Australian Government
Australian Institute of Health and Welfare

Validating diabetes case definition algorithms using primary health care data

Technical report

AIHW

Validating diabetes case definition algorithms using primary health care data

Technical report

Australian Institute of Health and Welfare
Canberra
Catalogue number CDK 24

The AIHW is a Corporate Commonwealth entity producing authoritative and accessible information and statistics to inform and support better policy and service delivery decisions, leading to better health and wellbeing for all Australians.

A complete list of the Institute's publications is available from the Institute's website www.aihw.gov.au.

ISBN 978-1-923085-84-8 (Online)

ISBN 978-1-923085-85-5 (Print)

Suggested citation
Australian Institute of Health and Welfare (2024) *Validating diabetes case definition algorithms using primary health care data: Technical report*, catalogue number CDK 24, AIHW, Australian Government.

Any enquiries relating to copyright or comments on this publication should be directed to:
Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601
Tel: (02) 6244 1000
Email: info@aihw.gov.au

Published by the Australian Institute of Health and Welfare

Please note that there is the potential for minor revisions of data in this report. Please check the online version at www.aihw.gov.au for any amendments.

Contents

Summary

Identifying cases of disease in large databases is important for surveillance, research, quality improvement and clinical care. Case definitions can be created from single or combinations of information including diagnostic codes, medications, condition-specific service claims and laboratory results.

General practice data are a rich information source which contains a range of information to enable identification of whether a person has a particular condition or not. Although general practice data are generally not available in linked data sets, these data often contain markers of diabetes status that are commonly found in such data sets. The aim of this report is to explore approaches for diabetes mellitus (diabetes) case definitions using markers of diabetes status including diabetes-specific prescriptions, pathology tests and Medicare Benefits Schedule (MBS) service items recorded in general practice data. The identified diabetes case definitions can be used (or refined where necessary) to identify people with diabetes, in linked data collections that include these diabetes status markers.

All analyses were conducted using a 10% sample of MedicineInsight, a general practice data set. The diabetes definition from the MedicineInsight condition flag (standard definition) was considered as the reference standard and case definitions using diabetes markers recorded in MedicineInsight were compared against this standard definition. Case definitions for diabetes, type 1 and type 2 diabetes were identified. While type 1 diabetes affects people of all ages, to improve the statistical power of identifying type 1 diabetes based on the assessed diabetes markers the analysis for type 1 diabetes was limited to people aged under 35. Analyses for all diabetes and type 2 diabetes included people of all ages.

The goal was to identify diabetes case definitions that minimise misclassification risk without compromising the predictive power (precision, as measured by the positive predictive value (PPV)) of the diabetes markers for detecting diabetes. Definitions with a very high PPV for identifying diabetes were preferred.

This information will help in understanding criteria for identifying people with diagnosed diabetes that can be applied to other data sets with similar diabetes markers, particularly linked data.

Key findings from the validation of algorithms for diabetes case definition

Findings from this analysis show that approaches using a combination of diabetes markers provide robust definitions for diabetes with very high precision (≥ 90%) and acceptable sensitivity (> 60%). The diabetes case definition with a minimum of one diabetes prescription and at least one HbA1c test each with a gap of 6 months to another HbA1c test had the highest precision (PPV 96%) and sensitivity of 61%. The probability that a person meeting this definition has diabetes (positive probability) is very high at 96% and the probability that a person not meeting this definition has diabetes (negative probability) is low (3%), suggesting that this definition is good for identifying people with diabetes. A sensitivity of 70% and PPV of 92% was observed for the algorithm with at least one diabetes prescription and 2 or more

HbA1c tests recorded any time during the study period (positive probability 93%, negative probability 2%).

Single markers such as diabetes-specific prescriptions and MBS items had high precision, but sensitivity was very low. This suggests that most people with each single marker had diabetes according to the MedicineInsight standard definition, but of those identified as having diabetes by the standard definition the proportion who had each single marker recorded was small.

Increasing the minimum number of records for each marker during the study period improves precision, but results in very low sensitivity.

Key findings from case definitions for diabetes type ascertainment

Using a minimum of 2 prescriptions for insulin only (without other diabetes medicines) at any time during the study period for people aged under 35 had the highest precision for identifying type 1 diabetes (91%) and sensitivity of 67%. The probability that a person not meeting this definition has type 1 diabetes is very low at 0.1%. The definition for type 1 diabetes with the highest sensitivity (79%) and PPV of 85% was having at least one prescription for insulin only (without other diabetes medicines) during the study period (negative probability 0.1%).

Using the study population containing all age groups, a minimum of one diabetes prescription with people prescribed insulin only excluded and one or more HbA1c test each with a 6-month gap of another HbA1c test had the highest precision (91%) of identifying type 2 diabetes and sensitivity of 61%. The probability that a person meeting this definition has type 2 diabetes is 91% and the probability that a person not meeting this definition has type 2 diabetes is 3%. The definition algorithm for type 2 diabetes with the highest sensitivity of 76% and PPV of 82% was having at least one diabetes prescription (excluding people with only insulin prescriptions) and at least one HbA1c test (positive probability 84%, negative probability 1%).

Conclusion

Findings from this analysis indicate that approaches using a combination of markers of diabetes status provide better capture of people with diagnosed diabetes.

We have identified potential case definition algorithms that can be used to identify people with diabetes in a cohort of people attending primary care. While the definitions for diabetes and type 2 diabetes included people of all ages, case definitions for type 1 diabetes were limited to people aged under 35. Therefore, the findings should be interpreted accordingly.

It is important to note that the diabetes case definitions and validity estimates observed in the current analysis might vary with those from other data sets due to some differences in the diabetes markers in the data used and other administrative data sets such as the Pharmaceutical Benefits Scheme (PBS) and MBS. Moreover, performance characteristics like precision are influenced by the prevalence of the condition in the study population, which might limit the generalisability of the findings in study populations with different prevalence estimates. Using the same case definition, PPV could decrease and negative predictive

value (NPV) increase in a setting where the prevalence of diabetes is lower than that observed in this study.

In this analysis the diabetes marker definition algorithms were compared to the MedicineInsight standard definition (diabetes definition from the MedicineInsight condition flags). This reference standard may have limitations if there is incomplete recording of diabetes in the MedicineInsight data fields used for the standard definition. A reference standard with limitations can introduce measurement error in the analysis and the performance of the definition algorithms depends on the quality of the reference standard.

Nevertheless, the algorithms in this report provide approaches for diabetes case definition that could be utilised (or refined where necessary) to identify people with diabetes in data collections that include these diabetes status markers. These insights can supplement the existing data sources, that is the National Diabetes Services Scheme (NDSS), in identifying people with diagnosed diabetes, particularly in linked data collections, thus enabling better estimation of its prevalence and further monitoring. This is important for implementing policies for prevention and management as well as proper resource allocation.

1 Introduction

The Australian Institute of Health and Welfare (AIHW) regularly reports on the prevalence, incidence, treatment and management of diabetes in Australia. Identifying data gaps and improving the quality of available data are essential components of this work.

The National Diabetes Services Scheme (NDSS) forms the foundation for monitoring people living with diabetes and has more than 1.45 million Australians registered for assistance with managing diabetes. The prevalence of diabetes has been documented in the Australian Bureau of Statistics' National Health Survey (NHS) and NDSS data and while each of these data sets may provide good estimates of people with diagnosed diabetes, each has limitations and some people with diagnosed diabetes may not be captured (AIHW 2009, 2023).

Previous projects have demonstrated that although the NDSS has a very high capture of people with diabetes there are people with diabetes who are not registered with the scheme. Registration with the NDSS is voluntary for eligible individuals and people with type 2 diabetes who access diabetes consumables to monitor their diabetes at home, or require insulin, are more likely to register. Therefore, while the NDSS's coverage of people with type 1 diabetes is good, some people with type 2 diabetes might not be registered. Also, Aboriginal and Torres Strait Islander (First Nations) people are under-represented in the NDSS.

The recent increase in linked data sets which combine administrative data from the Pharmaceutical Benefits Scheme (PBS), Medicare Benefits Schedule (MBS) and other health data provide an opportunity to further identify people with diagnosed diabetes. This work has the potential to supplement current data sources, that is the NDSS and the NHS, to provide a more comprehensive picture of the prevalence of diabetes in Australia. These linked data sets include the National Health Data Hub (NHDH), formerly known as National Integrated Health Services Information (NIHSI), the Person Level Integrated Data Asset (PLIDA) and Kidney and Diabetes Data Integration (KADDI). Finding approaches to identify people with diabetes based on markers of diabetes status such as diabetes-specific medicines and health services is important. However, some diabetes medicines can be prescribed for other conditions and health services such as HbA1c tests can be performed for monitoring blood glucose in the management of other conditions. Therefore, identifying and validating appropriate case definitions using diabetes status markers to identify people with diagnosed diabetes is important. Together with existing data sources, such as the NDSS, this will help improve the accuracy of identifying diagnosed diabetes for national monitoring and prevalence estimation.

This report presents findings from a 10% sample from the MedicineInsight data collection. MedicineInsight, a general practice data collection, is well suited for this study as it provides a longitudinal cohort where people with and without a recorded diagnosis of diabetes can be identified based on the existing MedicineInsight condition flags. Moreover, markers of diabetes status, such as diabetes-specific prescriptions, MBS billing items and pathology tests (for example HbA1c tests), important for creating case definitions for identifying diabetes, are also available in MedicineInsight. The diabetes definition from the MedicineInsight condition flag (standard definition) was considered as the reference standard

and case definitions using diabetes markers recorded in MedicineInsight were compared to this standard definition.

Aim of the project

The primary aim of this study was to identify and assess the performance of algorithms for diabetes case definitions using markers of diabetes status including diabetes-specific prescriptions, pathology tests and MBS billing services in MedicineInsight data. Specific objectives included to:

- determine the agreement between diabetes cases identified using definition algorithms based on markers of diabetes status and the MedicineInsight standard definition
- determine whether markers of diabetes status can be used to differentiate between type 1 and type 2 diabetes
- explore markers of diabetes status recorded prior to or after the date of diabetes diagnosis in a selected incident cohort.

Findings from this study can be used to inform other work that involves identifying people with diabetes in linked data sets such as the KADDI, which contain data collections like the PBS and the MBS that have similar diabetes markers. This information would be helpful in supplementing the existing data sources, such as the NDSS, in identifying people with diagnosed diabetes and improving the data available to regularly monitor the condition at the national level.

