POPULATION PREVALENCE OF DIABETES: VALIDATION OF A CASE DEFINITION FROM HEALTH ADMINISTRATIVE DATA USING A POPULATION-BASED SURVEY AND HOME BLOOD GLUCOSE SAMPLING

Aaron Leong, MD
Department of Epidemiology, Biostatistics and Occupational Health
McGill University, Montreal
April 2013

Supervisor
Elham Rahme, PhD
Co-supervisor
Kaberi Dasgupta, MD, MSc

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Epidemiology

© Aaron Leong, Montreal, Canada 2013
Prevalence of diabetes

Qualitative synthesis (Table 1) 10
Test performance of the NDSS criteria against self-report in surveys (Table 1) 12
Test performance of the NDSS criteria against medical records/laboratory data/prescription dispensation data (Table 1) 12
Meta-analysis 13
Additional analyses 14

DISCUSSION 15
Reference standards 17
Strengths and limitations 18
Generalizability 20

CHAPTER 1 CONCLUSION 28
Accuracy of alternative administrative algorithms 28
Validation of the NDSS case definition in Quebec (Chapter 2) 30

CHAPTER 2: ESTIMATING THE POPULATION PREVALENCE OF DIAGNOSED AND UNDIAGNOSED DIABETES 32

ABSTRACT 33

RESEARCH DESIGN AND METHODS 34

RESULTS 37
Stratified random sample 37
Validation of the NDSS case definition 38
Diabetes prevalence within the random sample 38
Estimating the population prevalence of diabetes 39

CONCLUSIONS 39
Principal findings 39
Correction factors 40
Prevalence of undiagnosed diabetes 42
Utility of mailed-in blood samples for diabetes screening 44
Strengths and limitations 44
### CHAPTER 2 CONCLUSION

Updating the meta-analysis (Chapter 1) with the Quebec validation study (Chapter 2)  
Applying correction factors to account for undiagnosed diabetes in Canada  
Specificity of the NDSS case definition  
Other limitations: Imperfect reference standards  
Other limitations: Measurement error of fasting glucose on mailed-in samples  
Other limitations: Selection bias from non-response  
Other limitations: Generalizability  
Overall conclusions

### REFERENCES
LIST OF TABLES

Chapter 1
Table 1: Test properties of the NDSS case definition 23

Chapter 2
Table 1. Selected baseline characteristics from RAMQ administrative databases and QSI survey item responses. 48
Table 2. 2-by-2 table for diabetes by NDSS criteria against (1) self-reported diabetes and (2) self-reported diabetes and/or undiagnosed diabetes (total diabetes) 49
Table 3. Test properties of NDSS criteria for self-reported diabetes and total diabetes 50
Table 4. Prevalence of physician-diagnosed diabetes and total diabetes 51

LIST OF FIGURES

Chapter 1
Figure 1: Flow diagram of selection strategy and article reviews 22
Figure 2: Forest plots of sensitivities and specificities of included validation studies by reference standard 23
Figure 3: Random-effects bivariate regression analysis of pooled test accuracies from 6 studies 26
Figure 4: Crude and adjusted prevalence of diabetes in Canada 27

Chapter 2
Figure 1. Flow diagram of participants in QSI survey and home fasting blood glucose sampling. 47
Supplemental figure S1: Participants’ instructions for home capillary glucose sampling 52
Chapter 2 conclusion

Figure 1: Random-effects bivariate regression analysis of pooled test accuracies from 7 studies (including Quebec study) 62

Figure 2: Prevalence of total diabetes in Canada 63

LIST OF SUPPLEMENTAL FILES

Chapter 1

Supplemental file 1: MEDLINE and EMBASE Search strategies 77
Supplemental file 2: The QUADAS tool 80
Supplemental file 3: Quality assessment by QUADAS 81
Supplemental file 5: Adjusted and unadjusted prevalence of diabetes in Canada 82
Supplemental file 6: Funding sources of included validation studies 83
Supplemental file 7: PRISMA checklist 84
ABSTRACT

Background: Previous validation studies of diabetes case definitions from health administrative data are scarce and have not systematically assessed undiagnosed diabetes. We aimed to determine the accuracy of a diabetes case definition (two physician claims or one hospitalization for diabetes, within a two-year period; Canadian National Diabetes Surveillance System, NDSS) using a systematic review of the published literature and by comparing cases identified from the Quebec health services administrative databases to those obtained from self-report, respectively. We also estimated the prevalence of physician-diagnosed diabetes by self-report in a health survey, and undiagnosed diabetes by measuring glucose levels on mailed-in capillary blood samples.

Methods: For the systematic review, we searched Medline (from 1950) and Embase (from 1980) databases for relevant validation studies published through August 2012 (keywords: “diabetes mellitus”, “administrative databases” and “validation studies”). Reviewers abstracted study data and assessed quality using standardized forms. As there was heterogeneity, a random-effects bivariate regression model was used to pool sensitivity and specificity estimates. To determine the accuracy of the NDSS case definition from the Quebec health services administrative databases, we obtained administrative data on a stratified random sample of 6,247 Quebec individuals (2009) whom we surveyed by telephone to query diabetes status and asked them to mail-in fasting capillary blood samples to a central laboratory for glucose testing. The NDSS case definition was compared with self-reported diabetes alone and with self-reported diabetes and/or elevated glucose level (≥7 mmol/l) for sensitivity, specificity, positive and negative predictive values, and statistical agreement. Population-level prevalence was estimated using the NDSS definition, corrected based on sensitivity and specificity estimates and sampling weights. In addition, we added the Quebec validation study to the meta-analysis and reported the final test properties of the NDSS case definition.

Results: The search strategy identified 1423 abstracts among which 11 studies were deemed relevant and reviewed; 6 of these reported sensitivity and specificity allowing pooling in a meta-analysis. Compared to surveys or
Prevalence of diabetes

medical records, sensitivity was 82.3% (95%CI 75.8, 87.4) and specificity was 97.9% (95%CI 96.5, 98.8). NDSS-based diabetes cases obtained from the Quebec health administrative databases compared to self-report revealed a sensitivity of the NDSS case definition of 84.3% (95%CI 79.3, 88.5) and specificity of 97.9% (95%CI 97.4, 98.4). Compared to self-report combined with glucose testing, sensitivity was 58.2% (95%CI 52.2, 64.6) and specificity was 98.7% (95%CI 98.0, 99.3). Adjusted for sampling weights, physician-diagnosed diabetes prevalence in Quebec was 7.2% (95%CI 6.3, 8.0) and total diagnosed and undiagnosed diabetes prevalence was 13.4% (95%CI 11.7, 15.0).

Conclusion: Including the Quebec study in an updated meta-analysis of 7 studies, the pooled sensitivity of the NDSS case definition was 82.6% (95%CI 77.1, 87.0) and specificity was 97.9% (95%CI 96.8, 98.6) for physician-diagnosed diabetes. The NDSS case definition is sufficiently accurate for surveillance purposes, in particular monitoring trends over time. The definition however misses approximately one fifth of physician-diagnosed cases and approximately 40% of those diagnosed and undiagnosed; it wrongly identifies diabetes in around 2% of the general population. Individuals with undiagnosed diabetes are likely to have a delay in diabetes treatment which implies a higher risk for diabetes-related complications. Diabetes prevalence estimated from health services administrative databases should be adjusted for the sensitivity and specificity of the case definition to better quantify yearly prevalence changes and account for undiagnosed diabetes.
ABRÉGÉ

Contexte: Les études qui valident les cas de diabète identifiés à partir des données administratives sont limitées et n’évaluent pas systématiquement le diabète non diagnostiqué. Nous avons mené une revue systématique pour déterminer la validité d’un algorithme utilisé pour identifier les cas de diabète (deux facturations de médecins ou une hospitalisation pour le diabète, pendant une période de deux ans; Système National de Surveillance du Diabète du Canada, SNSD). Nous avons aussi validé cet algorithme dans les bases de données du Québec en comparaison avec des données obtenues à partir d’une enquête publique. En plus, nous avons estimé la prévalence du diabète diagnostiqué à partir des données d’une enquête publique ainsi que la prévalence du diabète non diagnostiqué à partir du taux de la glycémie mesurée en utilisant des échantillons sanguins envoyés par la poste.


Résultats: Dans la revue systématique, la stratégie de recherche a identifié 1,423 résumés desquels 11 études de validation ont été choisies pour la revue. Six études ont été regroupées dans une méta-analyse. En
comparaison aux données des enquêtes et/ou des dossiers médicaux, la sensibilité était de 82,3% (IC 95%, 75,8, 87,4%) et la spécificité était de 97,9% (IC 95*: 96,5, 98,8%). Pour la validation de la définition dans les banques de données du Québec, la comparaison avec les données obtenues à partir de l’enquête a montré une sensibilité de 84,3% (IC 95*: 79,3, 88,5) et une spécificité de 97,9% (IC 95*: 97,4, 98,4). La comparaison avec les données de l’enquête combinées aux taux de glycémie, a montré une sensibilité beaucoup plus faible de 58,2% (IC 95% 52,2, 64,6) et une spécificité de 98,7% (IC 95*: 98,0, 99,3). Après ajustement pour le poids d’échantillonnage, la prévalence du diabète diagnostiqué était de 7,2% (IC 95% 6,3, 8,0) et la prévalence du diabète diagnostiqué et non-diagnostiqué était de 13,4% (IC 95% 11,7, 15,0).

**Conclusion:** Incluant l’étude du Québec dans une méta-analyse actualisée des 7 études, la sensibilité de la définition de cas de diabète diagnostiqué était de 82,6% (IC 95%, 77,1, 87,0) et la spécificité était de 97,9% (IC 95*: 96,8, 98,6). La définition semble être suffisamment précise pour une surveillance en santé publique, en particulier pour les analyses de tendance. Les personnes non-diagnostiquées sont susceptibles de subir un retard dans le traitement et d’encourir un risque plus élevé pour les complications liées au diabète. La prévalence du diabète estimée à partir des banques de données administrative doit être corrigée pour la sensibilité et la spécificité de la définition afin de mieux quantifier les changements annuels de prévalence et de tenir compte des cas de diabète non diagnostiqués.
CONTRIBUTION OF AUTHORS

This thesis amalgamates two manuscripts with linking sections. The first paper (chapter 1) is entitled “Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records”. The author list for the first paper is: Aaron Leong, Kaberi Dasgupta, Diane Lacaille, Sasha Bernatsky, Antonio Avina-Zubieta and Elham Rahme. We have submitted this paper to the PLOS ONE in April 2013.

The second paper (chapter 2) is entitled “Estimating the population prevalence of diagnosed and undiagnosed diabetes”. The author list for the second paper is: Aaron Leong, Kaberi Dasgupta, Jean-Louis Chiasson, and Elham Rahme. This paper was submitted to Diabetes Care in December 2012 and was accepted for publication in March 2013.

I am first author on both papers and have contributed substantially to this body of work. Specifically, I conducted the literature search and summarized the literature, formulated the study questions, performed the data analysis and reported the results. I took a leadership role in preparing all components of the manuscripts, manuscript writing and revisions.

KD is second author on both manuscripts and played a major role in manuscript revisions. She contributed to the formulation of the study designs, and was involved in all aspects of results reporting, manuscript preparation and revisions. SB contributed to the conceptualization of the study question and reviewed the final manuscript (chapter 1). DL and AVZ both assisted in formulating the aims of the systematic review and developing the search strategy in the first paper (chapter 1). JLC contributed to the formulation of the study design in the second paper (chapter 2). He led the collection of blood samples from all participants and the glucose testing in the central laboratories at Saint-Luc Hospital.

ER is senior author on both manuscripts; she led the formulation of study designs and coordinated the data-linkage between health administrative data, survey data and blood samples data. She supervised all aspects of data analysis, results reporting, and manuscript preparation and revisions. The study was funded by a grant from the Canadian Institutes of Health Research held by ER (Principal Investigator), and co-investigators KD and JLC.
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my research supervisors for their unparalleled supervision. The quality of the research presented in this thesis is a testament to the natural synergism among the three of us. I will forever benefit from their mentorship and career guidance: Dr. Elham Rahme, PhD and Dr. Kaberi Dasgupta, MD, MSc, Research Institute of the McGill University Health Centre (MUHC). I would also like to acknowledge the invaluable contribution of our research assistant: Youssef Habel, BSc MPhil, Research Institute of MUHC. Several esteemed professors have graciously shared their epidemiological and statistical knowledge, and assisted in the early stages of data analyses: Dr. Madhu Pai, MD, PhD, Dr. Sam Harper, PhD, Dr. Aurelie Labbe, PhD, Dr. Lawrence Joseph, PhD, McGill Epidemiology and Biostatistics Department.

The following collaborators from external academic institutions have been absolutely instrumental in this body of work: Dr. Jean Louis Chiasson, MD, Department of Médecine, Centre hospitalier de l’université de Montréal Hôtel-Dieu, Dr. Diane Lacaille, MD, MHSc and Antonio Avina-Zubieta, MD, MSc, PhD, Division of Rheumatology, University of British Columbia, and Mary-Doug Wright, MLS, Apex Information, Vancouver, British Columbia. I would like to thank the Régie de l’Assurance Maladie du Québec (RAMQ) for providing the administrative data and the personnel from the ISQ who oversaw the survey: Lucille Pica, Brigitte Beauvais, Johanne Théroux, Nathalie Plante and Robert Courtemanche, and Lyne Labrecque, Saint-Luc Hospital, for conducting the blood tests.

I appreciate the sincere advice from Dr. Mark Blostein, MD, PhD, program director of the Clinical Investigator Program, Dr. Pierre Ernst, MD, Intensive MSc in epidemiology for clinicians’ program director, Dr. Simon Wing, MD, head of endocrinology department, MUHC, Dr. Natasha Garfield, MD, endocrinology fellowship program director and Dr. Brent Richards, MD, my trusted research mentor. I am eternally indebted to friends, family, faculty members, colleagues, classmates, teaching assistants and administrative support staff for their support and genuine desire to watch me succeed. Intercalating this Master’s degree in epidemiology with my clinical
endocrinology fellowship training could only be realized with the strong commitment to fellow’s research of the McGill endocrinology department.

This research work was funded by a CIHR grant awarded to Principal Investigator, Dr. Elham Rahme, and Co-investigators, Drs. Kaberi Dasgupta and Jean-Louis Chiasson. I am salaried through a personal support award from RAMQ: Formation Complémentaire.
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATC</td>
<td>Anatomical Therapeutic Code</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>CCHS</td>
<td>Canadian Community Health Survey</td>
</tr>
<tr>
<td>CIHI</td>
<td>Canadian Institute of Health Information</td>
</tr>
<tr>
<td>CIHR</td>
<td>Canadian Institute of Health Research</td>
</tr>
<tr>
<td>HEDIS</td>
<td>Health Plan Employer Data and Information Set</td>
</tr>
<tr>
<td>HMO</td>
<td>Health Maintenance Organization</td>
</tr>
<tr>
<td>FPG</td>
<td>Fasting Plasma Glucose</td>
</tr>
<tr>
<td>ICD</td>
<td>International Statistical Classification of Diseases and Related Health Problems</td>
</tr>
<tr>
<td>QSI/ISQ</td>
<td>Quebec Statistical Institute/Institut de la statistique de Québec</td>
</tr>
<tr>
<td>MCBS</td>
<td>Medicare Current Beneficiary Survey</td>
</tr>
<tr>
<td>NDSS</td>
<td>National Diabetes Surveillance System</td>
</tr>
<tr>
<td>NPHS</td>
<td>National Population Health Survey</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative Predictive Value</td>
</tr>
<tr>
<td>ODB</td>
<td>Ontario Drug Benefit</td>
</tr>
<tr>
<td>ODD</td>
<td>Ontario Diabetes Database</td>
</tr>
<tr>
<td>OHIP</td>
<td>Ontario health Insurance Plan</td>
</tr>
<tr>
<td>PRISMA</td>
<td>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</td>
</tr>
<tr>
<td>RAMQ</td>
<td>Régie de l’assurance maladie du Québec</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive Predictive Value</td>
</tr>
</tbody>
</table>
QUADAS: Quality assessment tool for diagnostic accuracy studies

SD: Standard Deviation

SROC: Summary Receiver Operating Characteristic
INTRODUCTION

This thesis amalgamates two manuscripts that have been submitted independently for publication. Each manuscript is presented as a separate chapter (chapter 1 and 2) with connecting texts, entitled “Chapter conclusions”, that provide overall conclusions to the preceding chapter and logical bridges between chapters.

Claims-based health administrative data are frequently used for diabetes surveillance and prevalence estimation. However, studies that validate diabetes case definitions from administrative databases are limited. Several validation studies have reported test performance measures of various diabetes case definitions from administrative data (1-6) but rarely describe how test accuracies influence diabetes prevalence estimation or provide adjusted estimates of prevalence. Previous validation studies also do not systematically account for the prevalence of undiagnosed diabetes.

This thesis aims to address these information gaps. Chapter 1 serves as a comprehensive review of the relevant literature on the validation of a commonly used diabetes administrative definition (two physician billing claims or one hospitalization for diabetes, within a two year period; Canadian National Diabetes Surveillance System, NDSS). We determined the test accuracy of this case definition through systematic review and meta-analysis.

