25]	Nedkoff[2]	Quan[26]	Young[2]	Hebert[1]	Ngo[1]	Rector[1]	Miller[1]	Singh[19]	O'Connor[2]
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	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

verification using a reference standard of diagnosis?*																
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 (i.e. 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
Were the index test results interpreted without knowledge of	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

the results of the reference standard?*																
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 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[12].

Table 2: Study Characteristics and Test Measures of Studies for Physician Claims Data
(Superior performance characteristics within studies have been highlighted in bold.)

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)	Kappa
Canada	1995 - 1996	Wilcheky[23]	Medical Chart	Physician Claims	Using only diagnoses recorded in the claims of study physicians	ICD-9 250.0 - .9	2,752	51.78 (49.9, 53.6)	**98.41 (98.2, 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, 66.2)**	96.82 (96.5, 97.1)			
USA	1997 - 2001	Crane[14]	Clinician documentation in Electronic Medical Record progress notes	Physician Claims	At least one clinician-coded diagnoses	ICD 9 250.0, .1, .2, .3	1,441	93 (86, 100)	99 (99, 100)	91 (83, 99)		
USA	1998 - 1999	Borzecki[13]	Medical Charts	Physician Claims	At least one diagnosis in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over one year	ICD 9 250.x	1,176	97	96			0.92
					At least two diagnoses in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over one year	ICD 9 250.x						0.91
					At least one diagnosis in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over two years	ICD 9 250.x						0.89

					At least two diagnoses in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over two years	ICD 9 250.x						**0.93**
USA	1992 - 1995	Hebert[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 - .02, 366.41		**79.1**	94.3	71.4		
USA	1993 - 1994	O'Connor[20]	Telephone Survey	Physician Claims	Two or more ICD-9 diagnostic codes	ICD 9 250.x	1,976	92.22*	98.62*	76.15*	99.63*	
USA	1996 - 1998	Singh[19]	Self-reported Survey	Physician Claims	Veterans Affairs databases	ICD 9 250		76 (75 - 76)	98 (98 - 98)	91 (91 - 91)	95 (94 - 95)	0.79 (0.79 - 0.80)
USA	1997	Ngo[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 - 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 - 0.85)
USA	1997 - 2000	Miller[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 out-patient 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 out-patient 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 out-patient 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 out-patient codes	ICD 9 250, 357.2, 362.0, 366.41		63.8	**99.2**	**96.2**		

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
Note: *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.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.

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

Table 3: Study Characteristics and Test Measures of Studies for Hospital Discharge Data
(Superior performance characteristics within studies have been highlighted in bold.)

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)	Kappa
Canada	1995 - 2000	So[24]	Medical Chart	Hospital Discharge Data	Diabetes with Complications	ICD-9 250.1 - .9	93	**80 (51.91, 95.67)**	98.3 (95.15, 99.65)	80 (51.91, 95.67)	**98.3 (95.15, 99.65)**	
	2001 - 2004				Diabetes with Complications	ICD-10 E10.0 - .8, E11.0 - .8, E12.0 -.8, E13.0 -.8, E14.0 -.8		66.7 (38.38, 88.18)	**98.9 (96.00, 99.86)**	**83.3 (51.59, 97.91)**	97.2 (93.67, 99.10)	
Canada	2003	Quan[26]	Medical Chart	Hospital Discharge Data	Diabetes with Chronic Complications	ICD 9 250.4 - .7	4,008	**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, E110, .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[28]	Medical Chart	Hospital Discharge Data	Look back period: Index admission	ICD 9/ICD-9 CM 250	1,685	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	2,258	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**
					10-years			89.6	95.6 †	91.6	**94.5**	0.856
					15-years			**89.6**	95.5 †	91.5	**94.5**	0.855
Canada	1989 - 1990	Robinson[22]	Self-reported Survey	Hospital Discharge Data and Physician Claims	1, 2 or 3 physician claim or 1 hospitalization over 3 years	ICD 9 CM	2,651	72	98	76	98	0.72 (0.67 - 0.77)

PPV: Positive Predictive Value

NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
Note: *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.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

Table 4: Study Characteristics and Test Measures of Studies for both Physician Claims Data and Hospital Discharge Data
(Superior performance characteristics within studies have been highlighted in bold.)