Structure of the report

This report has the following chapters:

1) Chapter 1 outlines the purpose and structure of the report.
2) Chapter 2 provides details of the methods including MedicineInsight data, study population and period, and measures of validity.
3) Chapter 3 describes the algorithms for diabetes case definitions using markers of diabetes status.
4) Chapter 4 describes the case definitions for identifying type 1 and type 2 diabetes.
5) Chapter 5 is an exploratory analysis showing the frequency of the diabetes markers recorded before and after diagnosis of diabetes in a subset of people newly diagnosed with diabetes.

2 Data source and methods

This analysis uses a 10% sample from the MedicineInsight data collection. Data for people with and without diabetes (based on the MedicineInsight condition flags), and data for markers of diabetes status were obtained from MedicineInsight. Access to the MedicineInsight data collection was obtained through Project 2020–005 – Improving Australian National Diabetes Estimates. The project was approved by the AIHW Ethics Committee on 5 June 2020 (Project number: EO2020/2/1152).

MedicineInsight

MedicineInsight is a primary care data collection containing de-identified electronic health records (EHRs) from Australian general practices (Busingye et al. 2019). MedicineInsight was initially established by NPS MedicineWise in 2011, with core funding from the Australian Government Department of Health, to collect general practice data to support quality improvement in Australian primary care and post-market surveillance of medicines. From January 2023, the Australian Commission on Safety and Quality in Health Care (ACSQHC) became the custodian of the MedicineInsight program.

Out of the approximately 10,000 general practices nationally, almost 700 were participating in MedicineInsight across Australia in 2020. MedicineInsight covers all states, territories and remote areas. Practices in South Australia are under-represented and practices in Tasmania are over-represented, but otherwise the distribution of MedicineInsight practices in each state is similar to the distribution of all practices in each state or territory. Compared with MBS data, MedicineInsight patients are representative of the Australian patient population in terms of age and sex.

MedicineInsight uses third-party data extraction tools – GeneRic Health Network Information Technology for the Enterprise (GRHANITE) and Precedence Health Care's Inca – which de-identify, extract and securely transmit whole-of-practice data from within each practice's clinical information system (CIS), either Best Practice (BP) or Medical Director (MD). A whole-of-practice data collection, containing all available historic and current EHRs, is conducted when a practice joins MedicineInsight. Fields potentially containing identifying information, such as progress notes and correspondence, are not included in the extraction. The extraction tool collects incremental data regularly, resulting in an updated longitudinal data collection in which patients attending each practice can be tracked over time.

Patient-level data are de-identified at source meaning patients' personal identifiers, such as name, date of birth and address, are not extracted, although year of birth and postcode are extracted to enable calculation of age, geographical location, remoteness and Socio-Economic Indexes for Areas. Extracted data include patient demographics (year of birth, sex, postcode) and clinical data entered directly by health care professionals (diagnosis, observations, tests performed, medicines prescribed). Each patient is assigned a unique number which allows all the records held in the database to be linked to the associated patient. Further information is available online, https://www.safetyandquality.gov.au/our-work/indicators-measurement-and-reporting/medicineinsight.

MedicineInsight data used in this analysis include conditions, prescriptions, pathology tests (HbA1c tests), MBS billing information and clinical encounters.

Conditions

People with diabetes mellitus (including type 1, type 2 and unspecified) were identified based on the MedicineInsight condition flags or HbA1c (glycated haemoglobin) test results (at least 2 HbA1c test results ≥ 6.5% or 48 mmol/mol). No consistent national classification system is used in general practice to code conditions, and each CIS has its own classification or coding system. MedicineInsight condition flags are developed by clinical coders and reviewed by medical advisers and indicate those records where the conditions of interest, or their relevant synonyms, are reported in MedicineInsight. Both coded (Docle, Pyefinch) and noncoded (free text) conditions are searched for in the 3 diagnosis fields – 'Diagnosis', 'Reason for visit' or 'Reason for prescription'. Relevant terms used in the condition flag for diabetes are shown in Appendix A Table A1. Records identified by a free text string alone are not automatically flagged but are individually reviewed by a clinical coder to determine whether the text string refers to the condition indicated or is present in another context (for example, a search for 'diabetes' may identify ‘partner died from diabetes'). Each record is flagged accordingly. Records indicating ‘suspected’, ‘query’ or ‘?’ records of the condition are not flagged as the condition, unless otherwise specified.

Study population

The study comprised about 206,000 people who visited a participating MedicineInsight general practice at least 3 times in the 2 years preceding the August 2020 MedicineInsight data download (regular patients) (Appendix B). A cohort of regularly attending (active) patients was used as they are more likely to have complete data if they receive most of their care at the MedicineInsight practice, thereby allowing sufficient opportunity for recording of diabetes diagnosis, prescriptions or health services.

About 13,700 people with diabetes mellitus ever recorded (from the earliest date of the medical record to August 2020, the data download date or end of study) were identified based on the condition flags, and a very small number (n=200) were identified from HbA1c test results. Therefore, a total of about 13,900 individuals with diabetes were included in the study, equating to a prevalence of about 6.8%. Diabetes cases included type 1, type 2 or unspecified diabetes, but excluded gestational diabetes due to the transient nature of the condition. Characteristics of people with diabetes recorded are presented in Appendix B. Those without any record of diabetes were assigned to the non-diabetes group.

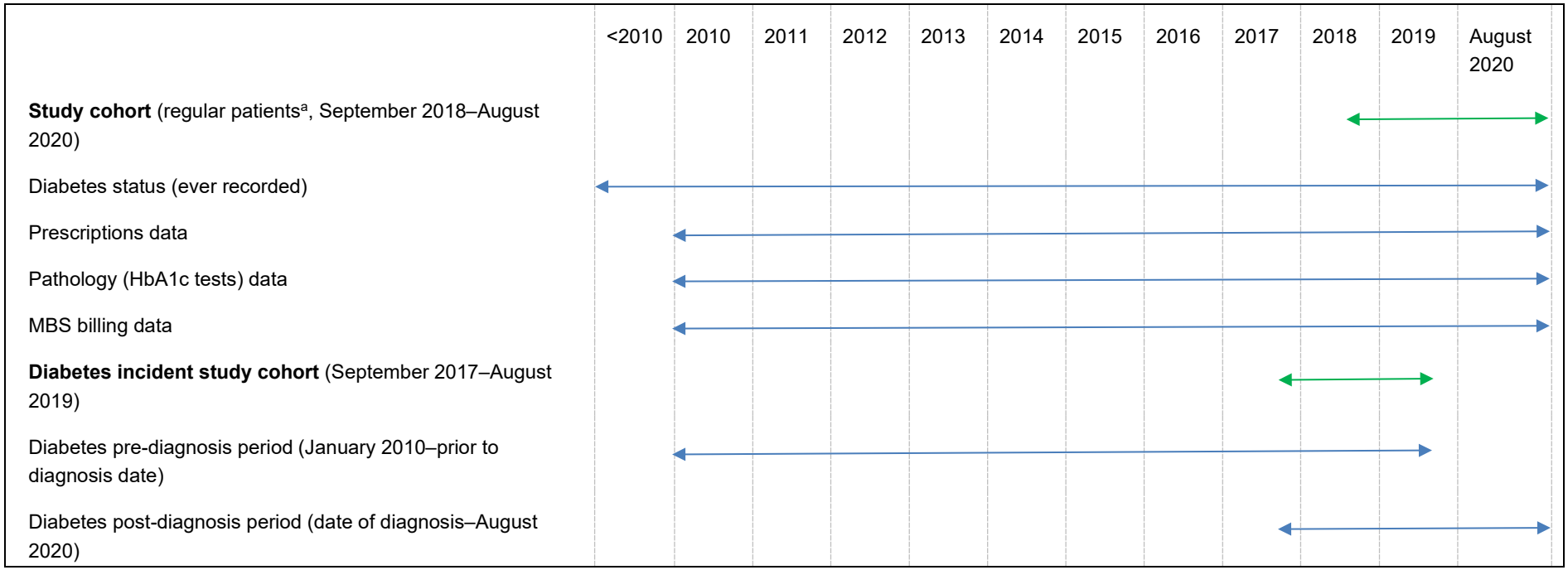
Study period

Data for identifying people with diabetes (reference standard) and those without diabetes in MedicineInsight were not restricted to any time period. All data from each individual’s earliest medical record to August 2020, the date of the data download, were assessed.

The study period for assessing markers of diabetes status in MedicineInsight, that is HbA1c tests, diabetes prescriptions and diabetes-related MBS service items, was from January 2010 to August 2020. Figure 2.1 shows the study period for the cohorts and the markers of

diabetes status assessed. Limiting the period for records of diabetes markers from January 2010 onwards was to account for potential limited use of clinical information systems by general practitioners in earlier years, which could affect completeness of the data.

Figure 2.1: Study period for the cohorts and the markers of diabetes status in MedicineInsight



[a] People who visited a participating MedicineInsight general practice at least 3 times in the 2 years preceding the August 2020 MedicineInsight data download.