Similar to previous validation studies, in chapter 2, we generated test properties of the NDSS case definition using self-reported diabetes from a population-based survey as a reference standard. Unlike previous validation studies that have only determined the prevalence of physician-diagnosed diabetes, we additionally estimated the prevalence of undiagnosed diabetes by measuring fasting plasma glucose values on mailed-in capillary blood samples. We estimated the population prevalence of total (diagnosed and undiagnosed) diabetes by applying sampling weights of participants and correction factors derived from the sensitivity and specificity of the NDSS case definition. We demonstrated that the utility of mailed-in blood samples for glucose testing is two folds: first, an efficient method of undiagnosed diabetes prevalence estimation and, second, an innovative and patient-centred diabetes screening tool for at-risk populations.
CHAPTER 1: Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records

Aaron Leong¹, MD; Kaberi Dasgupta¹,², MD, Sasha Bernatsky¹,², MD, PhD; Diane Lacaille³, MD; Antonio Avina-Zubieta³, MD, MSc, PhD and Elham Rahme¹,², PhD

AFFILIATIONS
¹Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada
²Department of Medicine, McGill University, Montreal, Quebec, Canada
³Division of Rheumatology, Department of Medicine, University of British Columbia, British Columbia, Canada

Correspondence: Elham Rahme, Division of Clinical Epidemiology, McGill University Health Centre, Division of Clinical Epidemiology, 687 Pine Avenue West, V Building, Montreal, Quebec H3A 1A1; tel: (514) 934-1934, ext. 44724; fax: (514) 934-8293; email: elham.rahme@mcgill.ca.

Disclaimer: Elham Rahme is Associate Professor in the Department of Medicine of McGill University and holds a Senior Investigator award from the Fonds de Recherche en santé du Québec. Kaberi Dasgupta is Associate Professor of Medicine at McGill University and holds the Fonds de recherche Santé du Québec-Société québécoise d’hypertension artérielle-Jacques de Champlain Award. Sasha Bernatsky is Assistant professor in the Department of Medicine, Division of Rheumatology and Clinical Epidemiology of McGill University. She is a scholar of the Canadian Arthritis Network and holds a Young Investigator Award from the Canadian Arthritis Network. Diane Lacaille is Associate Professor in the Division of Rheumatology at the University of British Columbia and a Senior Research Scientist at the Arthritis Research Centre of Canada. Antonio Avina-Zubieta is Assistant Professor in the Division of Rheumatology at the University of British Columbia, and...
holds the Network Scholar research training award from the Canadian Arthritis Network-The Arthritis Society and the BC Lupus Society. Aaron Leong is Fellow in Endocrinology and Metabolism and the Clinical Investigator Program at McGill University. He is concurrently MSc candidate in Epidemiology at McGill University.

Word count: 5,105

FUNDING

This work was supported by the Canadian Arthritis Network [10-01-RIPP-04] and the Canadian Institutes of Health Research [SSD-83167]

ACKNOWLEDGEMENTS

We would like to thank Ms. Mary-Doug Wright, B.Sc, M.L.S., of Apex Information, British Columbia, for executing the literature search for the systematic review.
ABSTRACT

Objectives: Health administrative data are frequently used for diabetes surveillance. We aimed to determine the sensitivity and specificity of a commonly-used diabetes case definition (two physician claims or one hospitalization for diabetes within a two-year period) and their potential effect on prevalence estimation.

Methods: Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we searched Medline (from 1950) and Embase (from 1980) databases for validation studies through August 2012 (keywords: “diabetes mellitus”; “administrative databases”; “validation studies”). Reviewers abstracted data with standardized forms and assessed quality using Quality Assessment of Diagnostic Accuracy Studies (QUADAS) criteria. A generalized linear model approach to random-effects bivariate regression meta-analysis was used to pool sensitivity and specificity estimates. We applied correction factors derived from pooled sensitivity and specificity estimates to prevalence estimates from national surveillance reports and projected prevalence estimates to year 2018.

Results: The search strategy identified 1423 abstracts among which 11 studies were deemed relevant and reviewed; 6 of these reported sensitivity and specificity allowing pooling in a meta-analysis. Compared to surveys or medical records, sensitivity was 82.3% (95%CI 75.8, 87.4) and specificity was 97.9% (95%CI 96.5, 98.8). The administrative definition underestimated prevalence when it was ≤10.6% and overestimated prevalence otherwise.

Conclusion: The administrative definition for physician-diagnosed diabetes misses up to one fifth of diabetes cases and wrongly identifies diabetes in approximately 2% of the population. This may be sufficiently sensitive and specific for surveillance purposes, in particular monitoring prevalence trends. Applying correction factors to adjust prevalence estimates from the administrative definition may be helpful to increase accuracy of estimates.
INTRODUCTION

Diabetes is a leading cause of blindness, renal failure and cardiovascular disease (7). The direct cost of diabetes and its complications put a substantial strain on healthcare system resources (8-10). The rise in diabetes prevalence has been largely driven by an ageing population, the obesity epidemic and a more sedentary lifestyle (11). In order to adequately project needs and costs of diabetes management, it is crucial to know its actual prevalence and track changes over time.

Administrative databases have become a popular tool for diabetes research and disease surveillance, as they are less prone to recall bias, and potentially more cost efficient, than nationwide surveys(12). Diabetes case identification algorithms can involve a combination of physician billing claims(4), hospitalization records(13), prescriptions data (13-15), and/or records of healthcare services utilization (16). However, the validity of this method for prevalence estimation or diabetes research has not been definitively established.

There are several potential information gaps that can affect prevalence estimation from administrative databases: first, diabetes algorithms that rely on physician billing claims require regular patient use of the health care system (i.e., medical visits at least once a year); second, data coding for diabetes must be accurate and comprehensive; third, some physicians are not on a fee-for-service plan exclusively (i.e., they either receive a salary or are on a mixed remuneration plan), visits to these physicians are not captured in some databases; and fourth, given that patients with diabetes commonly carry multiple comorbidities and are frequently managed by general practitioners (17), physicians may fill billing claims for conditions other than diabetes during their visits when addressing other active or acute conditions (18; 19). For these reasons, claims-based administrative algorithms may not be optimally sensitive.

The National Diabetes Surveillance System (NDSS) comprises regionally distributed diabetes surveillance systems across Canada. Anonymous administrative data relating to diabetes are sent to Health Canada for national analyses and population prevalence estimation. According to the
Prevalence of diabetes

NDSS case definition, a diabetes case fulfils at least one of the following two criteria: two physician billing claims within a two-year period or one hospitalization with an ICD code for diabetes (20). Similar claims-based definitions are used in parts of the United States. The objective of this systematic review and meta-analysis were [1] to determine the overall NDSS case definition performance (sensitivity and specificity) and [2] to estimate diabetes prevalence adjusted for the performance of the NDSS case definition.

METHODS

Search strategy

The systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (21). Two citation indices, Medline and Embase, were searched using an OVID platform. Keywords used included “administrative data”, “validation studies” and “diabetes mellitus” (Supplemental file 1 for search strings). The search strategy was limited to articles ever published through August 18, 2012 that were accessible via these search engines (i.e., from January 1, 1950 for Medline and from January 1, 1980 for Embase). The language of publication was not restricted. We also reviewed the bibliographies of relevant articles (i.e., citation tracking).

Abstract review and abstract exclusion criteria

Each abstract was reviewed independently (AL and ER). We used the following inclusion criteria: [1] test measures were reported; [2] the validated case definition was similar to the NDSS criteria; [3] the data sources were from administrative databases; [4] the study base was a representative sample of the general population and [5] the reference standard, via subject-specific record linkage, was adequate (e.g. self-report from population-based surveys, drug dispensation claims of anti-diabetic medication, laboratory data or primary care medical chart reviews). An example of an inadequate reference standard would be performing the validation test on a non-representative subsample of the study population. If the two investigators, AL and ER,
disagreed, they attempted to reach consensus by discussion. A third investigator (KD) was consulted to serve as a tie-breaking adjudicator.

**Full-text review, quality assessment and data extraction**

Study quality was evaluated using Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS) criteria (Supplemental file 2)(22), as well as consideration of the following potential biases: [1] verification bias (was there a comparison with an independent reference standard with no knowledge of the index test results?), [2] spectrum bias (was there ample representation of patients commonly seen in clinical practice?), [3] review bias (were the index test results interpreted independently of the reference standard results?) and [4] incorporation bias (did the index test form part of the reference standard?). Study data were abstracted using standardized forms that recorded the following information: study population, data sources, administrative algorithms, validation method, reference standards, funding sources, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and kappa statistic. If test measures or 95% confidence intervals were not reported in the original paper, estimates were calculated from data available. For example, we calculated the PPV from the sensitivity, specificity and prevalence when available using the following formula:

\[
PPV(\%) = \frac{\text{sensitivity} \times \text{prevalence}}{[(\text{sensitivity} + \text{specificity} - 1) \times \text{prevalence} + (1 - \text{specificity})]} \]

(23).

**Statistical Analysis**

Using STATA version 11, we generated forest plots and summary receiver operating characteristic (SROC) curves to visually inspect for heterogeneity. Forest plots were arranged according to reference standard used, namely, self-report from surveys or medical chart review. Given that the sensitivity and specificity of each study are calculated from correlated binary outcomes, we judged that simple pooling using weighted average of the sensitivity and specificity independently was inadequate. Thus, we performed a DerSimonian & Laird random-effects bivariate regression analysis using a
Prevalence of diabetes

generalized linear mixed model approach that took heterogeneity and correlation between sensitivities and specificities into account (24; 25). Pooled test accuracies were reported and hierarchical SROC curves were plotted (i.e., HSROC plots of sensitivity and specificity with 95% joint intervals in two-dimensional space). Confidence and prediction regions in the SROC space were constructed using the estimates from the bivariate normal distribution for the random-effects model.

Given the small number of studies, we were unable to perform meta-regression techniques or subgroup analyses to statistically describe the effect of study characteristics on the heterogeneity of test measures. Egger’s test and Begg’s funnel plots were not conducted because there was limited power to detect small-study effects of publication bias and these tests can be misleading in meta-analyses of diagnostic accuracy (26; 27).

Additional analyses

National surveillance reports of diabetes prevalence are not adjusted for the sensitivity and specificity of the diabetes case definition(28). To demonstrate the effect of such adjustments on reported national surveillance results, we adjusted the yearly Canadian population prevalence of diabetes cases(28), using the pooled sensitivity and specificity derived from our study. Based on the law of total probability and Bayes Theorem, the correction formula generated to adjust prevalence was as follows (29):

\[
\text{Adjusted prevalence (\%)} = \frac{[\text{unadjusted prevalence (\%)} + \text{specificity} - 1]}{\text{specificity} + \text{sensitivity} - 1}.
\]

RESULTS

Search results

The search strategy identified 1423 abstracts. Among these, 65 were determined to be potentially relevant for full text review, of which five articles were published in a language other than English (one Danish, one Hebrew,
two Italian and one Korean). The abstracts and method sections were translated by a native speaker of the language and determined not to be eligible for inclusion. A total of 43 studies were excluded for the following reasons: a validation was not performed; the study base was not representative of the general population; the validated case definition was too dissimilar from the NDSS case definition. Twenty two articles were considered for review and data extraction. A flow chart illustrating the selection process is shown in Figure 1.

At the time of full text review, five additional studies were excluded as they examined national registries rather than claims-based approaches (16; 30-33). Five more studies were excluded because of important divergence in the case definition from that used in NDSS. Divergent algorithms excluded physician claims or only used one physician claim from the case definition, or included dispensation of anti-diabetic medication, biochemical information or physician reporting of patient diagnosis in the case definition. Another article was excluded because the study base was not representative of the general population. Ultimately, 11 studies met the eligibility criteria and were included in the systematic review. Included studies are displayed in Table 1.

**Quality assessment**

The QUADAS scores (Supplemental file 3) ranged from 7 to 12 (median 11) out of a maximum of 14. The bias assessment identified two studies with potentially important bias. First, in the Solberg and colleagues’ study, investigators only reviewed the medical records of subjects who had tested positive in administrative data (34). Therefore, the PPV could be reported in the paper but not the sensitivity or specificity without introducing verification bias. Second, the Koleba and colleagues study used prescription data both as part of the diabetes definition and also as the reference standard (incorporation bias) (14). Regardless of quality assessment scores, all 11 studies are discussed in the systematic review.
**Qualitative synthesis (Table 1)**

All 11 studies were conducted in North American. Of these studies, eight were conducted in Canadian provinces (Three Ontario, two Manitoba, one Saskatchewan, one Alberta, and one Alberta/British Columbia) and three in Minnesota, USA.

Hux and colleagues (1) validated the Ontario Diabetes Database (ODD), a registry of diabetes cases identified by the NDSS criteria (age ≥ 20 years; n = 528,280), through record linkage to three independent sources: first, a medication dispensation database from a public medication reimbursement plan for individuals ≥ 65 years of age; second, survey data from the National Population Health survey (NPHS; a stratified random sample that included a query on diabetes); third, a random sample of medical charts (n = 3,317) at physician offices in the community. Another Ontario study (mean age 42.5 years; n = 19,442) by Harris and colleagues (6) examined the concordance of the ODD with two other data sources: a provincial ICD-code based registry developed for the Baseline Diabetes Database Initiative (BDDI) by the Ontario Ministry of Health; and anti-hyperglycemic medication prescriptions, laboratory test results and physician-recorded lists of medical diagnosis (i.e., problem lists) in electronic medical records from primary care practices residing in rural and urban areas of southwestern Ontario as part of the Delivery Primary Health-care Information (DELPHI) project (6). In a different Ontario study, Shah & Manuel (35) validated the ODD against self-report from the Canadian Community Health Survey (CCHS) (age ≥ 18 years; n = 1,812), a cross-sectional survey of health determinants, health status and health care use in the Canadian population (36).

In Manitoba, Robinson and colleagues (4) examined one, two or three physician contacts, defined by physician service claims or hospital summaries from the Manitoba Health Services Commission (MHSC) Data (age range 18 – 74 years; n = 2,792). The reference standard was self-report from the Manitoba Heart Health Project (MHHP), a population-based cross-sectional health survey. Lix and colleagues (37) validated 152 case definition algorithms derived from physician claims, prescription data and hospitalization data on chronic diseases, including diabetes, from Manitoba Health Services Insurance...
Plan (MHSIP) administrative data with CCHS (age ≥ 19 years; n = 5589). Of note, the Manitoba drug benefit program (i.e. Pharmacare) covers all Manitobans and reimbursement is scaled according to taxable income and amount of prescription drug cost.

Koleba and colleagues (14) determined the case capture rate of the NDSS definition in Saskatchewan using drug dispensation records (mean age 52.8 years; n = 145,696). Approximately 90% of the Saskatchewan population are eligible for public prescription benefits.

In Calgary, Southern and colleagues (38) validated the NDSS case definition and a more liberal definition involving single physician claims on a defined cohort of diabetes cases diagnosed by laboratory criteria of elevated fasting or post-prandial blood glucose values, or glycated hemoglobin levels (all ages; n = 25,419).

Chen and colleagues (3) performed their validation study on both rural and urban populations of Alberta and British Columbia and compared algorithms that varied in number of physician claims against medical records. General practitioner clinics were randomly selected from urban and rural areas and medical charts were randomly selected from within each clinic’s patient registries (mean age 52.8 years; n = 3,362).

The three Minnesota studies had different study designs to validate claims-based administrative algorithms similar to the NDSS criteria. First, O’Connor and colleagues (5) validated computerized insurance databases of Health Maintenance Organization (HMO) members in the Upper Midwest with self-report from a telephone survey (adults; mean age 40 years; n = 3,186). Discordant cases had their medical charts reviewed. Second, Hebert et al (2) used self-reported diabetes from the Medicare Current Beneficiary Survey (MCBS) to validate an administrative algorithm from claims data of Medicare beneficiaries. This study, however, was performed only on individuals ≥ 65 years of age who had comprehensive Medicare coverage. Thus, the specificity may be higher among these individuals because of more frequent physician encounters compared to a younger population. The claims data included those pertaining to home health agencies in addition to claims
Prevalence of diabetes

for hospitalizations and outpatient physician encounters. Third, Solberg and colleagues (34) reviewed medical records on a random sample of Medicare beneficiaries in Minnesota to verify the diabetes status of cases identified through NDSS-like case definitions from the Health Plan Employer Data and Information Set (HEDIS; age ≥ 19 years; n = 135 842).

**Test performance of the NDSS criteria against self-report in surveys (Table 1)**

Hux and colleagues reported a sensitivity of 85.0% (95%CI 81.0, 89.0) and a PPV of 64.0% (95%CI 59.0, 69.0) when the ODD was compared to self-report from NPHS (cycle three health component, 1998/1999; n = 4691). The 95% confidence intervals were estimated based on a diabetes prevalence of 6.8% in the ODD (1). Shah and Manual yielded a higher PPV of 74.8% (95%CI 72.8, 76.8%) when the ODD was compared to self-reported diabetes from CCHS (reference standard); as the study cohort was established entirely with diabetes cases identified from the ODD, the sensitivity of the ODD could not be calculated (35). In Manitoba, Robinson and colleagues reported a more modest sensitivity of 75.5% (95%CI 69.2, 81.8%) coupled with a high specificity [98.1% (95%CI 97.6, 98.6%)] (4). These test measures were similar to those reported by the two American studies by O’Connor and colleagues and Hebert and colleagues [sensitivities: 76.1% (95%CI 86.1, 84.1%) and 74.4% (95%CI 71.9, 76.9%), respectively; specificities: 99.6% (95%CI 99.3, 99.9%) and 97.5% (95%CI 97.1, 97.9%), respectively] (2; 5).