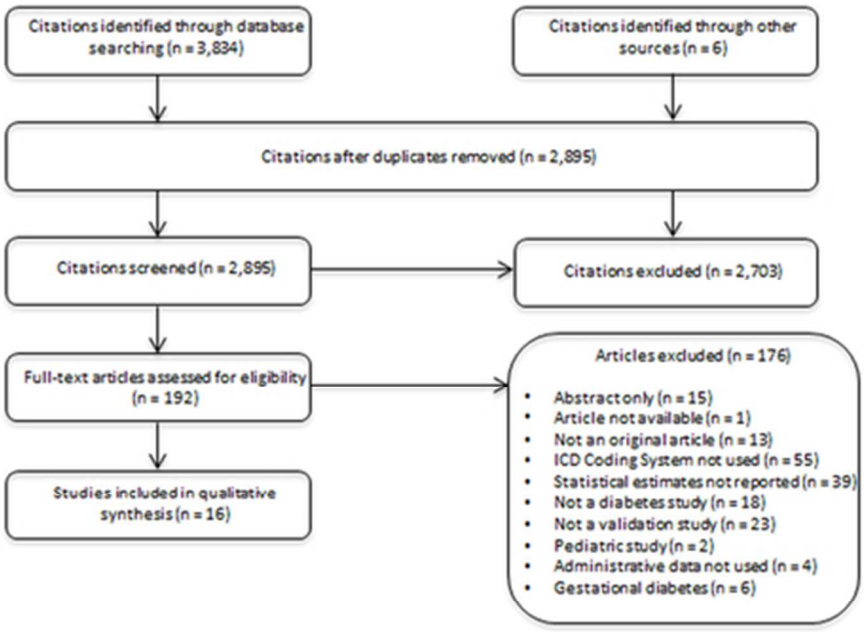
Country	Study Years	Author[Reference]	Reference	Type of Administrative Data	Diabetes Case Definition	ICD Codes Used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Kappa
Canada	1992 - 1999	Hux[21]	Medical Chart	Physician Claims and Hospital Discharge Data	One Physician Service Claims or One Hospitalization with diagnosis of diabetes	ICD-9 250.x	3,317	**91**	92*	61	**99***	
					Two Physician Service Claims or One Hospitalization with diagnosis of diabetes	ICD-9 250.x		86	**97***	**80**	98*	
Canada	2000 - 2002	Chen[25]	Medical Chart	Physician Claims and Hospital Discharge Data	3 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x - 14.x	3,362	**95.6 (92.5–97.7)**	92.8(91.9–93.7)	54 (49.6–58.5)	**99.6 (99.4–99.8)**	0.65 (0.61–0.69)
					2 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x - 14.x		86.4 (82.4–90.5)	**97.1 (96.5–97.7)**	**72.4 (67.5–77.3)**	98.8 (98.4–99.2)	**0.77 (0.73–0.81)**
				Physician Claims	3 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x - 14.x		**91.2 (87.9–94.6)**	97.6 (97.1–98.1)	72.1 (67.5–76.9)	**99.2 (98.9–99.5)**	**0.82 (0.78–0.85)**
					2 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x - 14.x		76.6 (71.5–81.6)	**99.3 (99.0–99.6)**	**90.9 (87.2–94.6)**	98 (97.5–98.4)	**0.82 (78.0–85.5)**

USA	1999	Rector[17]	Telephone surveys	Hospital Discharge Data and Physician Claims	One 1999 claim with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	3,633	**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			
					Two 1999 face-to-face encounter claims with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41		57	**99**			
	1999 - 2000				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,		87	96			

						362.0x, 366.41						
					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**			
Canada	1980 - 1984	Young[27]	Self-reported Survey	Hospital Admission and Physician Claims	[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC)] AND [Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment]	ICD 9-CM	1,000	82.7	96.3			
					[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC)	ICD 9-CM		82.1	**98.5**			

					or payment] AND [Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment]							
					[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC)] AND [Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment] AND [Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment]	ICD 9-CM		**83.9**	95.8			

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
Note: *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.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.