Markers of diabetes status

Prescriptions

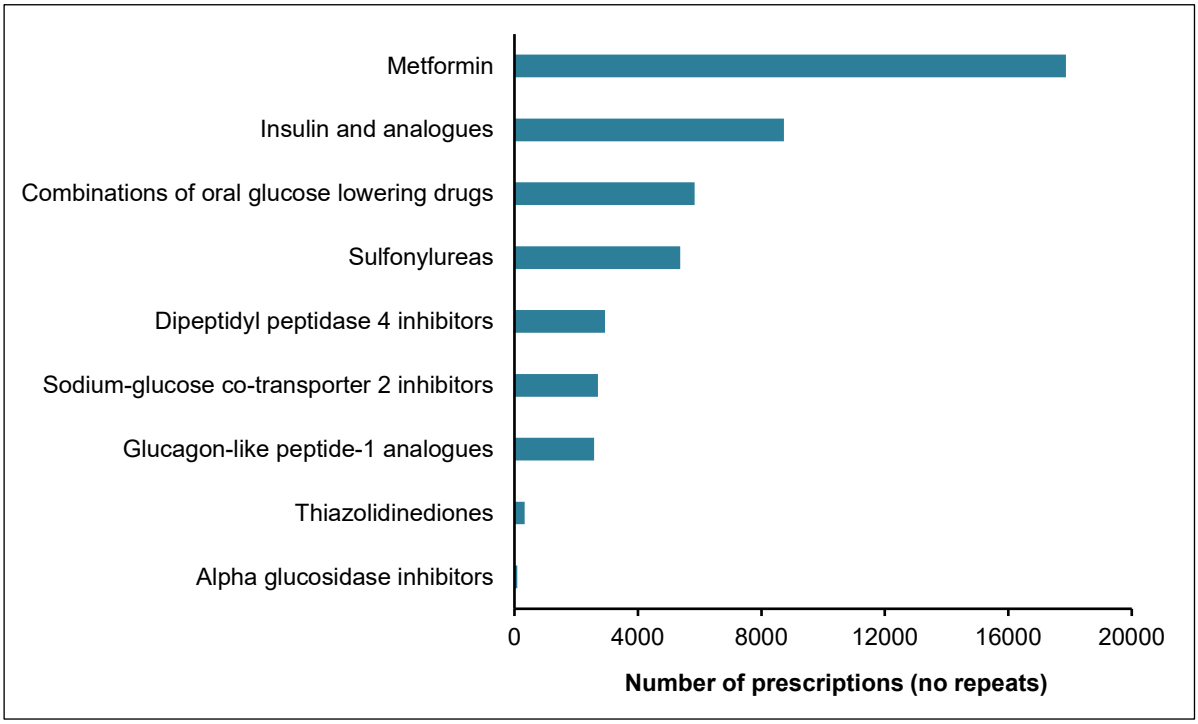
Prescription data in MedicineInsight consist mainly of medicines prescribed by general practitioners (GPs) but may also include medicines prescribed by specialists or hospital if recorded in the CIS by the GP. These prescriptions include medicines that are partly or wholly government rebated from the PBS and Repatriation PBS (RPBS), and also private (non-rebated) prescriptions. Private prescriptions are those paid for entirely by the patient or their private health insurer as they do not meet PBS/RPBS requirements related to the medicine prescribed, its indication for use, the amount supplied or the number of repeats.

Most medicines prescribed (77%) for the study cohort, between January 2010 and August 2020, were PBS subsidised. Similarly, a larger proportion of the diabetes prescriptions (96%) included in this analysis were PBS subsidised; 3% were private prescriptions (non-PBS subsidised) and the status was unknown for 1% of the diabetes prescriptions.

Prescriptions data contain the Anatomical Therapeutic Chemical (ATC) classification code which was used to identify medicines prescribed for diabetes treatment (ATC code: A10). For a few prescriptions where ATC codes were not available, the medicine active ingredient was used to identify diabetes-related prescriptions. Prescription data are available for both ‘issued’ (original script) prescriptions and the stated number of repeats recorded in the CIS. Issued prescriptions are used for this analysis. A list of diabetes medicines included in this analysis is provided in Appendix A Table A2.

From January 2010 to August 2020 there were just over 46,400 issued prescriptions (that is, excluding repeats) for any diabetes-related medicine recorded. Of the 206,000 people in the study cohort, about 14,000 (6.8%) were identified as having at least 1 diabetes prescription. The majority of the prescriptions were for metformin, followed by insulin and combinations of oral glucose lowering drugs (Figure 2.2).

Figure 2.2: Number of diabetes-related prescriptions (excluding repeats)



HbA1c tests from pathology results

Most Australian practices receive pathology test results electronically, transferred directly into the CIS from pathology providers. The pathology results table in MedicineInsight contains result values from specific pathology tests. Most of the common pathology test results are recorded using Logical Observation Identifiers Names and Codes (LOINC), and contain detailed results, often including whether the result is normal or abnormal depending on the normal ranges for that laboratory.

Glycated haemoglobin, haemoglobin A1c or HbA1c, is the main biomarker used to assess long-term glucose control in people living with diabetes. HbA1c tests were identified based on both LOINC and text search for HbA1c and other relevant synonyms. HbA1c data from the pathology results table were used in this analysis.

From January 2010 to August 2020, more than 200,000 HbA1c tests (for about 54,300 individuals) were recorded among the 206,000 people in the cohort.

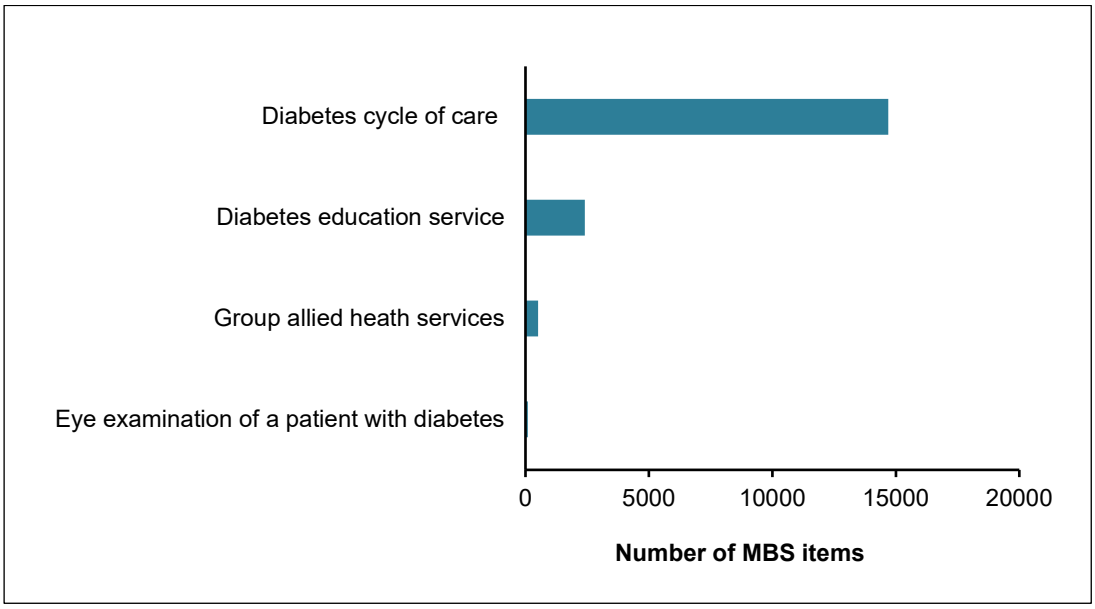
Medicare Benefits Schedule service items

The Medicare program (Medicare) provides access to Australian Government subsidised medical and hospital services listed through the MBS. MBS items in MedicineInsight listed for either established diabetes or for type 2 diabetes-specific services were included in the study (Appendix A Table A3). Of note, due to the incompleteness of the MBS billing data in MedicineInsight and to minimise duplicates the MBS items specific for HbA1c tests were not included in the analysis for diabetes-related MBS items because HbA1c tests from the pathology data were assessed.

MBS billing data are available in MedicineInsight when practices use integrated or compatible clinical and practice management software. Almost one-third of the MedicineInsight practices do not have billing data available, which might affect utility of diabetes-related MBS items for identifying diabetes cases.

About 17,700 diabetes-related MBS items (excluding HbA1c service items) were recorded for about 5,600 individuals in the study cohort over the period, from January 2010 to August 2020. The most recorded of these MBS service items were for annual diabetes cycle of care (Figure 2.3). The annual diabetes cycle of care consists of prevention and management activities for diabetes including patient education and self-care, medication review, health checks such as HbA1c, cholesterol, blood pressure, kidney health, weight, waist and body mass index, foot and eye assessments. Of note, the diabetes cycle of care items are no longer available on the MBS as of November 2022.

Figure 2.3: Number of diabetes-related Medicare services recorded



Measures of validity and reference standard

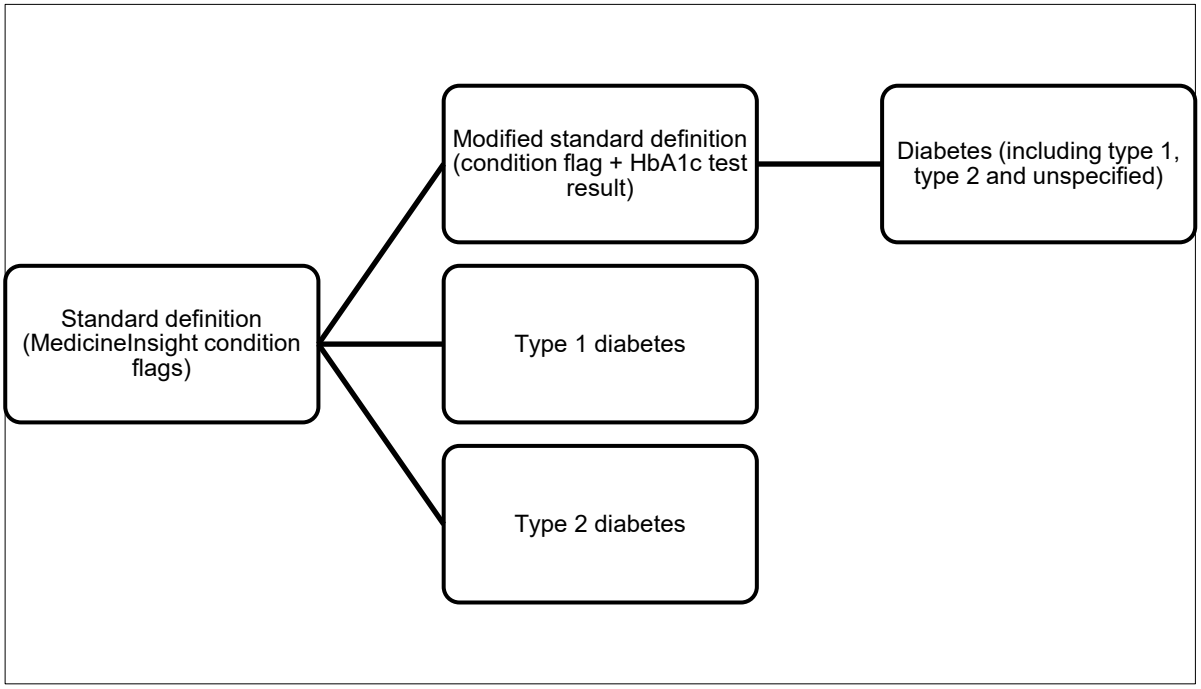
Reference standard

The gold (or reference) standard is the benchmark against which validity is determined. The reference standard is where the disease and non-disease status are known with a high level of certainty. In this case, no reference standard is available; however, the MedicineInsight data were considered as being a reasonable study population for this purpose as people with diabetes and those without diabetes can be identified from existing MedicineInsight condition flags (standard definition). Therefore, for this analysis, this formed the reference standard, hereafter called standard definition.