**Test performance of the NDSS criteria against medical records/laboratory data/prescription dispensation data (Table 1)**

Chen and colleagues in Alberta/British Columbia demonstrated high sensitivity [92.3% (95%CI 89.2, 95.5%)] and specificity [96.9% (95%CI 96.2, 97.5%)] of the NDSS criteria against medical records (3). Hux and colleagues reported a slightly lower sensitivity [86.1% (95%CI 82.0, 90.2%)] but comparable specificity [97.1% (95%CI 76.5, 97.7%)] of the ODD (1). A similar sensitivity [84.3% (95%CI 82.7, 86.3%)] and specificity [96.9% (95%CI 96.4, 97.5%)] of the ODD against electronic medical records were
found by Harris and colleagues (6). Southern and colleagues yielded a slightly lower sensitivity of 79.1% (78.9, 79.4%) when administrative data from Alberta Health Services were compared to laboratory data (38).

In general, high sensitivities were reported for the NDSS case definition against prescription data, such as the ODB (sensitivity; 91.0%, sample size not available to calculate the 95% CI) by Hux and colleagues for individuals ≥ 65 years of age (1). A similar sensitivity was reported by Koleba and colleagues among adults ≥ 20 years of age who were Saskatchewan Health Beneficiaries [sensitivity; 94.4% (94.2, 94.6)] when results were projected to the entire Saskatchewan population (14).

The NDSS case definition had generally good concordance with medical records (kappa 0.77 - 0.80) and self-reported diabetes from surveys (kappa 0.72 - 0.83). In general, higher concordance between case ascertainment techniques (e.g. diabetes cases from administrative case definitions and self-report from surveys) was reported by American studies compared to Canadian studies.

**Meta-analysis**

We were able to populate four-cell values of diagnostic two-by-two tables from available raw data of 6 studies (1-6). The reported sensitivities ranged from 74.4% to 92.3% (median 85.2%) and specificities ranged from 96.9% to 99.6% (median 97.3%, Figure 2). Studies validated by surveys (n = 3) had lower sensitivities (74.4% to 76.2%) than those validated by medical records (n = 3; 84.3% to 92.3%). The area under the curve (AUC) of the symmetric SROC was 97.7% (95% CI 97.1, 98.3%) and asymmetric SROC was 96.8% (95% CI 92.1, 100.0%) for all 6 studies.

By random-effects bivariate regression analysis, the overall pooled sensitivity was 82.3% (95% CI 75.8, 87.4%) and specificity was 97.9% (95% CI 96.5, 98.8%, Figure 3). The 95% prediction region, which is the confidence region for a forecast of the true sensitivity and specificity in a future study, ranged more widely from under 50% to over 90% for the predicted sensitivity and from approximately 80% to almost 100% for the
Prevalence of diabetes

predicted specificity. A multi-level hierarchical model and random-effects bivariate regression model for subgroups by validation method could not be performed because of the small number of studies.

Additional analyses

NDSS reports prevalence estimates of physician-diagnosed diabetes as the proportion of cases identified via the NDSS case definition in the population. This study demonstrated that the NDSS case definition is not gold standard and misclassifies ~ 20% of diabetes cases and ~ 2% of non-cases. From the Canadian 2009 NDSS report (28), the yearly population prevalence rates of NDSS-identified diabetes cases among adults aged ≥ 20 years between fiscal years 2002/3 and 2006/7 were adjusted by applying the following correction factors based on the pooled test accuracies (sensitivity and specificity) of the NDSS case definition:

\[
\text{Adjusted prevalence (\%) = \left[\text{reported unadjusted prevalence (\%) - 2.1\%}\right] / 80.2\%}.
\]

Figure 4 shows adjusted and unadjusted prevalence estimates plotted against time from fiscal year 2002/3 to 2006/7, respectively. These prevalence estimates were then projected to year 2018. The 95% margin of error for all adjusted and unadjusted prevalence estimates were ≤ 0.01% (population size, n ~ 25 000 000).

The impact of prevalence adjustment depended on the magnitude of diabetes prevalence. Unadjusted prevalence estimates were biased upwards by ~ 1% during the 5-year period (2002/3: unadjusted prevalence was 6.4% and adjusted prevalence was 5.3%; 2006/7: unadjusted prevalence was 8.0% and adjusted prevalence was 7.3%). However, the NDSS case definition underestimated the increase in prevalence over time as reflected by the steeper slope for adjusted prevalence against time (~ 0.4% per year) compared to unadjusted prevalence against time (~ 0.5% per year). Both unadjusted and adjusted prevalence equaled 10.6%, around year 2013. This crossover point occurred when the number of false positives equaled to the number of false
negatives. After year 2013, unadjusted prevalence estimates appear biased downwards.

As the PPVs were not consistently provided in the included studies, they were not pooled. Instead, we estimated the PPV based on the pooled NDSS sensitivity and specificity presented herein, using the following formula:

\[
PPV(\%) = \frac{\text{sensitivity} \times \text{prevalence}}{((\text{sensitivity} + \text{specificity} - 1) \times \text{prevalence} + (1 - \text{specificity}))}
\]

Using the pooled test measures reported herein, \(PPV(\%) = \frac{[82.3 \times \text{prevalence}(\%)]}{[0.802 \times \text{prevalence}(\%) + 2.1]}\). Assuming diabetes prevalence is between 5% and 10%, the PPV falls between 67.3% and 81.3%.

**DISCUSSION**

Our meta-analysis demonstrates that a commonly-used administrative database definition for diabetes (two physician outpatient billings and/or one hospitalization for diabetes 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%), based on the findings of 6 studies with complete data available. While this definition appears to miss approximately one fifth of diabetes cases and wrongly classifies 2.1% of non-cases in the population as diabetes cases, it is likely sufficiently sensitive for monitoring prevalence trends in the general population if its accuracy remains reasonably stable over time. In such situations, this administrative database definition can be particularly useful for tracking prevalence changes over time.

In a previous examination of administrative database definitions for diabetes, Saydah and colleagues (40) performed a literature review of validation studies on a variety of diabetes administrative definitions, gold standards and patient populations, from highly restrictive (e.g. only patients who underwent percutaneous coronary interventions) to nationally representative. The authors included 16 validation studies and reported that diabetes administrative definitions varied from moderately to very sensitive
Prevalence of diabetes

[46.0% to 97.0% (median 81.5%)] but were uniformly very specific [95.0% to
100.0% (median 99.0%)]. The authors did not perform a meta-analysis in that
study. Our study focused specifically on the evaluation of the NDSS
definition. It found that the sensitivity of this case definition ranged from
74.4% to 94.4% with a median of 81.7%; this median is similar to the median
sensitivity of all the diabetes administrative definitions examined by Saydah
and colleagues.

Validation studies included in this review evaluated an administrative
case definition based on physician billings and hospitalization codes and not
medication use. It has been suggested that the sensitivity of a claims-based
administrative algorithm could potentially be improved by incorporating
information from medication dispensation data, without compromising
specificity (14). However, some regions have restricted public medication
insurance coverage; therefore prevalence estimation from prescription data
may not always be representative of the general population. While medication
dispensation information may improve the sensitivity among those reimbursed
by the public healthcare system, not all people are covered by the government
drug plan. This can results in non-differentially bias through improving the
estimate in the group with coverage but not in the group of individuals without
coverage.

The high specificity of administrative case definitions cannot be under
appreciated as it contributes to a low false positive rate and high PPV, thus
reducing the potential for overestimating prevalence. A PPV above 70% has
been deemed sufficient for surveillance of other health outcomes (e.g.,
cerebrovascular accidents, congestive heart failure and
venothromboembolism) (41). We demonstrated that the PPV of the NDSS
case definition is generally higher than 70% assuming true diabetes prevalence
is > 5%. If diabetes prevalence is < 5%, over a third of diabetes cases may in
fact be falsely identified as diabetes cases. Conversely, a prevalence > 10%
reduces the false positive rate which renders the NDSS case definition more
efficient. In this situation, however, the NDSS case definition could
underestimate prevalence if the proportion of false negative cases exceeds the
proportion of false positive cases. Above all, the choice of administrative
Prevalence of diabetes

definition depends on the underlying prevalence of the disease and the goals of the surveillance system that might warrant maximizing the sensitivity at the expense of some specificity and PPV.

Sudden or marked changes in diabetes prevalence should prompt a re-validation of the test accuracy of the case definition (39). This is a priority especially if the case definition has not been systematically validated before. Thereafter, yearly change in diabetes prevalence can be adjusted and better quantified through applying correction factors derived from the test accuracies of the case definition. From fiscal year 2002/3 to 2006/7, the NDSS case definition underestimated the rise in diabetes prevalence in Canada by approximately 0.4% (78,625 diabetes cases) over the 5-year period (Supplemental file 4). The importance of applying correction factors grows over time as the bias appears to rise with increasing diabetes prevalence.

Administrative case definitions generally capture advanced physician-diagnosed diabetes cases and frequent users of health services. While estimating the prevalence of these cases is important for health economics and manpower distribution, infrequent users of health services and individuals with diabetes that have not been brought to medical attention are likely to have a disproportionate amount of downstream diabetes-related complications and attendant healthcare utilization. Indeed, none of the included validation studies accounted for undiagnosed diabetes. It has been previously estimated that approximately one third of diabetes cases remain undiagnosed (42-44). Accounting for undiagnosed diabetes not only increases the prevalence of diabetes considerably but also steepens the increase in diabetes prevalence over time as shown in Figure 4.

Reference standards

There are potential limitations for all reference standards used to validate administrative definitions for diabetes. The accuracy of primary care charts reviews depends largely on physician charting, availability of records, and the accurate interpretation of medical data during the review process. Medical chart reviews miss cases in the general population if diabetes
Prevalence of diabetes

screening is not routinely performed on every patient in the primary care setting. Poor participation by physicians also introduces bias, as physicians who agree to participate may have a keener interest in diabetes care, more thorough diabetes evaluations and follow-ups for patients in their practice and/or clearer medical charting.

In our study, the sensitivity of the NDSS case definition was shown to be generally higher when compared to medical records than to surveys. However, surveys also have potential limitations. Information bias could be introduced by patients’ poor recall, social desirability bias, poor understanding of survey questions, or incomplete knowledge of their diagnoses. Surveys can also suffer from participation biases that influences estimates towards or away from the null. Some asymptomatic individuals with low diabetes risk may be less willing to participate whereas certain patients with advance diabetes may be too unwell to participate. The validity of self-report from surveys is also affected by socio-demographic characteristics. The extremes of age are more likely to report having diabetes (45; 46) and the effect of sex could influence reporting in either direction (45-47). Both lower education(48) and poorly-controlled diabetes have been found to be associated with underreporting(49).

We acknowledge that the correction factors proposed herein were based on the premise that medical chart reviews and population-based surveys had perfect sensitivity and specificity. In the absence of a “gold standard” for validating administrative algorithms (50), Bayesian statistical approaches, that incorporate the uncertainties of non-gold standard case ascertainment techniques, could be undertaken to estimate the true population prevalence(51). Alternatively, a thorough assessment of sensitivity measurements obtained via different reference standards can be performed to corroborate prevalence estimates and surveillance results from administrative data (39; 52).

**Strengths and limitations**

This is the first systematic review on validation studies on the NDSS criteria from administrative data. Our systematic review was comprehensive as
it had a broad search strategy that bore no language or time restriction. During the review for selection, foreign language articles were partially translated by colleagues who were native speakers of these languages. The study selection during the abstract review was performed by two independent reviewers and discrepancies were adjudicated by a third reviewer. It was likely that only a small number of relevant articles were missed by our search strategy which was generic and based on the intercept of only a few keywords. The bibliographies of included studies were also perused. While only two major electronic databases (Medline and Embase) were examined, it was felt that other searches engines, such as Cochrane Central Register of Controlled Trials (CENTRAL), would unlikely yield any study of interest given that validation studies are not designed as randomized controlled trials. As validation studies generally require authors to have access to national administrative databases and surveys, it is also unlikely that additional studies would have been found from grey literature. The inclusion of unpublished studies might arguably reduce publication bias but expose the review to lower quality studies that potentially lack rigorous statistical techniques of published studies (53). All 11 included studies captured patient information at the population level with clear case definitions, were validated by reference standards encompassing a broad spectrum of patients and had QUADAS scores over 10. These studies were funded by large research agencies and academic centres (Supplemental file 5) with no reported disclosures from the private sector or special interest groups.

The heterogeneity observed in the meta-analysis likely arose from different reference standards used. Therefore, a random-effects bivariate regression model that accounted for heterogeneity and correlation between sensitivity and specificity was used to pool the test measure estimates. Other potential sources of residual heterogeneity are differences in socio-demographic characteristics, geographical location, year of study, health insurance arrangements, physician remuneration schemes, prescription subsidies and healthcare utilization, practices and access. These study characteristics were not explored in a meta-regression or subgroup analysis because of the small number of studies. Heterogeneity could also result from misclassification due to unmeasured confounders, such as human error in
Prevalence of diabetes

physician claims and hospitalizations coding. However, this was unlikely as the administrative databases studied have been previously validated and used widely for research studies and surveillance efforts.

**Generalizability**

As all included studies were conducted in North America, we assumed that the study bases were similar enough to make direct comparisons between studies. We found generally good concordance (kappa statistic > 0.7) between the administrative definition with medical records and the administrative definition with population-based surveys across studies, suggesting that public administrative data are a viable substitute for these other case ascertainment methods. Given that administrative data conveniently encompasses the entire population in identifying diabetes cases, it is particularly efficient for national surveillance. Indeed, maintaining a nationwide diabetes registry is expensive for a chronic disease as prevalent as diabetes.

However, while study bases were nested in the general population, the selected study samples were not always random and, thus, may not necessarily be representative of the total population. Mild variations in the statistical agreement between administrative data and medical records/surveys might be explained by differences in the constitution of the study bases. Higher concordance was reported between the administrative definition and medical records/surveys in the American studies which were conducted on well-defined populations (e.g., within a HMO). Conversely, slightly lower concordance was found in the Canadian studies that studied agreement between the NDSS case definition and self-report from surveys targeting the entire population via stratified random sampling.

Extrapolation of the pooled test measures of the NDSS criteria to other jurisdictions, with different healthcare systems, administrative databases, physician remuneration arrangements and patient populations, demands caution. This also highlights the need for jurisdictions to periodically evaluate the test accuracies of administrative algorithms on new populations. As the stability of sensitivity measurements is essential to monitor disease trends over
time, validation studies should be repeatedly performed at different time point within the same population to adjust prevalence estimates accordingly.

In sum, claims-based administrative algorithms are widely used across North America and play a vital role in Canadian diabetes surveillance strategies. Thus, establishing the criterion validity of the NDSS case definition is critical for healthcare professionals and public health researchers. We have shown that the NDSS case definition has an acceptable sensitivity and a reasonably high specificity for diabetes surveillance. By applying correction factors to reported diabetes prevalence from Canadian surveillance reports, we demonstrate that the NDSS case definition overestimates prevalence when it is \( \leq 10.6\% \) and does the converse when prevalence is \( > 10.6\% \); hence, correction factors can be applied to make proper quantifications of yearly prevalence. Even with the use of correction factors to account for the NDSS test accuracies, the administrative algorithm probably misses new or mild diabetes and is unable to identify undiagnosed diabetes cases. It does, however, capture advanced physician-diagnosed diabetes cases and frequent users of healthcare services. Estimating the population prevalence of these diabetes cases is important for health services, health economics, and budget and manpower allocation.
Prevalence of diabetes

Figure 1: Flow diagram of selection strategy and article reviews

Flow diagram is in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).
Table 1: Test properties of the NDSS case definition