186x139mm (72 x 72 DPI)

Appendix A: PRISMA Checklist 2009

Section	#	Checklist Item	Reported on Page Number
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 3
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix A
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Page 4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A

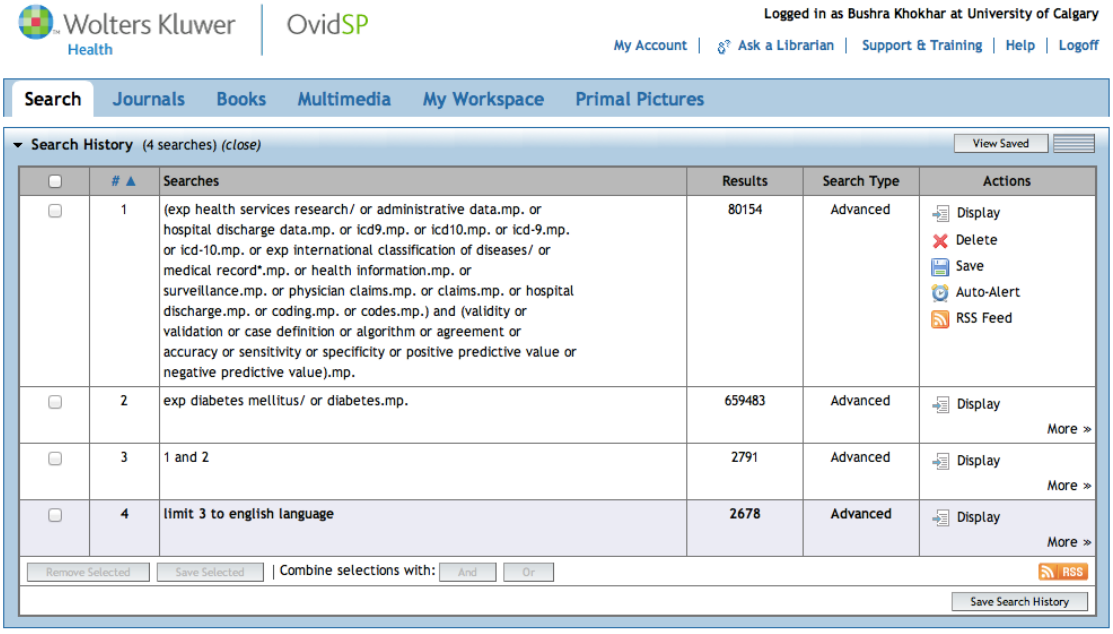
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g. I^2) for each meta-analysis.	N/A
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Tables 1 - 4
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Pages 6,7

Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Page 7
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Page 7
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Page 8

PRISMA 2009 Checklist was extracted from Liberati A. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventio Explanation and Elaboration. Ann. Intern. Med. 2009;339:b2700[11].

Appendix B: Search Strategies

Embase Search Criteria



Medline Search Criteria

BMJ Open

Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data in Adult Populations.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-009952.R4
Article Type:	Research
Date Submitted by the Author:	29-Jun-2016
Complete List of Authors:	Khokhar, Bushra; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health Jette, Nathalie; University of Calgary, O'Brien Institute for Public Health; University of Calgary, Clinical Neurosciences Metcalfe, Amy; University of Calgary, Department of Obstetrics and Gynecology; Alberta Children's Hospital Research Institute, Cunningham, Ceara Tess; University of Calgary, Community Health Sciences Quan, Hude; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health Kaplan, Gilaad; University of Calgary; University of Calgary, O'Brien Institute for Public Health Butalia, Sonia; University of Calgary, Community Health Sciences Rabi, Doreen; University of Calgary, Community Health Sciences; University of Calgary, O'Brien Institute for Public Health
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health informatics, Epidemiology, Diabetes and endocrinology, Health services research, Diagnostics
Keywords:	diabetes, validation studies, case definition, administrative data

Title: Systematic Review of Validated Case Definitions for Diabetes in ICD-9 and ICD-10 Coded Data in Adult Populations

Corresponding Author: Bushra Khokhar
Postal Address: 3280 Hospital Drive NW, 3rd floor TRW Calgary, Alberta T2N 4N1, Canada
E-mail: bushra.khokhar@ucalgary.ca
Telephone Number: (403) 210-3807

Author List: 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], and Doreen Rabi[1,2,6]

*Corresponding author

[1]Department of Community Health Sciences, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada

[2]O'Brien Institute for Public Health, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada

[3]Department of Clinical Neurosciences & Hotchkiss Brain Institute, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada

[4]Department of Obstetrics and Gynecology, University of Calgary, 1403 29 Street NW, Calgary, Alberta T2N 2T9, Canada

[5]Alberta Children's Hospital Research Institute, 3330 Hospital Drive NW, Calgary, Alberta T2N 4NI, Canada

[6] Division of Endocrinology, Department of Medicine, 1820 Richmond Road SW, Calgary, Alberta TCC CCC

Keywords: diabetes, validation studies, case definition, and administrative data

Word Count (excluding title page, abstract, references, figures and tables): 3,052words

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 International Classification of Diseases codes) for diabetes in adults was validated against a reference and statistical measures of the performance reported.

Results: The search yielded 2,895 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%, PPV ranged from 71.4 to 96.2%, NPV ranged from 95 to 99.6% and kappa 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 kappa 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 kappa ranged from 0.7 to 0.8).

Conclusion: 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 depends on the variations in the definition of primary diagnosis in ICD-coded discharge data, and/or the methodology adopted by the health care facility to extract information from patient records.

STRENGTHS AND LIMITATIONS

- 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 whose 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.

BACKGROUND

Diabetes is a chronic disease that has increased substantially during the past 20 years[1]. At present, diabetes is the leading cause of blindness[2], renal failure[3], non-traumatic lower limb amputations[4], and is a major risk factor for cardiovascular disease[5]. Because of 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 health care 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 analyze routinely collected health-related information from healthcare systems and provide reports on risk factors, care practices, morbidity, 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) 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) guidelines[11][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 [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 (e.g. ICD-9-CM).

Study Selection

Studies were evaluated in duplicate for eligibility in a two-stage procedure. In stage one, all identified titles and abstracts were reviewed and in stage two, 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 one, it was included in the full-text review in stage two. 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, kappa] 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 (e.g. self-report from population-based surveys or patient medical chart reviews); and (5) if it reported on original data. Studies validating diabetes in specialized populations (e.g. cardiovascular disease) were excluded to ensure the diabetes case definitions would be generalizable. Studies not employing a sole medical encounter data in their diabetes case definition (e.g. 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 kappa 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 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 2,895 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 United States[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 both 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, eight used medical records[13-14, 21, 23-26, 28], and eight 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 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, out 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 kappa in six. All 16 studies were categorized 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 kappa ranged from 0.8 to 0.9. Four out of the eight studies using physician claims data had a least one diabetes case definition where both 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 over time. This relationship is consistent with the study[18] 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 (in-patient or out-patient) was used, while the specificity and PPV were the highest

when most number of out-patient 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 kappa ranged from 0.6 to 0.9. Two out of the four studies using hospital discharge data had a least one diabetes case definition where both 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 kappa 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 both sensitivity and specificity exceed 80%.

Another factor affecting the statistical estimates is the number of claims being used in the definition. Rector et al.'s study[17] shows consistent results where the sensitivity is higher when at least one claims data is used in the definition, but the specificity is higher when at least two are used. Lastly, Young et al.'s study[27] demonstrates the highest sensitivity when 2 physician claims and 2 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 United States and four studies from Canada. Four studies use both 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 kappa 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 kappa 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 one 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 (NDSS) 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, 87.4) and specificity of 97.9% (95% CI 96.5, 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 standardized manner. Billing practices do vary by practitioner, which may in turn be influenced by the nature of physician reimbursement (salary versus 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 hospitalization, 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 standardized 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 hospitalization, within a certain period of time is appropriate. It is also important to recognize 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 out 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. Lastly, 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 charts 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 done.

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 over time 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 biologic or physiologic definition that denotes the presence or absence of disease, the administrative data definitions are a surrogate of disease, and denote presence or 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 health care 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.

Lastly, 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.

Generalizability

Fifteen out 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.

CONCLUSION

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 health care 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.

COMPETING INTERESTS

The authors declare that they have no competing interest.

AUTHOR'S CONTRIBUTIONS

Dr. Nathalie Jette wrote the protocol. Ms. Bushra Khokhar, Dr. Amy Metcalfe, and Dr. Ceara Tess Cunningham carried out the systematic review. Bushra Khokhar wrote the manuscript. Dr. Nathalie Jette, Dr. Hude Quan, Dr. Gilaad G. Kaplan, Dr. Sonia Butalia,

and Dr. Doreen Rabi provided final approval of the version to be published. All authors read and approved the final manuscript.