The standard definition for type 1 and type 2 diabetes was based on the MedicineInsight condition flag for each condition. However, for diabetes (including type 1, type 2 and unspecified) a modified standard definition was used where the MedicineInsight condition flag for diabetes and HbA1c test results (at least 2 HbA1c test results ≥ 6.5% or 48

mmol/mol) were considered the standard definition. Figure 2.4 shows the standard definition used for diabetes, type1 and type 2 diabetes.

Figure 2.4: Reference standard/standard definition for diabetes, type 1 and type 2 diabetes



Performance characteristics

Sensitivity, specificity, positive predictive value and negative predictive value are performance characteristics, which can be used to assess the validity of those markers being tested for identifying chronic conditions, which are derived from EHRs or administrative data. These measures are used to evaluate the ability of a disease marker to correctly detect cases of a disease and non-cases (that is, those without the disease) in the population. These measures of validity are typically calculated using a reference standard. Diabetes case definitions consisting of diabetes markers from MedicineInsight were compared to the MedicineInsight standard definition and case definitions could be met at any time during the study period (2010–2020). Performance characteristics including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and false identification rate (FIR) were reported.

Sensitivity (recall, or true positive rate) is defined as the proportion of people with diabetes, as identified by the MedicineInsight standard definition, who also have the marker of diabetes status (for example, prescriptions, HbA1c tests and MBS items) recorded.

Specificity (true negative rate) is the proportion of people without diabetes, as identified by the MedicineInsight standard definition, who do not have the diabetes marker recorded.

Positive predictive value (PPV, precision) is the proportion of people with the diabetes marker who also have diabetes as identified by the MedicineInsight standard definition.

Negative predictive value (NPV) is the proportion of people without the diabetes marker who do not have diabetes as identified by the MedicineInsight standard definition.

Of note, the interpretation of a positive or negative predictive value varies from setting to setting, according to the prevalence of the condition in the particular setting. For example, for a given sensitivity and specificity, PPV increases with increased diabetes prevalence while NPV decreases with increased prevalence.

False identification rate (FIR, or false discovery rate), which indicates the proportion of false positives among those who test positive (that is, the proportion of people misclassified as having diabetes by the marker), was also determined as an indication of misclassification risk, in some instances.

The formulae for calculating each measure of validity are shown in Box 2.1. Interpretation of performance characteristics is dependent on the context and purpose of the study. The values range from 0%, which is the lowest, to 100%, the highest. Thus, a perfect algorithm would have sensitivity, specificity, PPV and NPV values of 100%, and a FIR value of 0%.

Box 2.1 Calculating measures of validity

		Diabetes cases or non-diabetes controls identified through MedicineInsight condition flag (standard definition)	
		+	-
Diabetes status markers in MedicineInsight	+	A True Positive	B False Positive
	-	C False Negative	D True Negative

Sensitivity	=	A / (A + C)
Specificity	=	D / (B + D)
Positive predictive value (PPV)	=	A / (A + B)
Negative predictive value (NPV)	=	D / (C + D)
False identification rate (FIR)	=	B / (A + B) or 1 – PPV
Positive likelihood ratio (LR+)	=	sensitivity / (1 – specificity)
Negative likelihood ratio (LR-)	=	(1 – sensitivity) / specificity

Ideally, the optimal definition algorithm (single or combination of markers) will be able to identify a high proportion of the diabetes population (sensitivity) while minimising false positives or misclassification in the diabetes population (PPV). However, the main goal of this study is to minimise misclassification risk to as close to zero as possible without compromising the predictive power of the markers for detecting diabetes. A high PPV, implying low FIR, indicates low risk of misclassification. For this analysis, very high PPV (≥ 90%) was prioritised with sensitivity above 60% considered acceptable. Thus, sensitivity and PPV are our primary outcomes of interest.

We also estimated probabilities of having diabetes with (positive) and without (negative) meeting the definition for the suggested algorithms. Probabilities were calculated using Bayesian analyses, based on pre-test probability (which can be defined as prevalence) and likelihood ratios (LR), using the following formula (van Walraven et al. 2011):

O × LR / [(O × LR) + (1 – O)],

where O is the odds of disease in the study sample and LR is the positive [LR+] or negative [LR-] likelihood ratio of the algorithm (calculated as shown in Box 2.1 above).

Pre-test odds [O] are calculated as: prevalence / (1 – prevalence).

We sought algorithms that maximised PPV (≥ 90%) but with an acceptable sensitivity (> 60%) and were simple to apply (minimum requirements).

Data were presented in summary tables which included evaluation of markers as single use items and increasing uses, that is minimum thresholds from one to 6. Additionally, evaluation was undertaken to assess diabetes markers which had been used multiple times within a given time frame. For example, a prescription for a diabetes medicine within 6 months of a recording for another diabetes medicine.

Limitations

The data set is based on electronic health records collected to provide clinical care to a patient, and not for research purposes. All analyses are therefore dependent on the accuracy and completeness of data recorded in, and available for extraction from, the general practice CIS.

Conditions may be under-reported in the MedicineInsight data collection, depending on GP recording practices. For example, once chronic conditions are recorded in the medical record, and the patient is known to the GP, the GP may not routinely record the reason for prescribing, or the reason for the visit, at each visit. Moreover, for confidentiality reasons progress notes which may contain further information on diagnoses are not accessible. Some patients with diabetes may not have a recorded diagnosis in MedicineInsight and might get incorrectly assigned to the control group. Under-recording of diabetes would lead to low PPV estimates due to misclassification of people with diabetes. However, a validation study showed that the accuracy of condition definitions, including type 2 diabetes, in MedicineInsight is good (Havard et al. 2021).

The MedicineInsight standard definition was used as a reference standard in this analysis. This reference standard might not be perfect if there is incomplete recording of diabetes in the 3 MedicineInsight diagnosis fields – diagnosis, reason for visit or reason for prescription – used for the standard definition. A reference standard with limitations can introduce measurement error in the analysis and the performance of the definition algorithms would depend on the quality of the reference standard.

In Australia, people can visit multiple general practices for health services. Despite using a cohort of people who visited the same practice regularly in this study, it is possible that information on prescriptions, tests and MBS billing data might be incomplete if a patient visited another general practice or another health care setting such as a specialist or

hospital. Additionally, the incompleteness of billing data due to incompatibility of the clinical and billing software for some practices might affect sensitivity estimates for this diabetes marker in this study.

Not all markers of diabetes status in the MedicineInsight data collection are directly comparable to other administrative data sets. MedicineInsight is different from PBS administrative data as it contains prescriptions made by the GP both PBS subsidised and private (non-PBS subsidised), while prescriptions provided through other health care settings, such as specialists or hospitals, are not available.

Prescriptions may also be provided to a patient and never dispensed, resulting in no claim recorded in the PBS. This can lead to differences in validity measure estimates for diabetes case identification when using PBS and primary care data.

HbA1c tests from pathology results were used in this study but not the MBS items for HbA1c tests. Some of the HbA1c tests in the pathology table might be results provided to GPs through investigations undertaken by specialists or in the hospitals and are not captured as service items for the MBS. This can result in an increased sensitivity in the GP data which cannot be replicated using MBS administrative data. However, the incompleteness of the MBS billing data in MedicineInsight could limit the ability to identify people with diabetes using this diabetes marker. Also, the annual diabetes cycle of care MBS items ceased from November 2022, which might affect applicability of validity estimates for diabetes-related MBS items in data collections containing data for latest years.

The earliest date on which diabetes was recorded in one of the 3 diagnosis fields – diagnosis, reason for visit or reason for prescriptions – was defined as the first diagnosis date. This date might not be accurate for people who joined the MedicineInsight practice after their original date of diagnosis. Due to this, some patients may incorrectly be identified as part of the incident cohort with prescriptions or health services seemingly recorded before diabetes diagnosis.

These limitations might affect the accuracy of the results; however, the algorithms developed provide an effective approach for diabetes case identification and can be refined for other data sets with similar diabetes markers.

3 Diabetes case identification using markers of diabetes status

This chapter describes how well the relevant markers of diabetes status in MedicineInsight, that is prescriptions, HbA1c tests and MBS items (excluding HbA1c service items), can be used to correctly identify people with diabetes.

The data collection size, coverage and use of different data fields (diagnosis name, reason for prescription and reason for encounter) to identify diabetes for the standard definition make MedicineInsight ideal for validating algorithms for identifying diabetes using markers of diabetes status in primary care data.

The agreement between diabetes identified using diabetes marker definition algorithms and diabetes identified using the standard definition was assessed.

Diabetes case definitions

Identifying diabetes cases in large data collections is important for surveillance, research, quality improvement and clinical care. Diabetes case definitions can be created from single or combinations of information including diagnostic codes, diabetes-specific medications, service claims and laboratory results.

Several studies have examined the use of EHRs or administrative data to identify diabetes cases. Diabetes has been defined using the International Classification of Diseases (ICD) diabetes diagnostic codes from hospitalisation and physician data (Chen et al. 2010; Hirsch and Scheck McAlearney 2014; Nakhla et al. 2019; Williamson et al. 2014), pathology results indicating diabetes diagnosis and prescriptions for diabetes-related medicines including insulin and other glucose lowering drugs (Hirsch and Scheck McAlearney 2014; Rahimi et al. 2014; Williamson et al. 2014), utilisation of diabetes-specific service claims (Lipscombe et al. 2018), and reason for visit (Havard et al. 2021; Rahimi et al. 2014).

In a meta-analysis consisting of 6 studies, Leong and colleagues found a pooled sensitivity of 82% and 98% specificity for the commonly used administrative database definition for diabetes, that is 2 physician billing claims and/or one hospitalisation with a diabetes record within a 2-year period (Leong et al. 2013). These results demonstrate the utility of these methods for identifying diabetes cases with a minimum risk of misclassification. Examples of previously used algorithms for diabetes case definitions are provided in Box 3.1.

In the current study, similar approaches were adopted to validate algorithms (using markers of diabetes status including prescriptions, HbA1c tests and MBS items) for identifying diabetes against the MedicineInsight standard definition as the reference.