<table>
<thead>
<tr>
<th>Study population</th>
<th>Study Years</th>
<th>Author</th>
<th>Case definition algorithm</th>
<th>“Gold standard”/ comparator</th>
<th>Age (years)</th>
<th>N</th>
<th>Sens (%)</th>
<th>Spec (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontario, Canada</td>
<td>1992-1999</td>
<td>Hux (1)</td>
<td>Prescription data in ODB</td>
<td></td>
<td>≥65</td>
<td>-</td>
<td>91.0*</td>
<td>(81.0, 89.0)</td>
<td>64.0</td>
<td>(59.0, 69.0)</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Self-report in NPHS</td>
<td></td>
<td>≥20</td>
<td>4691</td>
<td>85.0†</td>
<td>(81.0, 89.0)</td>
<td>79.8</td>
<td>(76.0, 83.6)</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Medical records from GP offices</td>
<td></td>
<td>≥20</td>
<td>3317</td>
<td>86.1</td>
<td>(82.0, 90.2)</td>
<td>97.1</td>
<td>(96.5, 97.7)</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>2006-2008</td>
<td>Harris (6)</td>
<td>Medical records from EMR</td>
<td></td>
<td>Mean 42.5</td>
<td>19442</td>
<td>84.3*</td>
<td>(82.7, 86.3)</td>
<td>97.4*</td>
<td>(73.0, 76.6)</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>2000-2001</td>
<td>Shah (35)</td>
<td>NDSS case definition from CCHS</td>
<td></td>
<td>≥18</td>
<td>1812</td>
<td>74.8*</td>
<td>(72.8, 76.8)</td>
<td>98.2*</td>
<td>(98.0, 98.4)</td>
<td>0.77</td>
</tr>
<tr>
<td>Manitoba, Canada</td>
<td>1997-2002</td>
<td>Lix (37)</td>
<td>1) physician claims; 2) Hospitalization; 3) Prescription data from MHSIP</td>
<td></td>
<td>≥19</td>
<td>5589</td>
<td>75.5*</td>
<td>(69.2, 81.8)</td>
<td>97.8*</td>
<td>(66.1, 78.7)</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>1989-1990</td>
<td>Robinson (4)</td>
<td>1) or 3 physician claim or 1 hospitalization from MSHC (a)</td>
<td></td>
<td>Mean 18-74</td>
<td>2651</td>
<td>75.5*</td>
<td>(69.2, 81.8)</td>
<td>97.8*</td>
<td>(66.1, 78.7)</td>
<td>0.72</td>
</tr>
<tr>
<td>Saskatchewan, Canada</td>
<td>1991-2001</td>
<td>Koleba (14)</td>
<td>Prescription claims data</td>
<td></td>
<td>≥20</td>
<td>145696</td>
<td>94.4 (94.2, 94.6)</td>
<td>96.5 (96.1, 96.9)</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alberta and British</td>
<td>2001-2004</td>
<td>Southern (38)</td>
<td>1) 2 physician claim or 1 hospitalization (b); 2) 1 physician claim or 1 hospitalization from AHS</td>
<td>Laboratory data from CLS (c)</td>
<td>-</td>
<td>25419</td>
<td>79.1 (78.9, 79.4)</td>
<td>75.1 (74.8, 75.3)</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Columbia, Canada</td>
<td>2000-2002</td>
<td>Chen (3)</td>
<td>1) 2 physician claims or 1 hospitalization (d); 2) 2 physician claims; 3) 1 physician claim or 1 hospitalization; 4) 1 physician claim</td>
<td>Medical record from GP offices</td>
<td>Mean 35</td>
<td>3362</td>
<td>92.3 (89.2, 95.5)</td>
<td>96.9 (96.2, 97.5)</td>
<td>77.2</td>
<td>(99.0, 99.6)</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>Solberg (34)</td>
<td>2 physician codes or 1 hospitalization in 1 year OR 1 anti-diabetic medication from HEDIS</td>
<td>Prescription data, laboratory data and medical records</td>
<td>≥19</td>
<td>135842</td>
<td>96.5 (96.1, 96.9)</td>
<td>96.5 (96.1, 96.9)</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnesota, USA</td>
<td>1992-1995</td>
<td>O’Connor (5)</td>
<td>2 outpatient physician codes from HMO</td>
<td>Self-report in HMO survey</td>
<td>Adults Mean 39.9</td>
<td>1976</td>
<td>76.1* (68.1, 84.1)</td>
<td>99.6* (99.3, 99.9)</td>
<td>92.2* (86.7, 97.7)</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hebert (2)</td>
<td>14 algorithms that varied between 1 and 2 years, 1 or 2 Medicare claims</td>
<td>Self-report in MCBS</td>
<td>≥65</td>
<td>7562</td>
<td>74.4 (71.9, 76.9)</td>
<td>97.5 (97.1, 97.9)</td>
<td>84.1</td>
<td>(81.8, 86.4)</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Prevalence of diabetes
Prevalence of diabetes

AHS: Alberta Health Services; CCHS: Canadian Community Health Survey; CLS: Calgary Laboratory Services; EMR: Electronic Medical Records; GP: General Practitioner; HEDIS: Health Plan Employer Data and Information Set; HMO: Health Maintenance Organization; Kappa: Kappa statistic; MCBS: Medicare Current Beneficiary Survey; MHSC: Manitoba Health Services Commission; MHHP: Manitoba Heart Health Project; MHSIP: Manitoba Health Services Insurance Plan; NDSS: National Diabetes Surveillance System; NPHS: National Population Health Survey; NPV: Negative predictive value; ODB: Ontario Drug Benefit; ODD: Ontario Diabetes Database; OHI: Ontario health Insurance Plan; Sens: Sensitivity; SHB: Saskatchewan health beneficiaries; Spec: Specificity; PPV: Positive predictive value

\(^1\) Included in meta-analysis

\(^2\) Test estimates and 95% confidence intervals which were not reported in the original paper were manually calculated from available raw data

\(^3\) 95% confidence intervals were estimated based on a diabetes prevalence of 6.8% via the ODD and a sample size of 4691 from the NPHS cycle 3 1998/9

\(^4\) Test properties were recalculated to designate self-reported diabetes from survey as the reference standard

\(^5\) The sample size of the ODB were not explicitly stated in the paper. Hence, the 95% CIs could not be calculated

- Demographic data was not reported in the study

\(^6\) The authors validated one physician claim or one hospitalization which is similar but not identical to the NDSS criteria.

\(^7\) Test measures of the case algorithm closest to the NDSS criteria is displayed in the table

\(^8\) Test measures were calculated using the reference standard closest to current diabetes diagnosis criteria (i.e. CDA guidelines for diagnostic criteria of diabetes OR glycated hemoglobin \(\geq\) 6.7%) was displayed in the table

\(^9\) The sensitivity was calculated by projecting the case-control design (n=145 696) to the entire Saskatchewan population where 625 994 individuals would not have met the NDSS criteria of which 3443 would have been identified as having diabetes based in medication alone.
**Figure 2: Forest plots of sensitivities and specificities of the NDSS case definition reported by included validation studies**

<table>
<thead>
<tr>
<th>Reported sensitivities of the NDSS case definition</th>
<th>Reported specificities of the NDSS case definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study ID</td>
<td>ES (95% CI)</td>
</tr>
<tr>
<td>Charts</td>
<td>92.31 (89.15, 95.47)</td>
</tr>
<tr>
<td>Chen</td>
<td>86.12 (82.68, 89.55)</td>
</tr>
<tr>
<td>Hux</td>
<td>84.34 (82.72, 85.96)</td>
</tr>
<tr>
<td>Harris</td>
<td>74.38 (71.85, 76.92)</td>
</tr>
<tr>
<td>Survey</td>
<td>76.15 (68.15, 84.15)</td>
</tr>
<tr>
<td>Hebert</td>
<td>75.53 (69.39, 81.68)</td>
</tr>
<tr>
<td>O’ Connor</td>
<td>7.52 (63.49, 76.68)</td>
</tr>
<tr>
<td>Robinson</td>
<td>99.63 (99.35, 99.90)</td>
</tr>
<tr>
<td>Survey</td>
<td>97.81 (97.23, 98.39)</td>
</tr>
<tr>
<td>Hebert</td>
<td>97.51 (97.13, 97.89)</td>
</tr>
<tr>
<td>O’ Connor</td>
<td>96.96 (96.35, 97.56)</td>
</tr>
<tr>
<td>Harris</td>
<td>97.10 (96.49, 97.71)</td>
</tr>
<tr>
<td>Survey</td>
<td>96.67 (96.61, 97.13)</td>
</tr>
</tbody>
</table>

ES (95%CI): Summary estimate (95% confidence interval); Charts: Reference standard by medical chart review; Survey: Reference standard by patient self-report from population-based survey
Prevalence of diabetes

Figure 3: Random-effects bivariate regression analysis of the pooled test accuracies from 6 studies

The Hierarchical Summary Receiver Operator Characteristics (HSROC) curve displays the 95% confidence interval of the summary operating point and the 95% prediction region, which is the confidence region for a forecast of the true sensitivity and specificity in a future study. The shape of the prediction region is generated based on the assumption of a bivariate normal distribution for the random effects model. The Empirical Bayes estimate gives the best estimate of the true sensitivity and specificity of each study and these estimates will be shrunk towards the summary point compared with the study-specific estimates. The stronger the shrinkage, the greater the precision of the test estimate. The random-effects bivariate regression analysis could not be done for the subgroups stratified by validation method because the small number of studies.
Crude prevalence: prevalence of diabetes in Canada for fiscal years 2002/3 through 2006/7 obtained from the NDSS 2009 report (28); Adjusted prevalence: prevalence after applying correction factors \([\text{Prevalence(\%)} - 2.1 / 0.802]\); The margins of error for all adjusted prevalence and crude prevalence estimates were \(-0.01\%\) (n~25000000). Projected crude prevalence: future prevalence assuming an increase of 0.4\% per year; Adjusted projected prevalence: future prevalence after applying correction factors; Total diabetes: Estimated prevalence of physician-diagnosed and undiagnosed diabetes assuming 1/3 of total diabetes is undiagnosed. The crossover point of the crude and adjusted prevalence lines is \(-10.6\%\) around year 2013.
CHAPTER 1 CONCLUSION

Through systematic review and meta-analysis, we determined that the NDSS case definition has an acceptable sensitivity [random-effects bivariate regression pooled estimate; 82.3% (95%CI 75.8, 87.4)] and a high specificity [97.9% (95%CI 96.5, 98.8)] for diabetes surveillance. The administrative definition appears efficient in identifying physician-diagnosed cases that come in frequent contact with health services which is important for manpower allocation and budgetary implications. As misclassification is likely non-differential over time, the NDSS case definition can be particularly useful in evaluating prevalence trends. Assuming sensitivity and specificity of the administrative data based definition remain stable, changes in prevalence can be tracked over all observed time points.

Accuracy of alternative administrative algorithms

We additionally explored the accuracy of alternative administrative definitions to identify factors that may improve sensitivity, specificity or statistical agreement with reference standards. Several studies have compared the accuracy of algorithms based on the presence of two physician claims with algorithms based on a single physician claim; others have compared the accuracy of algorithms with and without hospitalization records.

Southern and colleagues found a higher sensitivity [86.4% (95%CI 86.2, 86.7)] when the liberal case definition (one claim or one hospitalization) was compared to the stricter NDSS case definition (one claims or one hospitalisation within 2 years) [78.5% (95%CI 78.3, 78.8)] validated against laboratory data (38). Chen and colleagues reported that the sensitivity increased considerably from 92.0% to 96.0% but the specificity decreased from 97.0% to 93.0% when one claim or one hospitalization was used in the algorithm instead of the full NDSS case definition (3). Similarly, Hebert and colleagues reported a higher sensitivity (79.0% vs. 74.0%) when one physician claim instead of two was used in the definition alongside a lower specificity (94.0% vs. 98.0%) (2). The single-claim definition had a lower sensitivity (92.4%) than the two-claims definition (97.1%) test properties from Hux and...
colleagues’ original data were back calculated (1). Lix and colleagues investigated the concordance between self-reported diabetes in the CCHS and administrative case definitions that varied in the number of physician billing claims, time span and inclusion of hospitalization data or prescription (kappa; 0.65 – 0.79) (37). The investigators concluded that one service claims had slightly lower statistical agreement with self-reported diabetes than two service claims. The same study demonstrated that removing hospitalization data from the algorithm decreased the kappa statistic slightly. Similar trends were found in the study by Hebert and colleagues (2). Robinson and colleagues described slightly lower kappa values when more physician contacts were required for diagnosis (0.72 for one vs. 0.71 for two vs. 0.68 for three).(51)

The validation study by Chen and colleagues varied the number of physician contacts and gap between claims against survey data. Changing the gap between claims from 1, 2 to 3 years did not change the sensitivity, specificity or kappa values significantly (3). In the study by Hebert and colleagues, varying the identification period from one to two years raised the sensitivity (63.0% vs. 74.0% respectively) while maintaining a robust specificity (99.0% vs. 98.0% respectively) (2).

In sum, single physician claim definitions have improved sensitivity; however, the trade-off in specificity leads to an unacceptable number of false positives and slightly poorer statistical agreement with reference standards. Some evidence support the incorporation of hospitalization records to improve the sensitivity and statistical agreement of claims-based algorithms with reference standards. Hence, continuing to advocate two physician claims and/or hospitalization for diabetes over a two-year period (i.e. the NDSS case definition) to prevent an overestimation of diabetes prevalence seems reasonable. The high specificity may be particularly important in evaluating health service delivery and establishing diabetes cohorts from administrative databases for associative or etiologic research.

Regardless of the case definition used, prevalence estimation should be adjusted for the test measure using the following correction formula:
Prevalence = \[\text{Proportion of positive cases} - (1-\text{specificity})\] / (\text{specificity} + \text{sensitivity}-1).

**Validation of the NDSS case definition in Quebec (Chapter 2)**

To our knowledge, there has not been a published study that validates the NDSS case definition in the province of Quebec. Quebec is one of two largest provinces in Canada comprising a large French Canadian population and a growing number of new immigrants from around the world. Socio-demographic factors and differences in healthcare delivery in this province may influence the accuracy of the NDSS case definition in case ascertainment. Thus, it is important to establish the test performance of the NDSS case definition within the Quebec context in order for public health researchers to make proper inferences from diabetes surveillance data.

While estimating the prevalence of these cases is important for health economics and manpower distribution, infrequent users of health services and individuals with diabetes that has not been brought to medical attention can have eventual downstream diabetes-related complications and attendant healthcare utilization. Indeed, none of the included validation studies accounted for undiagnosed diabetes. While validation studies on diabetes claims-based algorithms can be used to measure the prevalence of physician-diagnosed diabetes, undiagnosed diabetes has not been systematically assessed. As individuals with undiagnosed diabetes present to health services with advanced disease and diabetes-related complications, they can put a substantial strain on the Canadian healthcare system (54). Indeed, undiagnosed diabetes has the potential to contribute considerably to the burden of disease over time. Diabetes screening strategies aimed at early detection in asymptomatic populations, particularly those with an elevated diabetes risk, can afford healthcare providers and patients more time to enforce lifestyle modifications and initiate diabetes treatment before the onset of complications.

In chapter 2, we aim to estimate the total diabetes (physician-diagnosed and undiagnosed diabetes) prevalence in Quebec. First, we used a population-based health survey to validate the NDSS case definition from
Quebec administrative data and determine its test performance. Second, we determined the prevalence of undiagnosed diabetes using a novel diabetes screening approach involving mailed-in capillary blood samples. We then corrected prevalence estimates of diabetes by applying correction factors derived from Quebec-specific NDSS test accuracies to account for both physician-diagnosed and undiagnosed diabetes. We discussed the strengths and limitations of the mailed-in blood sampling in diabetes surveillance, prevalence estimation and diabetes screening. We also proposed strategies in which this screening test can be incorporated with existing diabetes screening practices to improve diabetes prevention.
CHAPTER 2: Estimating the population prevalence of diagnosed and undiagnosed diabetes

Aaron Leong¹, MD; Kaberi Dasgupta¹,², MD, MSc; Jean-Louis Chiasson³, MD and Elham Rahme¹,², PhD

¹Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada
²Department of Medicine, McGill University, Montreal, Quebec, Canada
³Department of Medicine, Université de Montréal and Centre de recherche du Centre Hospitalier de l’Université de Montréal, Montreal, Quebec, Canada

Short running title: Population prevalence of diabetes

Corresponding author: Elham Rahme, Division of Clinical Epidemiology, McGill University Health Centre, Division of Clinical Epidemiology, 687 Pine Avenue West, V Building. Montreal, Quebec H3A 1A1; tel: (514) 934-1934, ext. 44724; fax: (514) 934-8293; email: elham.rahme@mcgill.ca.

Word count: 3,695 words
ABSTRACT

Objective: Health administrative data are frequently used for diabetes surveillance but validation studies are limited and undiagnosed diabetes has not been considered in previous studies. We compared the test properties of an administrative definition with self-reported diabetes and estimated prevalence of undiagnosed diabetes by measuring glucose levels on mailed-in capillary blood samples.

Research Design and Methods: A stratified random sample of 6,247 individuals (Quebec province) was surveyed by telephone and asked to mail-in fasting blood samples on filter paper to a central laboratory. An administrative definition was applied (two physician claims or one hospitalization for diabetes, within a two-year period) and compared with self-reported diabetes alone, and with self-reported diabetes or elevated blood glucose level ($\geq 7$ mmol/l). Population-level prevalence was estimated using the administrative definition, corrected for its sensitivity and specificity.

Results: Compared to self-reported diabetes, sensitivity and specificity were 84.3% (95%CI 79.3, 88.5) and 97.9% (95%CI 97.4, 98.4), respectively. Compared to diabetes by self-report and/or glucose testing, sensitivity was lower at 58.2% (95%CI 52.2, 64.6) while specificity was similar at 98.7% (95%CI 98.0, 99.3). Adjusted for sampling weights, population-level prevalence of physician-diagnosed diabetes was 7.2% (95%CI 6.3, 8.0). Prevalence of total diabetes (physician-diagnosed and undiagnosed) was 13.4% (95%CI 11.7, 15.0), indicating that approximately 40% of diabetes cases are undiagnosed.

Conclusions: A substantial proportion of diabetes cases are missed by surveillance methods using health administrative databases. This is concerning as individuals with undiagnosed diabetes are likely to have a delay in treatment and thus a higher risk for diabetes-related complications.
The rapid rise in diabetes incidence and prevalence is placing substantial strains on health care systems in terms of management both of the disease itself and its complications (7). Accurate disease surveillance is therefore critical in order to make proper projections of healthcare needs and costs. In Canada, the National Diabetes Surveillance System (NDSS) tracks diabetes prevalence using physician billing and hospital admissions databases (1; 40; 55). Similar algorithms have been used in the United States and other countries (2; 4; 34; 40; 56-60). While these administrative health database definitions have been validated using medical records (1; 3), surveys (1; 2; 4; 5; 37), and medication utilization data (1; 14; 34), there has not been a validation study of claims-based administrative algorithms to account for previously-undiagnosed diabetes.