FUNDING SOURCES

Ms. Bushra Khokhar was supported by the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD) and The Western Regional Training Centre for Health Services Research (WRTC). Dr. Nathalie Jette 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. Dr. Ceara Tess Cunningham is funded by a Canadian Institute of Health Research doctoral research scholarship. Dr. Kaplan is a Population Health Investigator supported by Alberta Innovates - Health Solutions. Dr. Doreen Rabi is a Population Health Investigator supported by Alberta Innovates - Health Solutions.

DATA SHARING

Any additional data, such as study protocol, data extraction forms, etc. are available by emailing the first author at bushra.khokhar@ucalgary.ca

FIGURE LEGEND

Figure 1. Study Flow Chart.

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, and Dillon J. Early Diagnosis of Diabetes Through The Eye. *Photochem Photobiol* Published Online First 27 August 2015. doi: 10.1111/php.12524.
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 Sep 7;100:1134-46.
6) World Health Organization. Diabetes: The Cost of Diabetes. [http://www.who.int/mediacentre/factsheets/fs236/en/] Accessed August 27,

2014.
7) Jutte DR, Roos LL, and 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: 1792-99.
9) Centers for Disease Control and Prevention: National Notifiable Diseases Surveillance System (NNDSS). [http://wwwn.cdc.gov/nndss/case-definitions.html]. Accessed September 1, 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. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Ann. Intern. Med. 2009;339:b2700.
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, and 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:B10-21.
19) Singh JA. Accuracy of Veterans Affairs Databases for Diagnoses of Chronic Diseases. *Preventative Chronic Disease*. 2009;6:A126.
20) O'Connor PJ, Rush W A, 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;25:512-6.

22) Robinson JR, Young TK, Roos LL, et al. Estimating the burden of disease. Comparing administrative data and self-reports. *Medical Care*. 1997;35:932-47.
23) Wilchesky M, Tamblyn RM, and Huang A. Validation of Diagnostic Codes within Medical Services Claims. *J Clin epidemiol.* 2004;57:131-41.
24) So L, Evans D, and Quan H. ICD-10 coding algorithms for defining comorbidities of acute myocardial infarction. *BMC Health Serv Res*. 2006;6.
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;4: 1424-41.
27) Young TK, Roos NP, and 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;75:313-9.
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: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 August 28, 2014.
32) Roos LL, Roos NP, Cageorge SM, Nicol P. How good are the data? Reliability of one health care data bank. Med Care 1982;20(3):266–76.
33) Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. J Clin Epidemiol 2000;53(12):1258–67.
34) Carral F, Olveira 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;33:593–598.
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. *J Clin Epidemiol*. 1996;49:1407- 17.
39) Mackenbach JP, Looman CW, and van deer 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, Gutierrez-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.

Table 1. Study Quality Characteristics using QUADAS Tool

QUADAS Tool Item	Hux[21]	Robinson[2]	Borzecki[1]	Wilchesky[2]	Crane[1]	So[24]	Chen[25]	Nedkoff[2]	Quan[26]	Young[2]	Hebert[1]	Ngo[1]	Rector[1]	Miller[1]	Singh[19]	O'Connor[2]
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	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

verification using a reference standard of diagnosis?*																
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 (i.e. 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
Were the index test results interpreted without knowledge of	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

the results of the reference standard?*																
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 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[12].

Table 2: Study Characteristics and Test Measures of Studies for Physician Claims Data

(Superior performance characteristics within studies have been highlighted in bold.)

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)	Kappa
Canada	1995 - 1996	Wilcheky[23]	Medical Chart	Physician Claims	Using only diagnoses recorded in the claims of study physicians	ICD-9 250.0 - .9	2,752	51.78 (49.9, 53.6)	**98.41 (98.2, 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, 66.2)**	96.82 (96.5, 97.1)			
USA	1997 - 2001	Crane[14]	Clinician documentation in Electronic Medical Record progress notes	Physician Claims	At least one clinician-coded diagnoses	ICD 9 250.0, .1, .2, .3	1,441	93 (86, 100)	99 (99, 100)	91 (83, 99)		
USA	1998 - 1999	Borzecki[13]	Medical Charts	Physician Claims	At least one diagnosis in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over one year	ICD 9 250.x	1,176	97	96			0.92
					At least two diagnoses in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over one year	ICD 9 250.x						0.91
					At least one diagnosis in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over two years	ICD 9 250.x						0.89