Box 3.1: Examples of published methods in literature which have been used to identify diabetes cases using electronic medical records or administrative data sources

Hospitalisation/physician/pathology data

- ≥ 1 or ≥ 2 ICD 9 and 10 codes within 1–2 (or more) years
- ≥ 1 or ≥ 2 biochemical tests indicating diabetes (e.g., HbA1c ≥ 6.5%)
- 4 physician claims or 1 hospitalisation (ICD codes) within 1 year
- 4 physician claims within 2 years

Health care service utilisation

- ≥ 2 plasma glucose tests within 1–2 (or more) years
- ≥ 1 or ≥ 2 diabetes-related health services within 1–2 (or more) years

Diabetes medicines utilisation (prescribed or dispensed)

- ≥ 1 or ≥ 2 diabetes-related medicine prescriptions (insulin and other glucose lowering agents)
- ≥ 1 or ≥ 2 insulin prescriptions
- ≥ 1 or ≥ 2 oral hypoglycaemic drug prescriptions (including and excluding metformin)

Methods

People with diabetes recorded in the MedicineInsight data collection up to August 2020 were identified based on the MedicineInsight condition flags (Appendix A Table A1) and a very small number were identified from HbA1c test results (at least 2 test results ≥ 6.5% or 48 mmol/mol) (modified standard definition). Diabetes cases included type 1, type 2 and unspecified diabetes, but excluded gestational diabetes due to the transient nature of the condition. Those without any record of diabetes were assigned to the non-diabetes group (controls) to assess measures of accuracy including sensitivity, specificity, PPV, NPV and FIR.

To assess markers of diabetes status, HbA1c tests, diabetes prescriptions and diabetes-related MBS items recorded in MedicineInsight from January 2010 to August 2020 were included. Some medicines which are used for other conditions requiring blood glucose control (for example polycystic ovarian syndrome), or which are indicated for prevention of pre-diabetes progressing to diabetes were also assessed, including metformin, rosiglitazone and acarbose (RACGP 2020).

Each diabetes marker was assessed for the total number of records across the study period, with the number of records stratified into minimum thresholds of one record within the study period through to 6 and above records in the period. This stratification was used to determine if increasing the use of these markers would be a better predictor of diabetes status and reduce the number of individuals falsely identified as having diabetes. Additionally, analysis of diabetes markers recorded within a certain time threshold, that is 6, 12, 18 and 24 months was conducted (see Box 3.2 for further details).

Box 3.2: Assessment approach and records thresholds

Analysis period and time between markers

Minimum number of records	Marker(s) recorded any time during the study period (Jan 2010–Aug 2020)	Marker(s) with a gap of 6 months to another similar marker recorded during the study period	Marker(s) with a gap of 12 months to another similar marker recorded during the study period	Marker(s) with a gap of 18 months to another similar marker recorded during the study period	Marker(s) with a gap of 24 months to another similar marker recorded during the study period
1	**A**				
2		**B**			
3			**C**		
4				**D**	
5					**E**
6					

For example:

A = Number of people who had a minimum of one diabetes prescription (or HbA1c test or MBS services or combination of markers) recorded during the study period.

B = Number of people who had a minimum of 2 diabetes prescriptions where each prescription has a gap of 6 months to another diabetes prescription during the study period.

C = Number of people who had a minimum of 3 diabetes prescriptions where each prescription has a gap of 12 months to another diabetes prescription during the study period.

D = Number of people who had a minimum of 4 diabetes prescriptions where each prescription has a gap of 18 months to another diabetes prescription during the study period.

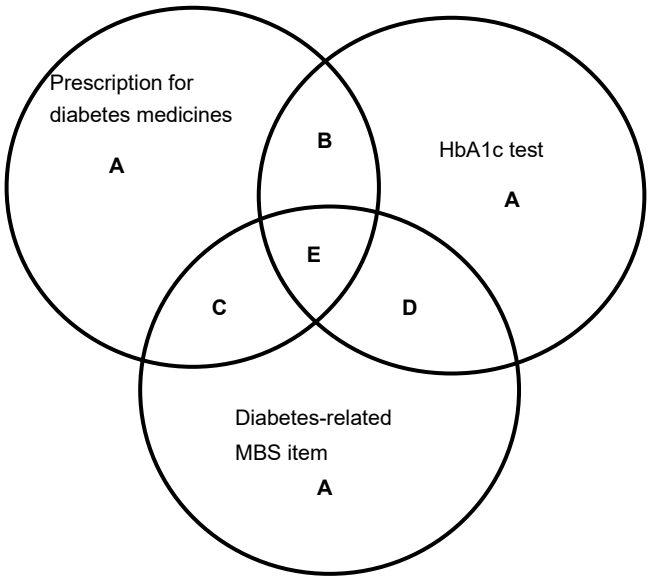
E = Number of people who had a minimum of 5 diabetes prescriptions where each prescription has a gap of 24 months to another diabetes prescription during the study period.

For the 3 markers of diabetes status (prescriptions, HbA1c tests and MBS services), a number of approaches were used to identify the best algorithm for identifying diabetes, namely (see Box 3.3 for illustration):

- single marker
- 2 markers combined using ‘and’ and ‘or’
- 3 markers combined using ‘and’ and ‘or’.

Combining diabetes markers using ‘or’ and ‘and’ was used to improve sensitivity and PPV estimates, respectively.

Box 3.3: Approaches used to identify diabetes case definitions



A = Each marker, diabetes prescriptions, Hba1c tests, diabetes-related MBS items, assessed individually (single marker).

B = Diabetes prescriptions and/or Hba1c tests (2 markers combined).

C = Diabetes prescriptions and/or diabetes-related MBS items (2 markers combined).

D = HbA1c tests and/or diabetes-related MBS items (2 markers combined).

E = Diabetes prescriptions and/or Hba1c tests and/or diabetes-related MBS items (3 markers combined).

Findings from validation of the algorithms for diabetes case definition

Single diabetes marker approach

When assessed individually, there were variations in sensitivity and PPV for the 3 markers of diabetes status, but very high specificity and NPV were observed for each marker. High specificity indicates that most people identified as not having diabetes by the MedicineInsight standard definition did not have a record of the diabetes marker. High NPV suggests a high probability of not having diabetes according to the standard definition for people who did not have a record of the diabetes marker. Sensitivity and PPV are the primary outcomes of interest and are the estimates mainly reported.

Prescriptions for diabetes medicines

Prescriptions for diabetes medicines including insulin and other glucose lowering drugs are an important marker for identifying people with diabetes in EHRs or administrative data sets. However, some diabetes medicines are used for other conditions which require blood glucose control, where there are other benefits such as weight loss; however, most of these medicines have limited applications outside treating diabetes. Information about the reason

for prescription is available in MedicineInsight, but this field is often incomplete, thus was not used to tease out the different indications for the medicines.

Results for diabetes prescriptions are presented in 2 ways: first, where people with prescriptions for all diabetes medicines are included in the analysis; second, where people prescribed only metformin are excluded from the analysis. Exclusion of people with prescriptions for metformin only was done to minimise the potential risk of misclassification as metformin can be prescribed for people with other conditions, for example polycystic ovarian syndrome.

Table 3.1 shows the misclassification risk and predictive power of diabetes status for people with at least one prescription for each individual diabetes medicine class. Medicines indicated for diabetes only showed low risk of misclassification demonstrated by very high PPV, equivalent to low FIR. The misclassification risk was high for glucagon-like peptide-1 analogues (GLP1a) (FIR about 30%) and metformin (FIR about 19%) which might reflect prescribing of these medicines for other conditions. Increasing off-label use of GLP1a for weight management, possibly via the private market (non-PBS), has been reported. Findings show that nearly 2 in 5 people (37%) with at least one record for GLP1a had at least one prescription of these medicines accessed via the private market. Therefore, caution should be exercised in identifying people with diabetes based on this class of medicines in primary care data.

Table 3.1: Misclassification risk and predictive power (%) of diabetes status based on number of prescriptions for individual diabetes medicine class

One or more prescriptions for each diabetes medicine class	With diabetes mellitus		Without diabetes mellitus		Sensitivity	PPV	FIR
	With prescription	Without prescription	With prescription	Without prescription			
Metformin	8,809	5,130	2,044	190,180	63.2	81.2	18.8
Insulin	3,268	10,671	123	192,101	23.5	96.4	3.6
Combinations of oral blood glucose lowering drugs	3,630	10,309	52	192,172	26.0	98.6	1.4
Sulfonylureas	3,456	10,483	54	192,170	24.8	98.5	1.5
Dipeptidyl peptidase 4 inhibitors	2,021	11,918	25	192,199	14.5	98.8	1.2
Sodium-glucose co-transporter 2 inhibitors	2,036	11,903	22	192,202	14.6	98.9	1.1
Glucagon-like peptide-1 analogues	1,151	12,788	486	191,738	8.3	70.3	29.7
Thiazolidinediones	231	13,708	<5	n.p.	1.7	98.7	1.3
Alpha glucosidase inhibitors	61	13,878	<5	n.p.	0.4	93.9	6.2

FIR: false identification rate; n.p.: not published; PPV: positive predictive value

Previous studies have used prescriptions for diabetes medicines, with or without other diabetes-related health service claims, to identify diabetes cases and showed good sensitivity and specificity using single and multiple prescriptions. Evidence from this study indicates good sensitivity and precision (both estimates at 81%) when a minimum of one diabetes prescription is recorded (Table 3.2). This suggests that most people identified with diabetes by the MedicineInsight standard definition had at least one diabetes prescription and also the majority of those with diabetes prescriptions had diabetes according to the standard definition. However, single prescriptions could represent a one-off attempt to control blood glucose and may not represent long-term diabetes treatment. When multiple

prescriptions (2 or more) are considered, findings show very high PPV (92%) but a decline in sensitivity of about 21 percentage points (from about 81% to 60%) is observed. Definitions based on multiple prescriptions might be required to minimise the risk of misclassifying people with diabetes but there is a trade-off with decline in sensitivity.

Misclassification risk can be mitigated further by using a threshold for the maximum gap (for example, 6 months) between prescriptions demonstrated by very high PPV (> 93%) (Table S1, available in the Data tables). However, a trade-off with decreased sensitivity (< 47%) was observed for analyses where thresholds of 6, 12, 18 and 24 months between prescriptions were assessed.