In the present study, we validated an administrative database diabetes definition and estimated the prevalence of physician-diagnosed diabetes using a population-based health survey. We estimated the prevalence of undiagnosed diabetes using a novel approach involving mailed-in capillary blood samples on which fasting glucose levels were measured at a central laboratory. Finally, using survey data and glucose levels to ‘correct’ the administrative database diabetes definition, we estimated the actual diabetes prevalence and compared this to what would have been estimated by the uncorrected definition.

**RESEARCH DESIGN AND METHODS**

The source population included residents of the Canadian province of Quebec. The Régie de l’assurance maladie du Québec (RAMQ), the government body that administers the public insurance plan for physician services and hospitalization costs, generated a stratified random sample (n = 6,247) for this study, over-sampling less populated regions (61). These individuals were surveyed by telephone or mail and asked to mail-in a self-collected capillary blood sample for glucose testing, as described below. Study procedures were approved by the Commission d’accès à l’information du Québec and the Institutional Review Board of the McGill University Health Centre.
The survey questionnaire was administered by trained Quebec Statistical Institute (QSI) staff from March 31 to July 14, 2009, including weekdays and weekends, daytime and evening by telephone. Prior to the telephone survey, a letter describing the study was sent, with an option to respond to the survey by mail. Participants were queried about diabetes with the following survey question: “Have you ever been told by a doctor or another health professional that you had diabetes?” Other data collected included socio-demographic factors, family history of diabetes, engagement in regular exercise, smoking, height and weight, and use of health care services. QSI survey questions were based on those of the Canadian Community Health Survey (62).

Participants received capillary blood sampling instructions (Supplemental figure S1), lancing materials and a specially-printed Whatman No. 3 filter paper that included two drops of quality control solution. Following an overnight fast, they were asked to clean their fingertip with an alcohol swab, puncture it with a lancet, and squeeze a drop of blood onto two circles printed on the filter paper. They were then asked to post the filter paper in the stamped and addressed envelope provided.

The blood samples were analyzed for glucose measurements at the central laboratory at Saint-Luc Hospital in Montreal. The centres of the blood spots were cut out using a hand-held 6 mm punch and these were placed in tubes filled with 400 microlitres of 2.5% trichloroacetic acid solution. The tubes were shaken manually at 10-minute intervals for one hour at room temperature and then centrifuged at 3,000 rpm for 10 minutes. Supernatants were transferred into sampling cups and immediately analyzed by observing the reaction with hexokinase. The eluate:reagent ratio was set at 1:11, and the reaction was monitored bichromatically (340 and 380 nm) at 37°C for six minutes. The results were recorded from the calibration curve established according to standards prepared on the date the filter papers were issued to subjects. As glucose values may have a maximum decay of 20% over time, the results were adjusted according to the value of the internal standard and the time elapsed between the dates of blood sampling and the date of measurement (63). This filter paper technique has been shown to have a
Coefficient of Variance of 3.6% within assays and CV of 4.2% between assays. Correlation with ordinary whole blood glucose dehydrogenase method is good \( r = 0.98 \) (64).

Survey and glucose results were linked to health administrative data for the period 1 January 1997 to 31 December 2009. The research team received no nominal information. We examined baseline characteristics of the entire sample, survey respondents, and the subgroup providing glucose samples. Variables include a socioeconomic (SES) measure, with sub-indices of social and material deprivation, that was developed by the Institut national de santé publique du Québec (INSPQ) based on Census Enumeration area data on education level, employment/population ratio and average income (65; 66). This SES measure is associated with higher risk of stroke mortality, mortality following myocardial infarction and disability in diabetes (67).

Within the survey sample, we compared the NDSS diabetes case definition of two or more physical billings for diabetes and/or one or more hospitalizations for diabetes within a two-year period from administrative databases with self-reported diabetes from the survey as the reference standard. Within the subgroup with glucose values, we compared the NDSS definition with a diabetes definition that included self-report and/or elevated fasting glucose. We selected a threshold of 7 mmol/l to define diabetes; this threshold is aligned with current clinical practice guidelines for diabetes diagnosis by fasting plasma glucose (68-70). While glucose measurements on capillary whole can be up to 15% lower than plasma due to the influence of hematocrit, the difference varies considerably (71). Thus we opted for this conservative cut-off value to avoid overestimating prevalence.

For these comparisons, we calculated the Kappa statistic (Fleiss’ formula), sensitivity, specificity, and positive and negative predictive values. We then applied the NDSS definition to the entire random sample and used the sensitivity and specificity estimates to adjust prevalence estimates. The correction formula used was the following: \( \frac{\text{Proportion of positive NDSS cases} - (1 - \text{specificity})}{\text{sensitivity} + \text{specificity} - 1} \) (72). Finally, we
extrapolated diabetes prevalence estimates from the sample to the Quebec population using appropriate sampling weights as provided by the QSI.

As an alternative approach, we imputed self-reported diabetes and glucose values for those in the stratified random sample who had not completed the survey, based on logistic regression models derived from RAMQ baseline characteristics. We used the Bayes posterior predictive distribution of the missing values from a non-informative prior to account for variation due to not being observed. Again, we extrapolated to the Quebec population using sampling weights.

The RAMQ data linkage with the QSI survey data was executed and retrieved using the statistical software SAS (version 9.2). Subsequent statistical analyses were performed with the statistical software STATA (version 11).

RESULTS

Stratified random sample

Among the 6,247 individuals from the sample, at the time of the study, 12 individuals had moved out of the province of Quebec, 10 were deceased and one was under the age of 20 years. Of the original random sample, 33.9% either had an incorrect telephone number or did not have a listed telephone number; 9.6% did not respond to telephone calls and 56.1% (n = 3,504) were contacted by telephone. Among individuals contacted by phone, 83.9% (n = 2,940; 47.1% of the original sample) completed the telephone-administered questionnaire. We received an additional 566 (9.1% of the original random sample) questionnaires completed by mail. The final response rate was thus 56.1%, comprising 3,506 survey respondents among whom 95.8% (n = 3,322; 53.2% of original sample) agreed to record linkage between survey and health administrative data. A total of 1,829 participants (29.3% of original sample; 52.2% of survey respondents) provided mailed-in blood samples of which 89.1% (n = 1,629; 26.1% of original random sample) were analysable. Non-analysable samples were mainly a result of inadequate quantities of blood (Figure 1).
Prevalence of diabetes

Individuals in the stratified random sample had a mean age of 49.7 years (SD 16.4) and were equally distributed between men and women (Table 1). The proportion of NDSS cases was highest among survey respondents who provided analyzable blood samples [10.3% (95%CI 8.8, 11.8)], followed by survey respondents as a whole [8.5% (95%CI 7.6, 9.4)] and the lowest in the original random sample [7.5% (95%CI 6.8, 8.2)]. Survey respondents and individuals who provided analyzable blood samples were quite comparable to the original random sample for other baseline characteristics. Likewise, survey data on Body Mass Index (BMI), ethnicity, family history of diabetes and frequency of physician visits did not differ importantly.

Validation of the NDSS case definition

Among survey respondents, the NDSS and self-reported diabetes definitions were concordant (kappa statistic 0.79, 95% CI 0.76, 0.83, Table 2). Table 3 shows the sensitivity, specificity and Positive Predictive Value (PPV) of the NDSS case definition using self-reported diabetes as the reference standard. Sex-specific analyses suggested similar sensitivity [women; 81.4% (95%CI 73.0, 88.1); men; 86.5% (95%CI 79.9, 91.5)], PPV [women 75.4% (95%CI 66.8, 82.8); men 79.5% (95%CI 72.4, 85.5)], and specificity [women; 98.2% (95%CI 97.4, 98.8); men; 97.7% (95%CI 96.7, 98.4)] for women and men. (When the self-report and/or elevated glucose definition was used as the gold standard, concordance with the NDSS definition was lower [kappa statistic 0.67 (95%CI 0.62, 0.71); Table 2]. The prevalence of diabetes among men was higher than among women [women 7.7% (95%CI 5.9, 9.5) and men 13.3% (95%CI 10.9, 15.7) by NDSS criteria; women 11.6% (95%CI 9.4, 13.8) and men 18.5% (95%CI 15.7, 21.3) by self-report and/or elevated glucose level].

Diabetes prevalence within the random sample

The proportion of NDSS positive cases in the random sample was 7.5% [95% (CI 6.9, 8.2); women 6.9% (95%CI 6.0, 7.8); men 8.2% (95%CI 7.2, 9.2)]. Adjusted for test properties as defined by self-report, diabetes
Prevalence of diabetes

prevalence was 6.6% [(95%CI 6.0, 7.2; women 6.4% (95%CI 5.5, 7.3); men 6.9% (95%CI 6.0, 7.8)]; multiple imputation methods yielded an estimate of 6.6% (95%CI 5.9, 7.2). Adjusted for test properties as defined by self-report and/or elevated glucose level, the prevalence estimate was 10.8% (95% CI 10.1, 11.5); multiple imputations methods yielded an estimate of 11.2% (95% CI 10.4, 12.0).

Estimating the population prevalence of diabetes

After adjusting the weighted prevalence estimate using the test properties of the NDSS criteria derived from self-reported diabetes [7.7% (95% CI 6.8, 8.6)], the resulting population prevalence estimate for diagnosed diabetes and total diabetes were 6.8% (95% CI 5.7, 7.9) and 11.2% (95% CI 9.6, 12.8) respectively. When the prevalence estimate for undiagnosed diabetes in Quebec was added to the weighted prevalence of diabetes by self-report, the total diabetes prevalence in Quebec was computed to be 13.4% (95% CI 11.7, 15.0).

CONCLUSIONS

Principal findings

Our study demonstrates that a widely-used administrative database definition for physician-diagnosed diabetes has a high concordance with self-reported diabetes as identified through survey (kappa statistic 0.79; sensitivity 84.3%; specificity 97.9%). However, concordance is lower with diabetes identified by self-report and/or glucose testing with mailed in samples (kappa statistic 0.66; sensitivity 58.2%; specificity 98.7%). A substantial proportion of diabetes cases are captured neither by administrative data nor by self-report, as both rely on physician diagnosis. Accounting for the sampling weights of survey respondents, the 2009 Quebec population prevalence of physician-diagnosed diabetes identified by means of the administrative algorithm or self-report was approximately 7%. The prevalence rose to over 11% with the
inclusion of previously undiagnosed diabetes. Thus, approximately 40% of individuals with diabetes in the province of Quebec appear to be undiagnosed.

The 84.3% sensitivity estimate of the NDSS case definition for physician-diagnosed diabetes detected in our study is similar to the 85% reported by Hux and colleagues in Ontario (1). In this previous study, the NDSS case definition was compared with self-reported diabetes from the National Population Health Survey. The sensitivities of the NDSS case definition of these studies were somewhat higher than that reported in Manitoba (4) (76%; reference standard was self-reported diabetes from Manitoba Heart Health Survey) and Minnesota (2,5) (76% and 74%; reference standard was self-reported diabetes from a Health Maintenance Organization survey and a Medicare health beneficiaries survey respectively). Specificities of these studies were uniformly high (>97%).

Interestingly, we determined the point estimate of the NDSS sensitivity for physician-diagnosed diabetes to be slightly better for men than women. Variations in the diagnostic accuracy of the NDSS criteria across subgroups were also identified in an earlier Canadian study by Koleba and colleagues (14). It has been proposed that women are more likely to present with multiple medical complaints (73) and, therefore, diabetes is not as frequently coded as the primary diagnosis even if they have diagnosed diabetes. Heterogeneity in NDSS criteria accuracy among subgroups based on demographic characteristics has also been reported in other studies (74).

**Correction factors**

Under the assumption that self-reported diabetes is a “gold” standard for physician-diagnosed diabetes, the following correction factor could be applied to prevalence estimates derived from the NDSS case definition: 
(proportion of diabetes cases – 0.021) / 0.822. Importantly, however, the NDSS case definition was found to have a higher false positive rate when compared to self-report diabetes alone than to diabetes identified by self-report and/or glucose testing (22.3% vs. 10.3%). Some participants may have failed to report physician-diagnosed diabetes, but would have been captured by the
Prevalence of diabetes

NDSS criteria or through mailed-in samples if fasting glucose values remained elevated. Failure to report diabetes on surveys could result from lack of comprehension (e.g., perceiving that treatment ‘cures’ disease), or the stigma of having a chronic disease commonly associated with poor lifestyle habits. Despite this shortcoming, self-report from surveys is a suitable case ascertainment technique for diabetes surveillance of the whole population as it potentially covers individuals who do not come into regular contact with health services (4; 75; 76).

Alternative reference standards used to validate claims-based administrative algorithms, such as medication dispensation administrative data or primary care chart reviews, may not be representative of the general population (1). Quebec prescription databases are only populated by individuals who are covered by the public drug insurance plan. These include persons ≥ 65 years, recipients of last-resort financial assistance, marginalized social groups, the self-employed and individuals in the work force who do not have private drug insurance (1; 77). In addition, medication dispensation data do not capture patients who are not on pharmacologic therapy. With respect to primary care chart reviews, as not all patients undergo glucose testing, some diabetes cases may be missed. Even if medical chart reviews are conducted in randomly-selected family physician offices, results may not always be generalizable to the entire population. This underscores the potential utility of mailed-in blood samples in systematic glucose testing for population prevalence estimation.

In contrast to other validation studies of administrative algorithms, we measured fasting glucose levels on mailed-in blood samples to generate a new reference standard that took into account undiagnosed diabetes cases. We are thus able to propose the following correction factor to estimate the population prevalence of total diabetes from the NDSS case definition: (proportion of diabetes cases – 0.011) / 0.574. Hux and colleagues (1) reported a 1998 diabetes prevalence in Ontario of approximately 6.8% which was very similar to the physician-diagnosed diabetes population prevalence found in our study. By applying our proposed correction formula, the 1998 Ontario population
Prevalence of diabetes

Prevalence of both physician-diagnosed and undiagnosed diabetes would have, in fact, been closer to 9.6%.

Prevalence of undiagnosed diabetes

Despite differences in sampling frame and statistical techniques to correct for non-participation, our estimates for the prevalence of undiagnosed diabetes were comparable to the Manitoba Heart Health Survey (MHHS) in 1998 and the National Health Nutrition Examination Survey (NHANES) in 1999-2002 (42; 43). The MHHS (n = 2,792), was conducted on a stratified random sample of the Manitoba population wherein over 60% of the participants had fasting glucose measured on venous blood samples. NHANES 1999-2002 (n = 10,291) had fasting glucose samples drawn on a subsample without self-reported diabetes in the United States. In both these studies undiagnosed diabetes comprised about a third of diabetes cases (42; 43).

NHANES 2003-2006 (n = 14,611) performed not only fasting glucose measurements but also glucose tolerance testing and glycated hemoglobin on a subsample. Self-reported diabetes by survey was 7.8% of the general population and undiagnosed diabetes estimated by fasting glucose was 2.5%. Compared to NHANES 1999-2002, it appeared that physician-diagnosed diabetes had increased from 6.5% and undiagnosed diabetes, estimated solely on fasting glucose values, had fallen slightly from 2.8%. However, when all three diagnostic criteria were considered, the proportion of undiagnosed diabetes in NHANES 2003-2006 rose to 5.4%, specifically, 0.3% additional cases by glycated hemoglobin and 2.3% by oral glucose challenge (44).

Given that we only obtained fasting samples in our study, we did not capture individuals with isolated elevations in post-prandial glucose. Indeed, the Decode Study on diabetes prevalence of 13 European cohorts and NHANES 2003-2006 reported that a substantial proportion of undiagnosed cases were only detected via an oral glucose challenge (44; 78). Likewise, Zhang and colleagues (79) reported a modest sensitivity of 67%, coupled with a high specificity of 98%, for fasting glucose $\geq 6.7$ mmol/l as an initial screening test when compared to a reference standard that involved further
evaluation with oral glucose tolerance testing. This suggests that fasting glucose thresholds lower than 7 mmol/l may potentially capture some additional diabetes cases that would only have been detected following a 2-hour glucose challenge (79).

Fasting glucose or glycated hemoglobin measurements are generally preferred to an oral glucose challenge as initial screening tests because of the ease of administration, greater acceptability to patients and lower cost (80). Consequently, patients who present with isolated elevations in post-prandial glucose are often missed in diabetes screening. These cases, however, likely represent individuals with early diabetes who are not uniformly treated pharmacologically in clinical practice.

Some clinical practice guidelines recommend verifying abnormal screening test results, in the absence of unequivocal hyperglycemia, with a second glucose test, before making a clinical diagnosis (68-70; 81). We acknowledge that, in our study, a single mailed-in fasting blood sample has potential sources of measurement error. Despite detailed written instructions, individuals may not follow these completely (e.g., fasting period); there may be differences in glucose levels from the first drop to the second drop of capillary blood due to more extravascular fluid in the former; the time elapsed from sampling to receipt of the mailed-in sample could affect measurement precision and adjustments made to glucose values according to internal standards could have residual errors. However, Palardy and colleagues (63) have demonstrated good correlation between blood glucose measurements derived from filter paper and venous whole blood glucose. The filter paper collection method provides accurate and reproducible measurements of glycated haemoglobin (82; 83). We had not measured glycated hemoglobin on the mailed-in blood samples because elevated glycated hemoglobin levels had not yet been internationally adopted as a clinical diagnostic criterion at the time of our study formulation.