					At least two diagnoses in National Department of Veterans Affairs (VA) database, Out-Patient Clinic file over two years	ICD 9 250.x						**0.93**
USA	1992 - 1995	Hebert[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 - .02, 366.41		**79.1**	94.3	71.4		
USA	1993 - 1994	O'Connor[20]	Telephone Survey	Physician Claims	Two or more ICD-9 diagnostic codes	ICD 9 250.x	1,976	92.22*	98.62*	76.15*	99.63*	
USA	1996 - 1998	Singh[19]	Self-reported Survey	Physician Claims	Veterans Affairs databases	ICD 9 250		76 (75 - 76)	98 (98 - 98)	91 (91 - 91)	95 (94 - 95)	0.79 (0.79 - 0.80)
USA	1997	Ngo[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 - 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 - 0.85)
USA	1997 - 2000	Miller[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 out-patient 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 out-patient 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 out-patient 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 out-patient codes	ICD 9 250, 357.2, 362.0, 366.41		63.8	**99.2**	**96.2**		

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
Note: *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.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

Table 3: Study Characteristics and Test Measures of Studies for Hospital Discharge Data

(Superior performance characteristics within studies have been highlighted in bold.)

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)	Kappa
Canada	1995 - 2000	So[24]	Medical Chart	Hospital Discharge Data	Diabetes with Complications	ICD-9 250.1 - .9	93	**80 (51.91, 95.67)**	98.3 (95.15, 99.65)	80 (51.91, 95.67)	**98.3 (95.15, 99.65)**	
	2001 - 2004				Diabetes with Complications	ICD-10 E10.0 - .8, E11.0 - .8, E12.0 -.8, E13.0 -.8, E14.0 -.8		66.7 (38.38, 88.18)	**98.9 (96.00, 99.86)**	**83.3 (51.59, 97.91)**	97.2 (93.67, 99.10)	
Canada	2003	Quan[26]	Medical Chart	Hospital Discharge Data	Diabetes with Chronic Complications	ICD 9 250.4 - .7	4,008	**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, E110, .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[28]	Medical Chart	Hospital Discharge Data	Look back period: Index admission	ICD 9/ICD-9 CM 250	1,685	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	2,258	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**
					10-years			89.6	95.6 †	91.6	**94.5**	0.856
					15-years			**89.6**	95.5 †	91.5	**94.5**	0.855
Canada	1989 - 1990	Robinson[22]	Self-reported Survey	Hospital Discharge Data and Physician Claims	1, 2 or 3 physician claim or 1 hospitalization over 3 years	ICD 9 CM	2,651	72	98	76	98	0.72 (0.67 - 0.77)

PPV: Positive Predictive Value

NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
Note: *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.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.

Table 4: Study Characteristics and Test Measures of Studies for both Physician Claims Data and Hospital Discharge Data
(Superior performance characteristics within studies have been highlighted in bold.)