Table 3.2: Predictive power (%) of diabetes status based on number of prescriptions for any diabetes medicine

Number of prescriptions for any diabetes medicine	With diabetes mellitus		Without diabetes mellitus		Se	Sp	PPV	NPV
	With prescription	Without prescription	With prescription	Without prescription				
One or more	11,336	2,603	2,632	189,592	81.3	98.6	81.2	98.7
Two or more	8,306	5,633	706	191,518	59.6	99.6	92.2	97.1
Three or more	5,861	8,078	235	191,989	42.1	99.9	96.2	96.0
Four or more	4,272	9,667	88	192,136	30.7	100.0	98.0	95.2
Five or more	3,109	10,830	39	192,185	22.3	100.0	98.8	94.7
Six or more	2,301	11,638	16	192,208	16.5	100.0	99.3	94.3

NPV: negative predictive value; PPV: positive predictive value; Se: sensitivity: Sp: specificity

Consistent with potential misclassification of diabetes status due to metformin prescriptions, findings show that exclusion of people with prescriptions for only metformin resulted in a very high PPV (92%) implying low misclassification risk, but sensitivity was low (56%) (Table 3.3). The decrease in sensitivity is because metformin is the first-line pharmacotherapy treatment for diabetes and most people with diabetes are prescribed this medicine. Definition algorithms with high sensitivity or PPV can be useful depending on the study question to be addressed.

Table 3.3: Predictive power (%) of diabetes status based on number of prescriptions for any diabetes medicines (excluding people with only metformin prescriptions)

Number of prescriptions for any diabetes medicine (excluding people with only metformin)	With diabetes mellitus		Without diabetes mellitus		Se	Sp	PPV	NPV
	With prescription	Without prescription	With prescription	Without prescription				
One or more	7,851	6,088	731	191,493	56.3	99.6	91.5	96.9
Two or more	5,825	8,114	151	192,073	41.8	99.9	97.5	96.0
Three or more	3,963	9,976	35	192,189	28.4	100.0	99.1	95.1
Four or more	2,745	11,194	14	192,210	19.7	100.0	99.5	94.5
Five or more	1,929	12,010	6	192,218	13.8	100.0	99.7	94.1
Six or more	1,384	12,555	<5	192,223	9.9	100.0	99.9	93.9

NPV: negative predictive value; PPV: positive predictive value; Se: sensitivity: Sp: specificity

HbA1c tests

Most people with diabetes identified by the standard definition had at least one HbA1c test recorded, indicating very high sensitivity (92%). However, only about 1 in 4 people (24%) with one or more HbA1c tests were identified as having diabetes by the standard definition, which indicates very low PPV (Table 3.4). Increasing the minimum number of HbA1c tests recorded during the study period to 6 shows very high precision (91%) of identifying people with diabetes based on HbA1c tests, but this is associated with a decrease in sensitivity (56%).

Table 3.4: Predictive power (%) of diabetes status based on number of HbA1c tests recorded

Number of HbA1c tests	With diabetes mellitus		Without diabetes mellitus		Se	Sp	PPV	NPV
	With HbA1c test	Without HbA1c test	With HbA1c test	Without HbA1c test				
One or more	12,873	1,066	41,379	150,845	92.4	78.5	23.7	99.3
Two or more	11,610	2,329	15,497	176,727	83.3	91.9	42.8	98.7
Three or more	10,526	3,413	6,728	185,496	75.5	96.5	61.0	98.2
Four or more	9,479	4,460	3,156	189,068	68.0	98.4	75.0	97.7
Five or more	8,574	5,365	1,501	190,723	61.5	99.2	85.1	97.3
Six or more	7,812	6,127	773	191,451	56.0	99.6	91.0	96.9

NPV: negative predictive value; PPV: positive predictive value; Se: sensitivity: Sp: specificity

Using a threshold of 6 months as the maximum gap between each HbA1c tests showed improvements in PPV but sensitivity decreased (Table 3.5). A minimum of 2 HbA1c tests, each separated by 6 months, had a PPV of 92% and sensitivity of 57%.

Table 3.5: Predictive power (%) of diabetes status based on number of HbA1c tests each with 6 months gap of another HbA1c test

Number of HbA1c tests each with 6 months gap of another test	With diabetes mellitus		Without diabetes mellitus		Se	Sp	PPV	NPV
	With HbA1c test	Without HbA1c test	With HbA1c test	Without HbA1c test				
One or more	9,870	4,069	3,656	188,568	70.8	98.1	73.0	97.9
Two or more	7,983	5,956	734	191,490	57.3	99.6	91.6	97.0
Three or more	6,693	7,246	273	191,951	48.0	99.9	96.1	96.4
Four or more	5,703	8,236	122	192,102	40.9	99.9	97.9	95.9
Five or more	4,954	8,985	60	192,164	35.5	100.0	98.8	95.5
Six or more	4,340	9,599	37	192,187	31.1	100.0	99.2	95.2

NPV: negative predictive value; PPV: positive predictive value; Se: sensitivity: Sp: specificity

Findings suggest that use of HbA1c tests as a single marker for identifying diabetes cases is not recommended due to high risk of misclassification since HbA1c tests can be performed for people with other conditions that require blood glucose monitoring, such as chronic kidney disease and impaired glucose. However, using at least 2 HbA1c tests each with 6 months maximum gap of another HbA1c test could partly mitigate the risk of misclassification, but the sensitivity is low.

Diabetes-related MBS items

All diabetes-related MBS service items, except MBS items for HbA1c tests, were assessed as part of this diabetes marker (Appendix A Table A3). MBS items related to diabetes allied health services were combined in a single item. All diabetes-related MBS item categories were associated with low risk of misclassification demonstrated by very high PPV (Table 3.6). However, very low sensitivity was observed, likely a reflection of the incomplete MBS billing information in MedicineInsight – thus fewer people with diabetes had this information recorded.

Table 3.6: Misclassification risk and predictive power (%) of diabetes status based on number of diabetes-related MBS items (excluding MBS items for HbA1c) for each category

One or more diabetes-related MBS item	With diabetes mellitus		Without diabetes mellitus		Sensitivity	PPV	FIR
	With MBS item	Without MBS item	With MBS item	Without MBS item			
Diabetes cycle of care	5,054	8,885	125	192,099	36.3	97.6	2.4
Allied health services	790	13,149	79	192,145	5.7	90.9	9.1
Eye examinations	61	13,878	5	192,219	0.4	92.4	7.6

FIR: false identification rate; MBS: Medicare Benefits Schedule; PPV: positive predictive value

A minimum of one diabetes-related MBS item was associated with very high PPV (96%) but very low sensitivity (39%) (Table 3.7). Increasing the minimum number of diabetes-related MBS items above one resulted in minimal changes to the PPV, implying that evidence of a diabetes-related MBS item, when available, is a good indicator of diabetes. However, poorer sensitivity was observed with increase in the minimum number of diabetes-related MBS items.

Although the precision of diabetes-related MBS items to identify people with diabetes in this study is high, the sensitivity is very low, which could have been affected by the incompleteness of the MBS billing information in MedicineInsight.

Table 3.7: Predictive power (%) of diabetes status based on number of diabetes-related MBS items (excluding MBS items for HbA1c)

Number of diabetes-related MBS items	With diabetes mellitus		Without diabetes mellitus		Se	Sp	PPV	NPV
	With MBS item	Without MBS item	With MBS item	Without MBS item				
One or more	5,422	8,517	208	192,016	38.9	99.9	96.3	95.8
Two or more	3,586	10,353	38	192,186	25.7	100.0	99.0	94.9
Three or more	2,485	11,454	19	192,205	17.8	100.0	99.2	94.4
Four or more	1,779	12,160	12	192,212	12.8	100.0	99.3	94.1
Five or more	1,274	12,665	9	192,215	9.1	100.0	99.3	93.8
Six or more	883	13,056	6	192,218	6.3	100.0	99.3	93.6

MBS: Medicare Benefits Schedule; NPV: negative predictive value; PPV: positive predictive value; Se: sensitivity: Sp: specificity

Combined data approaches for diabetes case identification

Combined approaches using inclusion criteria for diabetes medicine prescriptions, MBS items and HbA1c tests were also assessed. Definition algorithms where diabetes markers

were combined as pairs or all 3 markers using 'or' and 'and' (Box 3.3) were compared with the standard definition. Similar to the single marker approach, very high specificity and NPV were observed for definition algorithms combining the diabetes markers.

Approach with 2 diabetes markers combined using 'and'

The combination of one or more prescriptions for diabetes medicines and 2 or more HbA1c tests showed 70% sensitivity and very high PPV (92%) (Table 3.8). Increasing the minimum number of both prescriptions and HbA1c tests improved the PPV, but sensitivity decreased. This algorithm appears to provide a good case definition for diabetes.

Using at least one diabetes prescription and a minimum of one HbA1c test each with 6 months gap to another HbA1c test had very high PPV (96%) and 61% sensitivity (Table S2). A similar algorithm using a 12-month threshold (instead of 6 months) for HbA1c tests maintained a high PPV of 95% with a higher sensitivity of 67% (Table S3). Similar findings were observed when gaps of 18 and 24 months between HbA1c tests were used (Tables S4 and S5).

Table 3.8: Sensitivity and PPV (%) based on number of prescriptions for any diabetes medicine and HbA1c tests

	Number of prescriptions (any diabetes medicine)	Number of HbA1c tests					
		One or more	Two or more	Three or more	Four or more	Five or more	Six or more
Sensitivity							
	One or more	76.7	70.0	64.2	58.3	53.2	48.7
	Two or more	57.1	53.4	50.0	45.9	42.6	39.5
	Three or more	41.1	39.3	37.5	35.2	33.1	31.0
	Four or more	30.2	29.5	28.4	27.1	25.7	24.4
	Five or more	22.1	21.8	21.2	20.4	19.5	18.7
	Six or more	16.4	16.2	16.0	15.5	14.9	14.4
PPV							
	One or more	87.4	92.1	94.5	96.5	97.7	98.5
	Two or more	94.6	96.4	97.3	98.4	98.9	99.2
	Three or more	97.2	98.0	98.5	99.1	99.4	99.6
	Four or more	98.5	99.0	99.2	99.4	99.6	99.7
	Five or more	99.0	99.3	99.4	99.6	99.7	99.8
	Six or more	99.4	99.6	99.6	99.7	99.8	99.9

PPV: positive predictive value

Combination of diabetes prescriptions and diabetes-related MBS items revealed almost perfect PPV (≥ 99%) but very low sensitivity (< 35%), likely affected by poor coverage of the MBS billing information in MedicineInsight (Table S6). Similar findings were observed for the combination of HbA1c tests and diabetes-related MBS items (Table S7).

Approach with 2 diabetes markers combined using 'or'

Table 3.9 shows the predictive power of the combination of prescriptions for diabetes medicines (including/excluding people with only metformin prescriptions) or HbA1c tests using different thresholds for minimum number of records. A minimum of 6 records of either diabetes prescriptions or HbA1c tests had the highest PPV (91%), and sensitivity of about

68%. Precision estimates were similar when people with only metformin prescriptions were excluded in the prescriptions data, but sensitivity decreased.