The World Health Organization in 1999 (84) proposed a threshold of 6.1 mmol/l on fasting whole blood glucose for diabetes diagnosis. This lower threshold could arguably be chosen for initial diabetes screening by fasting whole blood glucose to improve sensitivity, although recent clinical practice recommendations have not commented on the distinction between whole
blood and plasma glucose in diabetes screening (68). Notably, lowering the threshold to 6.1 mmol/l would effectively double the weighted prevalence of total diabetes from 13.4% (95%CI 9.6, 12.8) to 25.9% (95%CI 23.5, 28.3). Thus, we still elected to perform our analysis with a cut-off of 7 mmol/l to avoid overestimating the prevalence of total diabetes. Importantly, cases detected via such a screening strategy require confirmation through a more comprehensive diabetes evaluation with venous plasma blood glucose measurements. At the end of our study, individuals who had fasting glucose $\geq 6.1$ mmol/l were informed by mail to consult a physician for a clinical assessment.

**Utility of mailed-in blood samples for diabetes screening**

Mailed-in blood samples for glycated hemoglobin and/or fasting glucose measurements arguably offer a more convenient, feasible and cheaper approach for diabetes screening compared to in-person clinical evaluations. In addition to estimating the population-level prevalence of undiagnosed diabetes, this method could be used for diabetes surveillance of high-risk individuals at one to three-year intervals (e.g., metabolic syndrome, past history of gestational diabetes, family history of diabetes) (69; 80) or used in combination with self-administered risk assessment tools [e.g., Finnish Diabetes Risk Score (FINDRISC) and Canadian Diabetes Risk Questionnaire (CANRISK) (69; 85)]. Such a mailed-in blood collection strategy could be especially beneficial in medically underserviced areas and among patients unable or reluctant to participate in regular follow-up because of time constraints or competing responsibilities. As we detected a higher proportion of undiagnosed diabetes among men who provided an analyzable blood sample, mail-in blood sampling may be particularly effective for diabetes screening in men.

**Strengths and limitations**

We observed that a higher prevalence of physician-diagnosed diabetes cases among those who provided mailed-in blood samples suggesting that
higher-risk individuals were generally more inclined to participate. Detection of undiagnosed diabetes was somewhat limited because some individuals were unwilling or unable to provide a sample or provided an insufficient amount. Difficulties with comprehension and acting on the written instructions resulting from low levels of literacy and numeracy may have been a factor (86). Other potential contributing factors may have been poor dexterity or hand-eye coordination, visual impairment, physical or mental limitations, and fear of pain from sufficiently squeezing the pricked fingertip. Some individuals may doubt the utility of mailed-in capillary blood collection, be less concerned with developing diabetes, or be uncomfortable with unsupervised self-testing. We endeavoured to correct for potential selection bias by adjusting for sampling weights and performing multiple imputations based on the baseline information derived from administrative data that we had for all individuals, including non-respondents. Nonetheless, we acknowledge that there could be residual confounding from unmeasured confounders that differed between respondents and non-respondents.

Despite these potential limitations and others discussed, our findings indicate that in a predominantly single-payer environment, public administrative data provides a powerful resource for population-based evaluation of the burden of physician-diagnosed diabetes. However, the incorporation of home capillary blood glucose sampling provides evidence of glucose abnormalities in a substantial proportion of respondents, indicating that population-wide diabetes screening practices need to be improved. Mailed-in blood glucose measurements could be a powerful addition to surveillance strategies for high diabetes risk patients and help clinicians prioritize patient evaluations. Increasing the detection of diabetes offers the potential for instituting early management strategies to stem the anticipated tide of diabetes-related complications from longstanding, untreated diabetes.

ACKNOWLEDGEMENTS SECTION

Author contributions: AL conducted the literature search and summarized the literature, formulated the study question, performed the data-analysis and reported results, and took a leadership role in preparing all components of the manuscript and its revisions. KD is Co-Investigator of the
Prevalence of diabetes

CIHR grant; she contributed to the formulation of the study design, and was involved in all aspects of results reporting, manuscript preparation and revision. She played a major role in manuscript writing. JLC is Co-Investigator of the CIHR grant; he contributed to the formulation of the study design, led the collection of blood samples from all participants and the glucose testing in the central laboratories at Saint-Luc Hospital. ER is Principal Investigator of the CIHR grant; she led the formulation of the study design, the data-linkage between health administrative data, survey data and blood samples data and supervised all aspects of data analysis, results reporting, and manuscript preparation and revision.

**Guarantor:** ER is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Conflict-of-interest:** All authors of this manuscript, AL, KD, JLC and ER, have no relevant conflict of interests to disclose.

**Acknowledgements:** The authors thank the Institut de la Statistiques du Québec personnel who oversaw the survey: Lucille Pica, Brigitte Beauvais, Johanne Théroux, Nathalie Plante and Robert Courtemanche. The authors would also like to thank Lyne Labrecque for conducting the blood tests at Saint-Luc hospital and the Régie de l’Assurance Maladie du Québec for providing the administrative data.

**Disclaimer:** Elham Rahme is Associate Professor in the Department of Medicine of McGill University and holds a Senior Investigator award from the Fonds de Recherche en santé du Québec. Kaberi Dasgupta is Associate Professor of Medicine at McGill University and holds the Fonds de recherche Santé du Québec-Société québécoise d’hypertension artérielle-Jacques de Champlain Award. Jean-Louis Chiasson is Professor of Medicine at Université de Montréal. Aaron Leong is a Fellow in Endocrinology and Metabolism, and MSc candidate in Epidemiology at McGill University. The study was funded by a grant from the Canadian Institutes of Health Research.
Figure 1. Flow diagram of participants in QSI survey and home fasting blood glucose sampling.

Prevalence of diabetes

- Random sample from RAMQ register: 6,247
  - Did not fulfill eligibility: 23 (0.4%)

**Respondents**
- Total survey response: 3,506 (56.1%)
  - Completed survey by phone: 47.1%
  - Completed survey by mail: 9.1%

**Non-respondents**
- Total survey non-response: 2,718 (43.5%)
  - Incorrect or no listed telephone numbers: 33.9%
  - Did not respond to telephone call: 9.6%
  - Refused to complete survey by telephone: 9.0%

- Refused record linkage: 184

**Consented record-linkage**
- 3,322 (53.2%)
  - Did not agree to send a blood sample: 596
  - Agreed but failed to send a blood sample: 1,080
  - Blood samples not analyzable: 201

**Analyzable blood samples**
- 1,598 (46.5%)

**Self-reported no diabetes**
- 1,367 (85.5%)

**Self-reported no diabetes but previous abnormal blood glucose**
- 71 (4.4%)

**Self-reported diagnosed diabetes**
- 160 (10.0%)

**Undiagnosed diabetes**
- FPG ≥ 7 mmol/l
  - 77 (5.6%)

- No diabetes
  - < 7 mmol/l
    - 1,290 (94.4%)
Table 1. Selected baseline characteristics from RAMQ administrative databases and QSI survey item responses.

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>Entire stratified random sample</th>
<th>Survey respondents&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Participants who provided analyzable blood glucose samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients, n</td>
<td>6,247</td>
<td>3,322</td>
<td>1,598</td>
</tr>
<tr>
<td>Mean age in years (SD)</td>
<td>49.7 (16.4)</td>
<td>51.2 (15.1)</td>
<td>52.4 (14.4)</td>
</tr>
<tr>
<td>Diabetes by NDSS criteria, n (%)</td>
<td>469 (7.5)</td>
<td>283 (8.5)</td>
<td>165 (10.3)</td>
</tr>
<tr>
<td>Self-reported diabetes, n (%)</td>
<td>-</td>
<td>261 (7.9)</td>
<td>160 (10.0)</td>
</tr>
<tr>
<td>Women, n (%)</td>
<td>3,206 (51.3)</td>
<td>1,767 (53.2)</td>
<td>845 (52.9)</td>
</tr>
<tr>
<td>Rural residence, n (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1,018 (16.3)</td>
<td>560 (18.9)</td>
<td>290 (18.1)</td>
</tr>
<tr>
<td>Hospitalization in last year, n (%)</td>
<td>617 (9.9)</td>
<td>354 (10.7)</td>
<td>184 (11.5)</td>
</tr>
<tr>
<td>Hypertension in the last year, n (%)</td>
<td>608 (9.7)</td>
<td>334 (10.1)</td>
<td>175 (11.0)</td>
</tr>
<tr>
<td>Ischemic heart disease in the last year, n (%)</td>
<td>226 (3.6)</td>
<td>129 (3.9)</td>
<td>66 (4.1)</td>
</tr>
<tr>
<td>Heart failure in the last year, n (%)</td>
<td>140 (2.2)</td>
<td>69 (2.1)</td>
<td>32 (2.0)</td>
</tr>
<tr>
<td>Cancer in the last year, n (%)</td>
<td>247 (4.0)</td>
<td>134 (4.0)</td>
<td>72 (4.5)</td>
</tr>
<tr>
<td>Material Deprivation Index (SD)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.9 (1.5)</td>
<td>2.9 (1.5)</td>
<td>2.9 (1.5)</td>
</tr>
<tr>
<td>Social Deprivation Index (SD)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.0 (1.5)</td>
<td>2.9 (1.5)</td>
<td>2.8 (1.5)</td>
</tr>
<tr>
<td>Survey respondent, n (%)</td>
<td>3,322 (53.2)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mean BMI in kg/m&lt;sup&gt;2&lt;/sup&gt; (SD)</td>
<td>-</td>
<td>26.3 (5.1)</td>
<td>26.5 (5.4)</td>
</tr>
<tr>
<td>Family history of diabetes, n (%)</td>
<td>-</td>
<td>1,417 (42.7)</td>
<td>722 (45.2)</td>
</tr>
<tr>
<td>Non-Caucasian ethnicity, n (%)</td>
<td>-</td>
<td>1,204 (36.2)</td>
<td>523 (32.7)</td>
</tr>
<tr>
<td>Self-reported regular physician visits (annually), n (%)</td>
<td>-</td>
<td>2,593 (78.1)</td>
<td>1,269 (79.4)</td>
</tr>
</tbody>
</table>

<sup>a</sup>This group represents the 3,322 survey respondents who agreed to have their responses and biochemical data linked to RAMQ information.

<sup>b</sup>Residence status of 13.7% individuals in the random sample were missing.

<sup>c</sup>Material and Social Deprivation Indices are scored from 1 to 5 based on quintiles and means (SD) of the quintiles are reported.
Table 2. 2-by-2 tables for diabetes by NDSS criteria against (1) self-reported diabetes and (2) self-reported diabetes and/or undiagnosed diabetes (Total diabetes).

<table>
<thead>
<tr>
<th>Self-reported diabetes&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes by NDSS criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>220</td>
<td>63</td>
<td>283 (8.5%)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>No</td>
<td>41</td>
<td>2,998</td>
<td>3,039</td>
</tr>
<tr>
<td>Total</td>
<td>261 (7.9%)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3,061</td>
<td>3,322</td>
</tr>
</tbody>
</table>

Kappa 0.79 (0.76, 0.83)

<table>
<thead>
<tr>
<th>Self-reported diabetes and/or undiagnosed diabetes (Total diabetes)&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes by NDSS criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>148</td>
<td>17</td>
<td>165 (10.3%)&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>No</td>
<td>105</td>
<td>1,328</td>
<td>1,433</td>
</tr>
<tr>
<td>Total</td>
<td>253</td>
<td>1,345</td>
<td>1,598</td>
</tr>
</tbody>
</table>

(15.8%)<sup>f</sup>

Kappa 0.67 (0.62, 0.71)

---

<sup>a</sup> 2 by 2 table derived from the 3,322 individuals who responded to the survey
<sup>b</sup> Prevalence by NDSS criteria: 283/3322 = 8.5%
<sup>c</sup> Prevalence by self-report: 261/3322 = 7.7%
<sup>d</sup> 2 by 2 table derived from the 1,598 individuals who provided analyzable blood sample
<sup>e</sup> Prevalence by NDSS criteria: 165/1598 = 10.3%
<sup>f</sup> Prevalence by self-report and/or elevated fasting glucose: 253/1598 = 15.8%
### Table 3. Test properties of NDSS criteria for self-reported diabetes and total diabetes

<table>
<thead>
<tr>
<th>Test properties</th>
<th>Self-reported diabetes from survey $^a$ percent (95% CI)</th>
<th>Total diabetes $^b$ percent (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>3,322</td>
<td>1,598</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>84.3 (79.3, 88.5)</td>
<td>58.5 (52.2, 64.6)$^c$</td>
</tr>
<tr>
<td>Specificity</td>
<td>97.9 (97.4, 98.4)</td>
<td>98.7 (98.0, 99.3)$^c$</td>
</tr>
<tr>
<td>PPV</td>
<td>77.7 (72.4, 82.4)</td>
<td>89.7 (84.0, 93.9)</td>
</tr>
<tr>
<td>NPV</td>
<td>98.7 (98.2, 99.0)</td>
<td>92.7 (91.2, 94.0)</td>
</tr>
<tr>
<td>ROC area</td>
<td>91.1 (88.9, 93.3)</td>
<td>78.6 (75.6, 81.7)</td>
</tr>
<tr>
<td>Kappa statistic</td>
<td>79.2 (75.8, 82.6)</td>
<td>66.6 (61.9, 71.4)</td>
</tr>
<tr>
<td>Prevalence (NDSS)</td>
<td>8.5 (7.6, 9.4)</td>
<td>10.3 (9.5, 11.1)</td>
</tr>
<tr>
<td>Prevalence (self-report/self-report + fasting glucose)</td>
<td>7.9 (7.0, 8.8)</td>
<td>15.8 (14.0, 17.6)</td>
</tr>
</tbody>
</table>

$^a$ Self-reported diabetes from the QSI survey was used as the reference standard to validate the NDSS case definition.

$^b$ “Total diabetes” represents the sum of physician-diagnosed diabetes and undiagnosed diabetes.

$^c$ These test measures do not reflect the actual sensitivity and specificity of the NDSS case definition for physician-diagnosed diabetes. They represent the correction factors of the NDSS case definition for both diagnosed and undiagnosed diabetes, bearing in mind that undiagnosed diabetes cases do not appear in health administrative database.

CI Confidence Interval
Table 4. Prevalence of physician-diagnosed diabetes and total diabetes

<table>
<thead>
<tr>
<th>Diabetes prevalence</th>
<th>Physician-diagnosed diabetes&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Total diabetes&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>percent (95% CI)</td>
<td>percent (95% CI)</td>
</tr>
<tr>
<td>Prevalence (NDSS) in entire random sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unadjusted prevalence</td>
<td>7.5 (6.8, 8.2)</td>
<td>7.5 (6.8, 8.2)</td>
</tr>
<tr>
<td>Adjusted prevalence</td>
<td>6.6 (6.0, 7.2)</td>
<td>10.8 (10.0, 11.6)</td>
</tr>
<tr>
<td>Multiple imputation to random sample</td>
<td>6.6 (5.9, 7.2)</td>
<td>11.2 (10.4, 12.0)</td>
</tr>
<tr>
<td>Weighted prevalence (NDSS)</td>
<td>7.7 (6.8, 8.6)</td>
<td>7.7 (6.8, 8.6)</td>
</tr>
<tr>
<td>Adjusted weighted prevalence (NDSS)</td>
<td>6.8 (5.7, 7.9)</td>
<td>11.2 (9.6, 12.8)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Weighted prevalence (self-report / self-report + fasting glucose)</td>
<td>7.2 (6.3, 8.0)</td>
<td>13.4 (11.7, 15.0)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Physician-diagnosed diabetes represents either NDSS positive case or self-reported diabetes from the QSI survey

<sup>b</sup> “Total diabetes” represents the sum of physician-diagnosed diabetes and undiagnosed diabetes

<sup>c</sup> The weighted prevalence by the NDSS case definition for the entire sample population of 6,247 individuals [7.7% (95% CI 6.8, 8.6)] was adjusted using the capture rate of the NDSS for total diabetes.

<sup>d</sup> Missing glucose values for non-participants generated by multiple imputation for all survey respondents before correcting for sampling weights.

CI Confidence Interval
Supplemental figure S1: Participants’ instructions for home capillary glucose sampling

Rationale for S1: This supplemental figure is a copy of the instructions for the home capillary glucose test that was mailed to all consenting participants. The instructions are easy to follow in a step by step fashion and depict the facility of home self-directed blood glucose testing.

INSTRUCTIONS FOR THE BLOOD SUGAR TEST

- The blood test should be done before breakfast after 10 hours of fasting. It is permitted to drink water.
- If you are uncomfortable pricking your own finger, ask a family member or friend to help you or do this for you.

If you have any questions or if you need help, do not hesitate to call us at 418-691-2404 or toll free at 1 800-561-0213 (for calls outside the Quebec City area).

STEP 1  ➔ Wash your hands with soap in warm water and dry well.
            ➔ Seat yourself comfortably at a table.

STEP 2  ➔ Take out the alcohol swab, the three (3) plastic finger-pricking devices* and the blotting paper with circles.
            ➔ Place these on the table in front of you.

STEP 3  ➔ Clean the tip of either the middle or ring finger (that you will prick) with the alcohol swab.
            ➔ Let the alcohol dry completely.
            ➔ If you are right-handed, prick the finger on your left hand.
            ➔ If you are left-handed, prick the finger on your right hand.