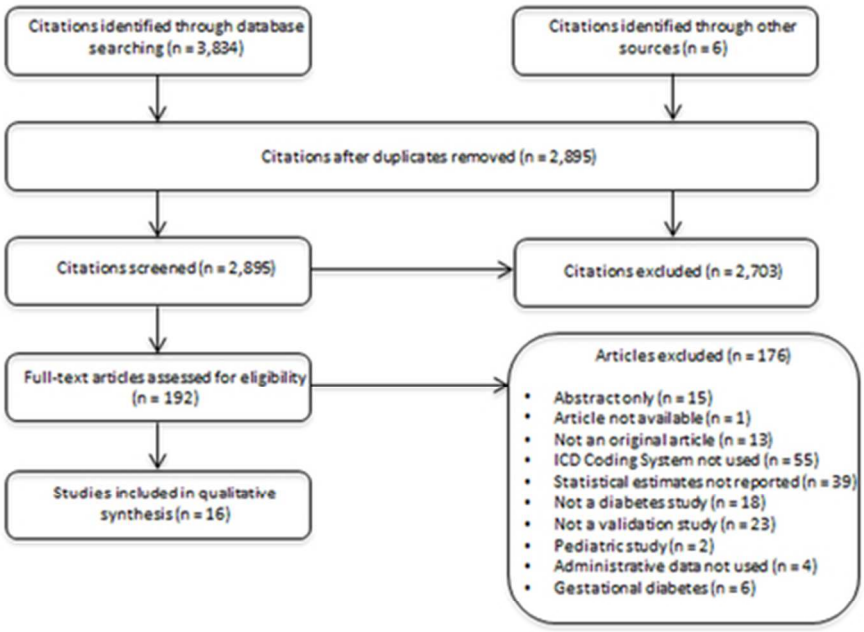
Country	Study Years	Author[Reference]	Reference	Type of Administrative Data	Diabetes Case Definition	ICD Codes Used	Study, N	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Kappa
Canada	1992 - 1999	Hux[21]	Medical Chart	Physician Claims and Hospital Discharge Data	One Physician Service Claims or One Hospitalization with diagnosis of diabetes	ICD-9 250.x	3,317	**91**	92*	61	**99***	
					Two Physician Service Claims or One Hospitalization with diagnosis of diabetes	ICD-9 250.x		86	**97***	**80**	98*	
Canada	2000 - 2002	Chen[25]	Medical Chart	Physician Claims and Hospital Discharge Data	3 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x - 14.x	3,362	**95.6 (92.5–97.7)**	92.8(91.9–93.7)	54 (49.6–58.5)	**99.6 (99.4–99.8)**	0.65 (0.61–0.69)
					2 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x - 14.x		86.4 (82.4–90.5)	**97.1 (96.5–97.7)**	**72.4 (67.5–77.3)**	98.8 (98.4–99.2)	**0.77 (0.73–0.81)**
				Physician Claims	3 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x - 14.x		**91.2 (87.9–94.6)**	97.6 (97.1–98.1)	72.1 (67.5–76.9)	**99.2 (98.9–99.5)**	**0.82 (0.78–0.85)**
					2 Years Observation Period Data	ICD 9 250.xx, ICD 10 E10.x - 14.x		76.6 (71.5–81.6)	**99.3 (99.0–99.6)**	**90.9 (87.2–94.6)**	98 (97.5–98.4)	**0.82 (78.0–85.5)**
USA	1999	Rector[17]	Telephone surveys	Hospital Discharge Data and Physician Claims	One 1999 claim with dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41	3,633	**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			
					Two 1999 face-to-face encounter claims with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41		57	**99**			
	1999 - 2000				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,		91	95			

						362.0x, 366.41						
					Two 1999 or 2000 face-to-face encounter claims with primary dx	ICD 9 250.xx, 357.2x, 362.0x, 366.41		77	**98**			
Canada	1980 - 1984	Young[27]	Self-reported Survey	Hospital Admission and Physician Claims	[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC)] AND [Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment]	ICD 9-CM	1,000	82.7	96.3			
					[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment] AND [Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment]	ICD 9-CM		82.1	**98.5**			

					[Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC)] AND [Hospital admissions of provincial residents claims for which are submitted to the Manitoba Health Services Commission (MHSC) AND Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment] AND [Claims by the physician to the Manitoba Health Services Commission (MHSC) or payment]	ICD 9-CM		**83.9**	95.8			

PPV: Positive Predictive Value
NPV: Negative Predictive Value
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD 10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
Note: *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.
Positive predictive value is the probability that subjects with a positive screening test truly have the disease/condition.
Negative predictive value is the probability that subjects with a negative screening test truly do not have the disease/condition.
Kappa is an inter-rater agreement statistic to evaluate the agreement between two classifications on ordinal or nominal scales.



186x139mm (72 x 72 DPI)

Appendix A: PRISMA Checklist 2009

Section	#	Checklist Item	Reported on Page Number
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 3
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix A
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Page 4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g. I^2) for each meta-analysis.	N/A
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Tables 1 - 4
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Pages 6,7

Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Page 7
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Page 7
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Page 8

PRISMA 2009 Checklist was extracted from Liberati A. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventio Explanation and Elaboration. Ann. Intern. Med. 2009;339:b2700[11].

Appendix B: Search Strategies

Embase Search Criteria

Medline Search Criteria