A minimum of 3 records of either diabetes prescriptions or HbA1c tests each with a 6-month gap to another HbA1c test had 69% sensitivity and very high PPV (94%) (Table S8). When people with only metformin prescriptions were excluded, a minimum of 2 records of either diabetes prescriptions or HbA1c tests each with 6 months gap of another HbA1c test had sensitivity of 71% and very high PPV (92%) (Table S9).

Table 3.9: Predictive power (%) of diabetes status based on combination of prescriptions for any diabetes medicine or HbA1c tests

	Sensitivity	PPV
Number of prescriptions (any diabetes medicine) OR HbA1c test		
One or more	97.0	24.2
Two or more	92.5	43.9
Three or more	86.1	62.2
Four or more	79.5	75.8
Five or more	73.4	85.2
Six or more	67.6	91.0
Number of prescriptions (excluding people with only metformin) OR HbA1c test		
One or more	95.5	24.2
Two or more	89.1	44.1
Three or more	81.9	62.5
Four or more	75.1	76.4
Five or more	68.3	86.0
Six or more	62.9	91.6

PPV: positive predictive value

A minimum of 2 records of either a diabetes prescription or diabetes-related MBS item had a sensitivity of about 70% and very high PPV (93%) (Table 3.10). Increasing the minimum number of records above 2 showed improvements in PPV, but sensitivity became poorer. When people with only metformin prescriptions were excluded, a minimum of one record of either a diabetes-related prescription or MBS item showed a 70% sensitivity and very high PPV (91%).

A minimum of 2 records of either diabetes prescriptions or diabetes-related MBS items, each with a gap of 6 months to another MBS item, had 60% sensitivity and 92% PPV (Table S10). Findings were similar when the 6-month threshold for MBS items was replaced with a 12-month threshold. While improvement in PPV (97%) was observed with exclusion of people with only metformin prescriptions, sensitivity (58%) decreased when one or more records of either diabetes prescriptions or MBS items within 6 months of another diabetes MBS item was assessed (Table S11).

Combination of HbA1c tests or diabetes-related MBS items was not associated with improved estimates for PPV and sensitivity (Table S12). However, when a 6-month gap between HbA1c tests was used, 2 or more records of either diabetes-related MBS items or HbA1c tests within 6 months of another test had a sensitivity of about 62% and PPV of 92% (Table S13).

Two or more records of either diabetes prescriptions or HbA1c tests with each marker having a 6-month threshold had a sensitivity and PPV of about 62% and 92%, respectively (Table S14).

Table 3.10: Predictive power (%) of diabetes status based on combination of diabetes-related prescriptions or MBS items (excluding MBS items for HbA1c)

	Sensitivity	PPV
Number of prescriptions (any diabetes medicine) OR MBS item		
One or more	86.4	81.1
Two or more	69.5	92.8
Three or more	54.7	96.7
Four or more	43.4	98.3
Five or more	34.6	99.0
Six or more	27.7	99.4
Number of prescriptions (excluding people with only metformin) OR any MBS item		
One or more	70.1	91.3
Two or more	55.6	97.6
Three or more	42.5	99.1
Four or more	32.9	99.4
Five or more	25.7	99.5
Six or more	20.3	99.8

MBS: Medicare Benefits Schedule; PPV: positive predictive value

Approach with 3 diabetes markers combined using 'and'

When the 3 markers of diabetes status were combined using and, almost perfect PPV estimates (> 99%) were observed, implying that almost all people who had the 3 diabetes markers recorded had diabetes according to the standard definition (Table S15). However, sensitivity was very low (< 34%), indicating that of those identified as having diabetes by the standard definition the proportion who had all the 3 diabetes markers recorded was small. The very low sensitivity in this analysis might reflect the incompleteness of the MBS data in this study leading to few people with diabetes having all the 3 diabetes markers recorded.

The findings suggest that this algorithm that combines the 3 diabetes markers may not be suitable for identifying people with diabetes in this data set.

Approach with 3 diabetes markers combined using 'or'

When the 3 diabetes markers were combined using or, a minimum of 6 records of either diabetes prescriptions or MBS items or HbA1c tests had very high PPV (91%) and 69% sensitivity (Table 3.11).

When a gap of 6 months between each similar marker was applied, at least 2 records of either diabetes prescriptions or MBS items or HbA1c tests were associated with a sensitivity of 63% and PPV of 92% (Table S16).

A minimum of 3 records of either diabetes prescriptions or HbA1c tests with a 6-month maximum threshold or diabetes-related MBS items with a 6-month maximum threshold had sensitivity and PPV of 69% and 94%, respectively (Table S17). When people with metformin prescriptions only were excluded from the analysis, at least 2 records of either diabetes prescriptions or HbA1c tests each separated by 6 months or MBS items each separated by 6 months had 71% sensitivity and 92% precision (Table S18).

Table 3.11: Predictive power (%) of diabetes status based on number of prescriptions for diabetes medicines or HbA1c tests or diabetes-related MBS items (excluding MBS items for HbA1c)

	Sensitivity	PPV
Number of prescriptions (any diabetes medicine) OR diabetes-related MBS items OR HbA1c tests		
One or more	97.1	24.1
Two or more	92.6	43.9
Three or more	86.5	62.2
Four or more	80.1	75.8
Five or more	74.2	85.2
Six or more	69.0	91.0
Number of prescriptions (excluding people with only metformin) OR diabetes-related MBS items OR HbA1c tests		
One or more	95.6	24.2
Two or more	89.4	44.1
Three or more	82.4	62.5
Four or more	76.0	76.4
Five or more	69.7	86.1
Six or more	64.7	91.6

MBS: Medicare Benefits Schedule; PPV: positive predictive value

Potential algorithms for diabetes case definition

Table 3.12 presents potential algorithms for diabetes case definition using markers of diabetes status in the MedicineInsight data collection. The specificity and NPV estimates for all the suggested algorithms were very high (> 97%). High specificity indicates that most people identified as not having diabetes by the standard definition were also identified as not having diabetes by the diabetes markers definition algorithms. High NPV implies that most people identified as not having diabetes by the diabetes markers definition algorithm did not have diabetes according to the standard definition.

The case definition with a minimum of one diabetes prescription and at least one HbA1c test each with a gap of 6 months to another HbA1c test had the highest precision (96%) and sensitivity of 61%. Using this algorithm, the positive probability of diabetes was 96% (probability that a person meeting this definition has diabetes) and the negative probability was 3% (probability that a person not meeting this definition has diabetes).

The second highest precision of 95% and sensitivity of 67% was observed for the algorithms with at least one diabetes prescription and at least one HbA1c test each with a 12-month gap to another HbA1c test (positive probability 95%, negative probability 3%).

The algorithm with the highest sensitivity (71%) while maintaining precision above 90% was at least 2 records of either diabetes prescriptions (excluding people with only metformin) or HbA1c tests each with a 6-month gap to another HbA1c test (positive probability 92%, negative probability 2%). Two algorithms had a sensitivity of 70% with other performance estimates similar to the previous definition, namely:

- a minimum of one diabetes prescription and at least 2 HbA1c tests
- a minimum of 2 records with either diabetes prescriptions or diabetes-related MBS items.

For these 2 algorithms, the probability that a person meeting each definition has diabetes is 93% (positive probability) and the probability of having diabetes without meeting each definition is 2% (negative probability).

Table 3.12: Validation of diabetes case definition algorithms against people identified with diabetes using MedicineInsight standard definition; all study population (prevalence 6.8%), 2010 to 2020

Algorithm for diabetes case identification	Performance characteristics (%)					
	Se	sp	PPV	NPV	Pr+	Pr-
One or more diabetes prescriptions AND at least one HbA1c test each with 6 months gap of another HbA1c test	61.1	99.8	96.3	97.3	96.0	3.0
One or more diabetes prescriptions AND at least one HbA1c test each with 12 months gap of another HbA1c test	67.0	99.7	94.5	97.7	94.6	2.5
Three or more records with either diabetes prescriptions OR HbA1c tests each with 6 months gap of another HbA1c test	68.6	99.7	94.1	97.8	94.7	2.4
One or more diabetes prescriptions AND at least two HbA1c tests	70.0	99.6	92.1	97.9	93.2	2.3
Two or more records with either diabetes prescriptions OR diabetes-related MBS item	69.5	99.6	92.8	97.8	93.2	2.4
Two or more records with either diabetes prescriptions (excluding people with only metformin) OR HbA1c tests each with 6 months gap of another HbA1c test	71.2	99.5	91.6	98.0	91.8	2.2

MBS: Medicare Benefits Schedule; NPV: negative predictive value; PPV: positive predictive value; Pr+: positive probability, probability of having diabetes with the algorithm; Pr-: negative probability, probability of having diabetes without the algorithm; se: sensitivity; sp: specificity.

4 Diabetes type ascertainment

This chapter examines whether the 3 markers of diabetes status – prescriptions, HbA1c tests and MBS service items – recorded in MedicineInsight can be used to correctly identify people with diagnosed type 1 or type 2 diabetes.

The agreement between type 1 or type 2 diabetes identified using diabetes marker definition algorithms and type 1 or type 2 diabetes identified using the standard definition was assessed.

Type 1 and type 2 diabetes case definitions

The ability to distinguish accurately between type 1 and type 2 diabetes is important for providing accurate estimates for the burden of each type and for clinical quality improvement, given the difference in treatment approaches between the 2 conditions. Miscoding, misclassification, and misdiagnosis issues have been reported in identifying diabetes type in health records, therefore identification and classification of cases can be challenging (RCGP 2011).

A number of validated algorithms distinguishing type 1 and type 2 diabetes have been developed using EHRs or administrative data. These algorithms usually rely on diagnostic codes such as ICD 9/10 (Klompas et al. 2013; Sharma et al. 2016; Teltsch et al. 2019; Wells et al. 2020), demographic characteristics (age of diabetes onset) and prescriptions for diabetes medicines (Ke et al. 2020; Lethebe et al. 2019; Ng et al. 2008; Vanderloo et al. 2012; Weisman et al. 2020).