STEP 4  ➔ Hold the pricking device between your two fingers as shown in the photo.
            ➔ Twist and pull off the blue tab of the pricking device.

STEP 5  ➔ Place the tip of the pricking device firmly on the side of the fingertip you have cleaned with the alcohol swab.
            ➔ Squeeze the pricking device like a little stapler.
            ➔ This will make a small puncture in your finger. You will not see a needle.

STEP 6  ➔ Squeeze the tip of the finger that you pricked to obtain a generous drop of blood.
            ➔ Let the drop of blood fall into white circle no. 7 on the blotting paper making sure it spreads beyond the edge of the circle and is absorbed on both sides of the paper.
            ➔ Squeeze a second drop of blood into white circle no. 8, if possible.
            ➔ Do not drop any blood into the circles that are already red.
            ➔ Throw away the pricking device in an ordinary garbage disposal.

STEP 7  ➔ Let the blotting paper (with the blood) air dry for 30 minutes.
            ➔ Place it in the plastic baggie and fill out the required information on the label of the baggie.
            ➔ Please mail it back in the pre-stamped, pre-addressed envelope within the next 10 days.

* Please note that if you have any difficulty in using the first pricking device, two others have been provided. Thank you again for your invaluable cooperation with this study!
CHAPTER 2 CONCLUSION

*Updating the meta-analysis (Chapter 1) with the Quebec validation study (Chapter 2)*

In chapter 1, we examined 11 North American validation studies in a systematic review and pooled 6 of them in a meta-analysis. Compared to surveys and/or medical records, the NDSS case definition has a pooled sensitivity of 82.3% (95%CI 75.8, 87.4%) and a pooled specificity of 97.9% (95%CI 96.5, 98.8%) for physician-diagnosed diabetes. In chapter 2, similar to methods used by the studies included in the meta-analysis, we validated the NDSS case definition from Quebec public administrative databases with self-reported diabetes from a population-based health survey. The sensitivity and specificity of the NDSS case definition for physician-diagnosed diabetes were 84.3% (95%CI 79.3, 88.5) and 97.9% (95%CI 97.4, 98.4) respectively. These NDSS test measures reported in the Quebec study (chapter 2) were quite comparable to the pooled NDSS test measures of the 6 validation studies in the meta-analysis (chapter 1).

Of note, the Quebec validation study of the NDSS case definition meets the inclusion criteria of our systematic review and meta-analysis. Given that we had the raw data available from this Quebec study to populate four-cell values of diagnostic two-by-two tables, we pooled this Quebec study with the other 6 studies (1-6) in an updated meta-analysis. We performed the same DerSimonian & Laird random-effects bivariate regression analysis (24; 25), plotted updated HSROC curves and reported updated pooled test accuracies. By visual inspection, the differences between the updated and previous HSROC were imperceptible (Figure 1).

The updated pooled test measures of all 7 studies were as follows: sensitivity was 82.6% (95%CI 77.1, 87.0%) and specificity was 97.9% (95%CI 96.8, 98.6%). These test measures were near equivalent to the original pooled estimates from 6 studies (excluding the Quebec study). The 95% confidence intervals of the pooled sensitivity did shrink from a span of approximately 12% to 10% when the Quebec study was included.
Likewise, updated correction factors using the same formula (29)
\[
\text{Adjusted prevalence (\%) = \left[\text{unadjusted prevalence (\%) + specificity - 1}\right] / 
\left(specificity + sensitivity - 1\right)}
\]
were near equivalent to the original correction factors:
\[
\text{Adjusted prevalence (\%) = \left[\text{reported unadjusted prevalence (\%) - 2.1\%}\right] / 80.4\%}
\]

By applying correction factors derived from the pooled test accuracies, in chapter 1, we demonstrated that when diabetes prevalence was less than 10.6\%, the case definition underestimated the prevalence but did the opposite when prevalence was beyond 10.6\%. Using updated correction factors, the crossover point at which the case definition neither overestimated nor underestimated the diabetes prevalence was quite similar at 10.7\%. This is evidence that the NDSS test performance from Quebec administrative databases was quite consistent to the reported NDSS test performance in earlier validation studies conducted in other parts of North America.

**Applying correction factors to account for undiagnosed diabetes in Canada**

In chapter 2, we also compared the NDSS case definition to self-reported diabetes combined with glucose testing to yield a sensitivity of 58.2\% (95\%CI 52.2, 64.6) and specificity of 98.7\% (95\%CI 98.0, 99.3). Adjusted for sampling weights, physician-diagnosed diabetes prevalence in Quebec was 6.8\% (95\%CI 5.7, 7.9) and total diabetes (physician diagnosed and undiagnosed) prevalence was 11.2\% (95\%CI 9.6, 12.8), implying that over 40\% of diabetes cases remained undiagnosed.

By applying correction factors that take undiagnosed diabetes into account, reported NDSS population prevalence can be adjusted to estimate the prevalence of total diabetes (undiagnosed and physician-diagnosed) (29) using the following formula:
\[
\text{Adjusted prevalence (\%) = \left[\text{reported unadjusted prevalence (\%) - 1.3\%}\right] / 56.9\%}
\]
We adjusted the yearly Canadian population prevalence of diabetes cases (28) with the above correction factors to account for both the test performance of the NDSS case definition and undiagnosed diabetes from fiscal year 2002/3 to 2006/7 (figure 2).
In year 2006/7, unadjusted diabetes prevalence was 8.0%; adjusted physician-diagnosed diabetes was 7.3% and total diabetes was 11.8%. Hence, the estimated prevalence of undiagnosed diabetes in Canada was 4.5% (38.1% of all diabetes case). The prevalence of undiagnosed diabetes estimated using these proposed correction factors is quite comparable to other North American studies that have employed in-person glucose testing to estimate the prevalence of undiagnosed diabetes (42; 43).

We observed an increase in physician-diagnosed diabetes prevalence in Canada by ~0.4% per year between years 2002/3 to 2006/7. Using a linear model, we projected diabetes prevalence to the future and estimated that physician-diagnosed diabetes prevalence would reach ~13% in year 2018. Assuming diabetes detection and screening mechanisms remain unchanged over time, we applied the same correction factors to diabetes prevalence and anticipated that the undiagnosed diabetes and total diabetes prevalence would reach ~7% and ~20% respectively in year 2018 (figure 2). The rising prevalence of physician-diagnosed and undiagnosed diabetes in Canada is a stark reflection of a major epidemic facing North America (87). The diabetes epidemic is certainly a global health crisis represented by mounting rates of diabetes that can be observed in all continents, from developing to developed nations (88; 89). Current approaches to prediction, prevention and medical care are evidently inadequate in controlling this escalating clinical and public health issue.

Specificity of the NDSS case definition

As discussed in chapter 1 Conclusions, high specificity of the definition may be particularly desirable for certain roles of administrative databases, such as evaluating health service delivery and establishing diabetes cohorts for associative or etiologic research. Mechanisms compromising the sensitivity of claims-based administrative definitions have been described in the introduction of chapter 1. Here, we discuss potential mechanisms involved in suboptimal specificity of physician claims in administrative databases:
Physicians may mistakenly fill service claims for diabetes when conducting in-person diabetes screening for at-risk patients, even if laboratory results indicate that the patient does not have diabetes. If the same billing error is repeated within a two-year period, a non-case may eventually fulfill the NDSS criteria. Physicians may fill service claims for diabetes when treating patients with “pre-diabetes” who are in the strictest sense non-cases. Some of these patients may in fact be managed similarly to early diabetes (e.g. the use of anti-diabetic medication). Gestational diabetes may be erroneously mislabelled or miscoded as diabetes (the chronic condition) in hospital discharge summaries or service claims. During data extraction from health administrative databases, the distinction between the diagnostic code for gestational diabetes and diabetes (the chronic condition) may have been overlooked in some of the included studies of the systematic review (chapter 1). In the Quebec validation study (chapter 2), we ensured that physician claims for gestational diabetes were excluded from our analysis. Notably, among survey respondents, we identified only 18 cases of gestational diabetes out of the 3,039 non-cases by self-report, suggesting that gestational diabetes may only have a trivial impact on diabetes prevalence estimation.

We were not able to distinguish between Type 1 and Type 2 diabetes on physician claims or hospitalization records. Given that physician follow-up visits for type 1 diabetes patients are generally more closely spaced than type 2 diabetes patients, the sensitivity of the NDSS case definition for type 1 diabetes may be much higher than that for type 2 diabetes. Given that the prevalence of type 2 diabetes (90%) far exceeds type 1 diabetes (10%), especially among adults ≥ 20 years of age (90), validation studies probably evaluate the administrative definition’s ability to capture Type 2 diabetes.

*Other limitations: Imperfect reference standards*

The primary goal of validation studies is to quantify the degree of misclassification by administrative definitions which is influenced by flaws in data collection procedures or imperfect definitions. However, the primary limitation of validation studies is misclassification by reference standards, specifically nationwide surveys and primary care chart reviews. Reference
standards are assumed to provide perfect sensitivity and specificity when validating administrative definitions through conventional Frequentist methods.

Lack of participation occurs more commonly among the non-diseased which may inflate the diabetes prevalence estimate and, as discussed in chapter 2, underreporting of diabetes in surveys may occur as a result of an individual’s lack of knowledge of their disease or the stigma of having a chronic disease commonly associated with poor lifestyle habits; this can potentially underestimate diabetes prevalence.

Misclassification in primary care chart reviews is probably differential as younger, asymptomatic and healthier patients tend not to be screened for diabetes while screening procedures are generally directed towards frequent users of health services, older individuals carrying multiple comorbidities and those with clinical evidence of advanced diabetes. Medical records can only be a gold standard if all patients within the primary care setting routinely undergo thorough diabetes evaluations. Furthermore, even if medical chart reviews are conducted in randomly-selected family physician offices, a low physician participation rate can bias results.

In the absence of a gold standard (50), Bayesian statistics can be employed to compare the NDSS case definition with self-reported diabetes in chapter 2. Bayesian methods are supposedly more realistic in determining disease prevalence as they do not assume that data sources provide perfect diagnostic information on disease ascertainment. The Bayesian approach relaxes the assumption of perfect sensitivity and specificity through incorporating prior uncertainties of non-gold standard tests while making simultaneous inferences about the accuracy of the NDSS case definition, the “reference standard” and diabetes prevalence. Other studies have also employed Bayesian methods to estimate the population prevalence of certain chronic diseases from administrative databases (29; 91-93).

Other limitations: Measurement error of fasting glucose on mailed-in samples

Potential sources of measurement error from a single capillary fasting blood glucose sample have been described in chapter 2. These include: [1]
regression to the mean (attenuation or regression dilution bias), [2] inadequate fasting period, [3] influence of extravascular fluid in capillary sampling, [4] time lapse from home sampling to laboratory testing. In addition, we recognize that we obtained fasting whole blood and not plasma glucose measurements on mailed-in samples. Yet, we chose the widely-accepted threshold of 7 mmol/l for fasting plasma glucose to define diabetes instead of 6.1 mmol/l for fasting whole blood glucose which was recommended by WHO in 1999 (84). The rationale for selecting 7 mmol/l was to avoid overestimating diabetes prevalence given the wide variability of whole blood glucose compared to plasma glucose. As a result, we likely have a non-trivial proportion of false negatives and underestimated diabetes prevalence in the study.

Mailed-in glucose sampling requires further validation with in-person glucose testing on a subset of the study cohort to adjust for measurement errors and to determine the most appropriate threshold for fasting glucose to define diabetes in the population. We did not conduct this extension project because in-person glucose testing was costly and labour intensive. Alternatively, we considered identifying incident NDSS diabetes cases in the year following the study among those who provided mailed-in blood samples. The statistical distribution of fasting glucose from mailed-in blood samples among incident diabetes cases shortly after the study may provide inferences on the most appropriate cut-off for diagnosing diabetes based on this new diabetes screening method. However, we concluded that this supplementary analysis was beyond the research objectives of this thesis and thus reserved for a future study.

Other limitations: Selection bias from non-response

In chapter 2, we corrected for selection bias by adjusting for sampling weights and performing multiple imputations based on the baseline information from administrative data. While respondents and non-respondents did not differ greatly on important baseline characteristics (Table 1 of chapter 2), we acknowledge that there could be residual confounding from unmeasured confounders:
[1] The health survey and self-testing instructions were only available in French and English, language barriers could have limited the participation of new immigrants and some ethnic minorities. In addition, we were not provided individual-level data on race/ethnicity for the entire original sample.

[2] We were unable to perform telephone surveys on those who did not have landlines, although we did attempt to reach these subjects by mail. However, it was unclear whether not having a landline was related to diabetes status or health status, or influenced prevalence estimation.

Despite the numerous potential uses of mailed-in blood samples for diabetes screening and prevalence estimation of undiagnosed diabetes (highlighted in chapter 2), the rather low response rate (25.6%) limits its utility as a solution for imperfect reference standards for validating administrative definitions. Furthermore, individuals with physician-diagnosed diabetes were more inclined to return an analyzable blood sample; thus, the response rate among non-diabetes cases was slightly lower at 24.8% (1,328 out of 5,778).

Other limitations: Generalizability

The correction factors proposed herein to account for the accuracy of the NDSS case definition and undiagnosed diabetes apply primarily to single payer health care systems that maintain national health administrative databases. They may not be generalizable to other healthcare systems, e.g. multiple private insurer systems, mixed private-public systems or multitiier systems, where it may be difficult to track all health service utilization for a given individual. Nevertheless, if administrative databases from both the private and public systems are available, and matched through subject-specific linkage, the application of correction factors may be valid.

In the systematic review, the small number of studies that were included prohibited the exploration of region-specific characteristics that may have contributed to the heterogeneity of the meta-analysis through meta-regression or subgroup analyses. For example, diagnostic coding practices vary across regions in Canada. In Alberta, physician claims allow inclusion of up to 3 diagnostic codes, which may have an impact on the sensitivity of the
NDSS case definition. Thus, the applicability of our results may not be as far reaching as originally intended.

A linear model was employed to extrapolate diabetes prevalence to year 2018. While the growth in diabetes prevalence from fiscal year 2002/3 to 2006/7 appears to be linear, selecting this model for future prediction is arguably arbitrary. Considering the aging population, longer life expectancies, obesity epidemic and changing ethnic landscape (90), the Canadian population may be accruing high risk groups faster than anticipated (e.g. exponential growth). The accuracy of the NDSS case definition may vary in older populations and ethnic minorities who face unique barriers to health services. Applying the same correction factors to adjust prevalence estimates to these subpopulations and to future Canadian populations may not be appropriate due to demographic change. This emphasizes the importance of repeating validation studies of administrative definitions at regular intervals (e.g. every 5 to 10 years). Furthermore, we have restricted our study to evaluating diabetes prevalence and have not considered diabetes incidence. The NDSS case definition may have a vastly different accuracy for incident diabetes; therefore the correction factors may not apply for incidence estimates.

**Overall conclusions**

Despite these limitations, this research work demonstrates that administrative data can be a powerful resource for diabetes surveillance. The NDSS case definition is sufficiently sensitive and specific for prevalence estimation and monitoring trends over time. Correction factors can be applied to improve the accuracy of prevalence estimation although the validity of correction factors depends on whether reference standards provide perfect sensitivity and specificity.

We also confirmed the utility and feasibility of an innovative diabetes screening strategy involving mailed-in blood samples to detect undiagnosed diabetes in the community. This diabetes screening test can be used to estimate undiagnosed diabetes prevalence in the general population. Mailed-in blood collection for glucose testing or glycated hemoglobin measurements can possibly be combined with existing diabetes prevention strategies (e.g.
diabetes risk scores) to improve early detection of diabetes in at-risk populations.

The impact of diabetes screening on the general population (e.g. low and moderate risk individuals) in the prevention of diabetes-related mortality and cardiovascular outcomes remains uncertain (94). However, given the long latent phase of the disease and negligible potential harm (95), early detection of diabetes in asymptomatic individuals with an elevated diabetes risk through regular screening at 3-to-5 year intervals, may reduce this growing burden of disease, and allow for early institution of diabetes treatment to prevent diabetes-related complication (69; 96; 97).
Figure 1: Random-effects bivariate regression analysis of pooled test accuracies from 7 studies (including Quebec study)

The Hierarchical Summary Receiver Operator Characteristics (HSROC) curve displays the 95% confidence interval of the summary operating point and the 95% prediction region, which is the confidence region for a forecast of the true sensitivity and specificity in a future study. This updated HSROC was visually very similar to the previous HSROC of 6 studies that excluded the Quebec study.
Prevalence of diabetes

Figure 2: Prevalence of total diabetes in Canada

Crude prevalence: prevalence of diabetes in Canada for fiscal years 2002/3 through 2006/7 obtained from the NDSS 2009 report (28); Adjusted prevalence: prevalence after applying correction factors \(((\text{Prevalence}(\%)) - 2.1) / 0.804\); Projected crude prevalence: future prevalence assuming an increase of 0.4% per year; Adjusted projected prevalence: Future prevalence after applying correction factors that accounts for the NDSS test performance; Total diabetes: Estimated prevalence of physician-diagnosed and undiagnosed diabetes after applying correction factors that account for both the NDSS test performance and undiagnosed diabetes \(((\text{Prevalence}(\%)) - 1.31) / 0.569\).
REFERENCES


27. Deeks JJ, Macaskill P, Irwig L: The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of clinical epidemiology 2005;58:882-893


34. Solberg LI, Engebretson KI, Sperl-Hillen JM, Hroscikoski MC, O'Connor PJ: Are claims data accurate enough to identify patients for performance measures or quality improvement? The case of diabetes, heart disease, and
Prevalence of diabetes


35. Shah BR, Manuel DG: Self-reported diabetes is associated with self-management behaviour: a cohort study. BMC health services research 2008;8:142


68
41. Carnahan RM, Moores KG: Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. Pharmacoepidemiology and drug safety 2012;21 Suppl 1:82-89


50. German RR: Sensitivity and predictive value positive measurements for public health surveillance systems. Epidemiology 2000;11:720-727


Prevalence of diabetes

claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions. Health services research 2004;39:1839-1857


71


72


Prevalence of diabetes


86. Mayor S: Nearly half of adults in England don't understand health information material, study indicates. BMJ 2012;345:e8364


Prevalence of diabetes


97. Charles M, Simmons RK, Williams KM, Roglic G, Sharp SJ, Kinmonth AL, Wareham NJ, Griffin SJ: Cardiovascular risk reduction following diagnosis of diabetes by screening: 1-year results from the ADDITION-
Prevalence of diabetes

Supplemental file 1: MEDLINE and EMBASE Search strategies

MEDLINE
Database: Ovid MEDLINE(R) 1950 to Present with Daily Update, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <November 15, 2010>
Database: Ovid MEDLINE(R) 1950 to Present with Daily Update, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <November 15, 2010>
Search Strategy:
--------------------------------------------------------------------------------
1 administrative data.ti,ab. (2391)
2 administrative database:.ti,ab. (1329)
3 Databases, Factual/ (30843)
4 factual database/ (30843)
5 Databases as Topic/ (7263)
6 database/ (0)
7 Medical Record Linkage/ (2778)
8 administrative databank:.ti,ab. (2)
9 factual database:.ti,ab. (17)
10 factual databank:.ti,ab. (2)
11 factual data.ti,ab. (53)
12 exp medical records/ (69234)
13 exp medical record/ (69234)
14 exp medical records systems, computerized/ (18265)
15 (medical record or health record or medical records or health records).ti,ab. (45693)
16 medical transcription:.ti,ab. (73)
17 exp Registries/ (42227)
18 registry/ (39948)
19 (registry or registries).ti,ab. (41377)
20 (utilisation data: or utilisation data: or claims data: or managed care data: or physician billing data: or hospitalization data: or linked data:).ti,ab. (5478)
21 (administrative healthcare data: or administrative health care data: or administrative health data: or administrative health data:).ti,ab. (142)
22 (medical records based index or claims based index).ti,ab. (5)
23 (register and (link or links or linked or linkage or linking)).ti,ab. (1850)
24 or/1-23 [ADMINISTRATIVE DATA (BROAD)] (206366)
25 Validation Studies/ (47964)
26 validation study/ (0)
27 Validation Studies as Topic/ (588)
28 Validation Studies.pt. (47964)
29 (validat: or validity).ti,ab. (236536)
30 or/25-29 [VALIDATION STUDIES] (257352)
--------------------------------------------------------------------------------------
98 exp Diabetes Mellitus/ (261426)
99 diabet:.ti,ab. (319880)
100 98 or 99 [DIABETES MELLITUS (BROAD)] (366136)
101 24 and 30 and 100 [ADMINISTRATIVE DATA (BROAD) +
VALIDATION STUDIES + DIABETES MELLITUS (BROAD)] (438)
102 diabetes mellitus.ti,ab. (97761)
103 98 or 102 [DIABETES MELLITUS (NARROW)] (287652)
104 24 and 30 and 103 [ADMINISTRATIVE DATA (BROAD) +
VALIDATION STUDIES + DIABETES MELLITUS (NARROW)] (296)
SAVED EN 2143-2438 (277 UNIQUE)
105 31 and 103 [ADMINISTRATIVE DATA (NARROWEST) + DIABETES
MELLITUS (NARROW)] (214)
106 104 or 105 (476)
107 105 not 104 (180) SAVED EN 2439-2618 (165 UNIQUE)
108 (comorbidity index or comorbidity indexes or comorbidity indices).ti,ab.
[COMORBIDITY INDEXES] (995)
109 24 and 30 and 108 [ADMINISTRATIVE DATA (BROAD) +
VALIDATION STUDIES + COMORBIDITY INDEXES] (44) SAVED EN
2619-2662 (42 UNIQUE)
110 31 and 108 [ADMINISTRATIVE DATA (NARROWEST) +
COMORBIDITY INDEXES] (63)
111 109 or 110 (88)
112 110 not 109 (44) SAVED EN 2663-2706 (41 UNIQUE)

MEDLINE / EMBASE
Database: EMBASE <1980 to 2010 Week 45>, Ovid MEDLINE(R) 1950 to
Present with Daily Update, Ovid MEDLINE(R) In-Process & Other Non-
Indexed Citations <November 15, 2010>
Search Strategy:
--------------------------------------------------------------------------------
1 administrative data.ti,ab. (4927)
2 administrative database:.ti,ab. (2833)
3 Databases, Factual/ (41452)
4 factual database/ (41452)
5 Databases as Topic/ (79396)
6 database/ (72133)
7 Medical Record Linkage/ (92752)
8 administrative databank:.ti,ab. (4)
9 factual database:.ti,ab. (41)
10 factual databank:.ti,ab. (6)
11 factual data.ti,ab. (110)
12 exp medical records/ (159208)
13 exp medical record/ (159208)
14 exp medical records systems, computerized/ (108239)
15 (medical record or health record or medical records or health records).ti,ab.
(96352)
16 medical transcription:.ti,ab. (150)
17 exp Registries/ (73671)
18 registry/ (71392)
19 (registry or registries).ti,ab. (88310)
20 (utilization data: or utilisation data: or claims data: or managed care data: or
physician billing data: or hospitalization data: or linked data:).ti,ab. (11879)
21 (administrative healthcare data: or administrative health care data: or
administrative health data: or administrative health data:).ti,ab. (296)
Prevalence of diabetes

22 (medical records based index or claims based index).ti,ab. (10)
23 (register and (link or links or linked or linkage or linking)).ti,ab. (3848)
24 or/1-23 [ADMINISTRATIVE DATA (BROAD)] (466927)
25 Validation Studies/ (72226)
26 validation study/ (24262)
27 Validation Studies as Topic/ (24850)
28 Validation Studies.pt. (47964)
29 (validat: or validity).ti,ab. (504774)
30 or/25-29 [VALIDATION STUDIES] (533582)

----------------------------------------------------------------------------------
98 exp Diabetes Mellitus/ (665833)
99 diabet:.ti,ab. (696125)
100 98 or 99 [DIABETES MELLITUS (BROAD)] (839087)
101 24 and 30 and 100 [ADMINISTRATIVE DATA (BROAD) + VALIDATION STUDIES + DIABETES MELLITUS (BROAD)] (1027)
102 diabetes mellitus.ti,ab. (211612)
103 98 or 102 [DIABETES MELLITUS (NARROW)] (706730)
104 24 and 30 and 103 [ADMINISTRATIVE DATA (BROAD) + VALIDATION STUDIES + DIABETES MELLITUS (NARROW)] (810)
105 31 and 103 [ADMINISTRATIVE DATA (NARROWEST) + DIABETES MELLITUS (NARROW)] (554)
106 104 or 105 (1276)
107 105 not 104 (466)
108 (comorbidity index or comorbidity indexes or comorbidity indices).ti,ab. [COMORBIDITY INDEXES] (2258)
109 24 and 30 and 108 [ADMINISTRATIVE DATA (BROAD) + VALIDATION STUDIES + COMORBIDITY INDEXES] (89)
110 31 and 108 [ADMINISTRATIVE DATA (NARROWEST) + COMORBIDITY INDEXES] (128)
111 109 or 110 (180)
112 110 not 109 (91)

----------------------------------------------------------------------------------
129 remove duplicates from 104 (592)
130 from 129 keep 1-316 [EMBASE: ADMINISTRATIVE DATA (BROAD) + VALIDATION STUDIES + DIABETES MELLITUS] (316) SAVED EN 3634-3949 (314 UNIQUE)
131 remove duplicates from 107 (291)
132 from 131 keep 1-127 [EMBASE: ADMINISTRATIVE DATA (NARROWEST) + DIABETES MELLITUS] (127) SAVED EN 3950-4076 (127 UNIQUE)

----------------------------------------------------------------------------------
Supplemental file 2: The QUADAS tool

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Supplemental file 3: Quality assessment by QUADAS

<table>
<thead>
<tr>
<th>QUADAS Question no.</th>
<th>Chen (3)</th>
<th>Harris (6)</th>
<th>Hebert (2)</th>
<th>Hux (1)</th>
<th>O’Connor (5)</th>
<th>Robinson (4)</th>
<th>Koleba (14)</th>
<th>Lix (37)</th>
<th>Shah (35)</th>
<th>Solberg (34)</th>
<th>Southern (38)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>6*</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>7</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>10*</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>11*</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>13</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Unclear</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>14</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
</tbody>
</table>

Score (max 14) | 11 | 11 | 12 | 11 | 12 | 12 | 7 | 11 | 11 | 7 | 12

Bias assessment (max 5) | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 5 | 5 | 2 | 5

*Questions selected from QUADAS to constitute the “Bias Assessment”. QUADAS questions are displayed in Supplemental file 2.
Supplemental file 4: Adjusted and unadjusted prevalence of diabetes in Canada

<table>
<thead>
<tr>
<th>Fiscal Year</th>
<th>Unadjusted Prevalence (%)</th>
<th>Adjusted Prevalence (%)</th>
<th>Difference in Prevalence (%)</th>
<th>Cases, n</th>
<th>Adjusted cases, n</th>
<th>Difference in cases, n</th>
<th>Population, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002/3</td>
<td>6.4</td>
<td>5.4</td>
<td>-1.0</td>
<td>1,540,001</td>
<td>1,293,996</td>
<td>-246,005</td>
<td>24,134,536</td>
</tr>
<tr>
<td>2003/4</td>
<td>6.8</td>
<td>5.9</td>
<td>-0.9</td>
<td>1,657,282</td>
<td>1,437,778</td>
<td>-219,504</td>
<td>24,534,007</td>
</tr>
<tr>
<td>2004/5</td>
<td>7.2</td>
<td>6.4</td>
<td>-0.8</td>
<td>1,781,879</td>
<td>1,584,653</td>
<td>-197,226</td>
<td>24,919,442</td>
</tr>
<tr>
<td>2005/6</td>
<td>7.6</td>
<td>6.9</td>
<td>-0.7</td>
<td>1,916,172</td>
<td>1,736,332</td>
<td>-179,840</td>
<td>25,318,884</td>
</tr>
<tr>
<td>2006/7</td>
<td>8.0</td>
<td>7.4</td>
<td>-0.6</td>
<td>2,061,995</td>
<td>1,894,615</td>
<td>-167,380</td>
<td>25,753,921</td>
</tr>
<tr>
<td>Change in 5 years</td>
<td>1.6</td>
<td>2.0</td>
<td>0.4</td>
<td>521,994</td>
<td>600,619</td>
<td>78,625</td>
<td>1,619,385</td>
</tr>
</tbody>
</table>

Diabetes prevalence rates from fiscal year 2002/3 to 2006/7 of individuals aged ≥20 years from the NDSS 2009 report were adjusted using the following correction formula: \[[\text{prevalence} \% - 2.1]/0.802\]. The adjusted cost of diabetes per year is consistently lower than the estimated cost calculated from diabetes cases identified by the NDSS. However, the increase in adjusted diabetes prevalence over the 5-year time span is greater by 0.4% than the crude prevalence. This amounts to an additional 78,625 diabetes cases that would not have been accounted for without the application of correction factors.
## Supplemental file 5: Funding sources of included validation studies

<table>
<thead>
<tr>
<th>Title</th>
<th>population</th>
<th>Author</th>
<th>Funding sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigating concordance in diabetes diagnosis between primary care charts (electronic medical records) and health administrative data: a retrospective cohort study</td>
<td>Ontario, Canada</td>
<td>Harris (6)</td>
<td>Institute for Clinical Evaluative Sciences (ICES)</td>
</tr>
<tr>
<td>Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm.</td>
<td></td>
<td>Hux (1)</td>
<td>Medical Research Council of Canada</td>
</tr>
<tr>
<td>Self-reported diabetes is associated with self-management behaviour: a cohort study</td>
<td></td>
<td>Shah &amp; Manuel (35)</td>
<td>Canadian Diabetes Association; CIHR</td>
</tr>
<tr>
<td>Population-based data sources for chronic disease surveillance.</td>
<td>Manitoba, Canada</td>
<td>Lix (37)</td>
<td>CIHR</td>
</tr>
<tr>
<td>Estimating the burden of disease. Comparing administrative data and self-reports</td>
<td></td>
<td>Robinson (4)</td>
<td>Heart and stroke foundation</td>
</tr>
<tr>
<td>Prescription drug data and the national diabetes surveillance system case definition.</td>
<td>Saskatchewan, Canada</td>
<td>Koleba (14)</td>
<td>CIHR</td>
</tr>
<tr>
<td>Validating ICD coding algorithms for diabetes mellitus from administrative data</td>
<td>Alberta and British Columbia, Canada</td>
<td>Chen (3)</td>
<td>CIHR</td>
</tr>
<tr>
<td>Validity of administrative data claim-based methods for identifying individuals with diabetes at a population level</td>
<td></td>
<td>Southern (38)</td>
<td>Canadian Diabetes Association</td>
</tr>
<tr>
<td>Identifying persons with diabetes using Medicare claims data</td>
<td>Minnesota, USA</td>
<td>Hebert (2)</td>
<td>Association of Schools of Public Health/ Communicable disease Centre; Prevention Cooperative Agreement</td>
</tr>
<tr>
<td>Identifying diabetes mellitus or heart disease among health maintenance organization members: Sensitivity, specificity, predictive value, and cost of survey and database methods</td>
<td></td>
<td>O’Connor (5)</td>
<td>Health Partners Research Foundation</td>
</tr>
<tr>
<td>Are claims data accurate enough to identify patients for performance measures or quality improvement? The case of diabetes, heart disease, and depression.</td>
<td></td>
<td>Solberg (34)</td>
<td>Robert Wood Johnson Foundation</td>
</tr>
</tbody>
</table>
Supplemental file 6: PRISMA checklist

<table>
<thead>
<tr>
<th>Section/topic</th>
<th>#</th>
<th>Checklist item</th>
<th>Reported on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>### TITLE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>1</td>
<td>Identify the report as a systematic review, meta-analysis, or both.</td>
<td>1</td>
</tr>
<tr>
<td>### ABSTRACT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structured summary</td>
<td>2</td>
<td>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.</td>
<td>3</td>
</tr>
<tr>
<td>### INTRODUCTION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>3</td>
<td>Describe the rationale for the review in the context of what is already known.</td>
<td>4</td>
</tr>
<tr>
<td>Objectives</td>
<td>4</td>
<td>Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).</td>
<td>5</td>
</tr>
<tr>
<td>### METHODS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol and registration</td>
<td>5</td>
<td>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.</td>
<td>N/A</td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td>6</td>
<td>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.</td>
<td>5</td>
</tr>
<tr>
<td>Information sources</td>
<td>7</td>
<td>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.</td>
<td>5</td>
</tr>
<tr>
<td>Search</td>
<td>8</td>
<td>Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.</td>
<td>5 and Supplement 1</td>
</tr>
<tr>
<td>Study selection</td>
<td>9</td>
<td>State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).</td>
<td>6</td>
</tr>
<tr>
<td>Data collection process</td>
<td>10</td>
<td>Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.</td>
<td>6</td>
</tr>
<tr>
<td>Data items</td>
<td>11</td>
<td>List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.</td>
<td>6</td>
</tr>
<tr>
<td>Risk of bias in individual studies</td>
<td>12</td>
<td>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.</td>
<td>6</td>
</tr>
<tr>
<td>Summary measures</td>
<td>13</td>
<td>State the principal summary measures (e.g., risk ratio, difference in means).</td>
<td>6</td>
</tr>
<tr>
<td>Synthesis of results</td>
<td>14</td>
<td>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.</td>
<td>7</td>
</tr>
</tbody>
</table>
Prevalence of diabetes

<table>
<thead>
<tr>
<th>Section/topic</th>
<th>#</th>
<th>Checklist item</th>
<th>Reported on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of bias across studies</td>
<td>15</td>
<td>Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).</td>
<td>6 and 7</td>
</tr>
<tr>
<td>Additional analyses</td>
<td>16</td>
<td>Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.</td>
<td>8</td>
</tr>
</tbody>
</table>

### 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. | 8                 |
| 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. | 8, 9, 30 and 31 (Table 1) and Supplement 5 |
| 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). | 9, 10, 11 and 12 |
| 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. | 30 and 31 (Table 1) and Figure 2 |
| Synthesis of results     | 21 | Present results of each meta-analysis done, including confidence intervals and measures of consistency. | 13 and 14; Figures 3 and 4 |
| Risk of bias across studies | 22 | Present results of any assessment of risk of bias across studies (see Item 15). | Supplement 3 |
| Additional analysis      | 23 | Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). | 14 and 15 |

### 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). | 15 and 17          |
| 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). | 19, 20 and 21 |
| Conclusions              | 26 | Provide a general interpretation of the results in the context of other evidence, and implications for future research. | 21                 |

### 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. | 2                 |