While type 1 diabetes must be treated with insulin, type 2 diabetes can be managed with lifestyle modification, insulin, or other glucose lowering medications. However, type 1 diabetes patients typically require insulin treatment within a few months of diagnosis. Additionally, type 1 diabetes usually develops during childhood or adolescence while type 2 diabetes typically occurs in adulthood and becomes more common with increasing age. However, recent evidence shows that type 2 diabetes is increasingly diagnosed at younger ages. Given that both age at diagnosis and medication use separately cannot definitively differentiate between people with type 1 and type 2 diabetes, a combination of the 2 is often used to distinguish between the 2 types. Box 4.1 shows some examples of algorithms that have been used for type 1 and type 2 diabetes case definitions.

While type 1 diabetes affects people of all ages, to assess whether the 3 markers of diabetes status can be used to identify people with type 1 diabetes the analysis was limited to people aged under 35. This was done to improve the predictive power of the diabetes markers to identify people with type 1 diabetes since age is widely used as one of the discriminative features between type 1 and type 2 diabetes (RCGP 2011). The analysis for type 2 diabetes included all age groups. Therefore, these assessments are independent, and findings should be interpreted separately for each type of diabetes.

Box 4.1: Examples of published methods in literature which have been used to identify type 1 or type 2 diabetes cases using electronic medical records or administrative data sources

Type 1 diabetes

- ≥ 1 or ≥ 2 type 1 diabetes specific diagnostic codes e.g., ICD 9/10
- ≥ 1 prescription for fast-acting insulin and analogues for injection
- Age at diagnosis < 30 or < 35 years and continual insulin within 6/12 months of diagnosis
- Age at diagnosis ≥ 30 or ≥ 35 years and continual insulin treatment from diagnosis
- A diagnostic code of type 1 diabetes only and prescription for insulin only
- Ratio of type 1 to type 2 diabetes codes ≥ 4 and at least 1 insulin prescription within 90 days
- Continual insulin treatment starting within 1 month of diagnosis
- Age at first diagnosis (< 22 or < 30 years)

Type 2 diabetes

- ≥ 1 or ≥ 2 type 2 diabetes-specific diagnostic codes e.g., ICD 9/10
- Taking oral anti-diabetic medicines or not taking insulin
- Age at diagnosis < 30 or < 35 years and not on continual insulin treatment within 6/12 months of diagnosis
- Diagnosis ≥ 30 or ≥ 35 years and not on continual insulin treatment from diagnosis
- A diagnostic code for type 2 diabetes only and prescriptions for other antidiabetic agents with or without insulin
- A diagnostic code for diabetes and prescription for other antidiabetic agents with no insulin prescription

Methods

Similar markers of diabetes status (prescriptions, HbA1c tests and MBS items) recorded in MedicineInsight as used for diabetes mellitus were assessed for type 1 and type 2 diabetes separately. Similar methods and inclusion criteria for these markers were used.

People with type 1, type 2 and/or unspecified diabetes recorded in MedicineInsight up to August 2020 were identified based on the MedicineInsight condition flags (standard definition). Individuals were defined as having type 2 diabetes if they had at least one record of type 2 diabetes in their medical record or at least one record of only unspecified diabetes in accordance with the previously validated definition of type 2 diabetes in MedicineInsight (Havard et al. 2021). Individuals were defined as having type 1 diabetes if they had at least one record of type 1 diabetes in their medical record. Individuals who had both type 1 and type 2 diabetes recorded were not included in either of these 2 types of diabetes.

Of about 13,900 people with diabetes, nearly 12,500 were identified as having type 2 diabetes, about 860 had type 1 diabetes and the diabetes type was not clear for nearly 580 people. Characteristics of people with record of type 2 diabetes are presented in Appendix B. For either type 1 or type 2 diabetes, people not assigned to the group were considered as non-cases (non-type 1 or non-type 2 diabetes group) to assess the performance of algorithms including sensitivity, specificity, PPV, NPV and FIR.

As already mentioned, to improve the predictive power of type 1 diabetes status the analysis was limited to people aged under 35. The number of people aged under 35 in this study was

almost 84,600 and about 340 were identified as having type 1 diabetes, a prevalence of about 0.4%. The characteristics of people aged under 35 and those with record of type 1 diabetes are presented in Appendix B.

As done previously, each diabetes marker was assessed for the total number of records across the study period, with the number of records stratified into minimum thresholds of one record within the study period through to greater than 6 records (Box 3.2). Analysis of diabetes markers recorded within a certain time threshold, that is 6, 12, 18 and 24 months was also conducted. Similar approaches used to identify the definition algorithm for diabetes were utilised, including a single marker or combination of markers (Box 3.3).

Type 1 or type 2 diabetes identified through definition algorithms using diabetes markers was compared with type 1 or type 2 diabetes identified using the MedicineInsight standard definition.

Findings from validation of the algorithms for type 2 diabetes case definition

Single diabetes marker approach

Prescriptions for diabetes medicines

Table 4.1 shows the risk of misclassification (FIR) and predictive power of type 2 diabetes status for each diabetes medicine class. Medicines indicated for only type 2 diabetes showed low risk of misclassification. The risk of type 2 diabetes misclassification was high for glucagon-like peptide-1 analogues (FIR 36%), insulin (FIR 35%) and metformin (FIR 23%) which reflects prescribing of these medicines for other conditions. Although some patients with type 2 diabetes are treated with insulin, it is the main treatment for type 1 diabetes, and this is consistent with the observed high misclassification risk. The high misclassification risk of glucagon-like peptide-1 analogues might reflect the increasing off-label use of this class of medicines for weight management, possibly via private market access.

Table 4.1: Misclassification risk and predictive power (%) of type 2 diabetes status based on number of prescriptions for individual diabetes medicines

One or more prescriptions for each diabetes medicine class	With type 2 diabetes		Without type 2 diabetes		Sensitivity	PPV	FIR
	With prescription	Without prescription	With prescription	Without prescription			
Metformin	8,357	4,140	2,496	191,170	66.9	77.0	23.0
Combinations of oral blood glucose lowering drugs	3,447	9,050	235	193,431	27.6	93.6	6.4
Sulfonylureas	3,300	9,197	210	193,456	26.4	94.0	6.0
Insulin	2,203	10,294	1,188	192,478	17.6	65.0	35.0
Dipeptidyl peptidase 4 inhibitors	1,908	10,589	138	193,528	15.3	93.3	6.7
Sodium-glucose co-transporter 2 inhibitors	1,885	10,612	173	193,493	15.1	91.6	8.4
Glucagon-like peptide-1 analogues	1,048	11,449	589	193,077	8.4	64.0	36.0
Thiazolidinediones	214	12,283	20	193,646	1.7	91.5	8.6
Alpha glucosidase inhibitors	55	12,442	10	193,656	0.4	84.6	15.4

FIR: false identification rate; PPV: positive predictive value

Among people identified as having type 2 diabetes by the standard definition, 81% had at least one diabetes prescription recorded, indicating high sensitivity (Table 4.2). Of the people with at least one diabetes prescription, about 73% had type 2 diabetes as identified by the standard definition, suggesting a high potential for misclassification risk.

Table 4.2: Predictive power (%) of type 2 diabetes status based on number of prescriptions for any diabetes medicines

Number of prescriptions for any diabetes medicine	With type 2 diabetes		Without type 2 diabetes		Se	Sp	PPV	NPV
	With prescription	Without prescription	With prescription	Without prescription				
One or more	10,123	2,374	3,845	189,821	81.0	98.0	72.5	98.8
Two or more	7,256	5,241	1,756	191,910	58.1	99.1	80.5	97.3
Three or more	5,144	7,353	952	192,714	41.2	99.5	84.4	96.3
Four or more	3,741	8,756	619	193,047	29.9	99.7	85.8	95.7
Five or more	2,741	9,756	407	193,259	21.9	99.8	87.1	95.2
Six or more	2,019	10,478	298	193,368	16.2	99.9	87.1	94.9

NPV: negative predictive value; PPV: positive predictive value; Se: sensitivity: Sp: specificity

As has been done by other investigators, exclusion of people with only insulin prescriptions from the analysis to minimise misclassification revealed sensitivity of 79% and a slight increase in PPV (from 73% to 76%) (Table 4.3). For both analyses (including/excluding people with only insulin prescriptions), increasing the minimum number of prescriptions required to identify type 2 diabetes increased precision, but there was a trade-off with decreased sensitivity.

High PPV (> 84%) and low sensitivity (< 40%) were observed for analyses where thresholds of 6 and 12 months between prescriptions were assessed (Table S19). These findings suggest that use of diabetes prescriptions solely as a case definition for type 2 diabetes might not be suitable in these data.

Table 4.3: Predictive power (%) of type 2 diabetes status based on number of prescriptions for any diabetes medicines (excluding insulin)

Number of prescriptions for diabetes medicines (excluding insulin)	With type 2 diabetes		Without type 2 diabetes		Se	Sp	PPV	NPV
	With prescription	Without prescription	With prescription	Without prescription				
One or more	9,923	2,574	3,098	190,568	79.4	98.4	76.2	98.7
Two or more	6,816	5,681	1,055	192,611	54.5	99.5	86.6	97.1
Three or more	4,573	7,924	488	193,178	36.6	99.8	90.4	96.1
Four or more	3,166	9,331	274	193,392	25.3	99.9	92.0	95.4
Five or more	2,216	10,281	175	193,491	17.7	99.9	92.7	95.0
Six or more	1,578	10,919	124	193,542	12.6	99.9	92.7	94.7

NPV: negative predictive value; PPV: positive predictive value; Se: sensitivity: Sp: specificity

HbA1c tests

While about 93% of people identified as having type 2 diabetes by the standard definition had at least one HbA1c test recorded during the study period (sensitivity), the precision was very low, with only 21% of those with one or more HbA1c tests having type 2 diabetes as identified by the standard definition (Table 4.4). The highest PPV, of about 82% with a sensitivity of 57%, was observed when a minimum of 6 HbA1c tests was recorded. This finding suggests that HbA1c tests individually might not be suitable for identifying type 2 diabetes.

Table 4.4: Predictive power (%) of type 2 diabetes status based on number of HbA1c tests

Number of HbA1c tests	With type 2 diabetes		Without type 2 diabetes		Se	Sp	PPV	NPV
	With HbA1c test	Without HbA1c test	With HbA1c test	Without HbA1c test				
One or more	11,590	907	42,662	151,004	92.7	78.0	21.4	99.4
Two or more	10,451	2,046	16,656	177,010	83.6	91.4	38.6	98.9
Three or more	9,503	2,994	7,751	185,915	76.0	96.0	55.1	98.4
Four or more	8,572	3,925	4,063	189,603	68.6	97.9	67.8	98.0
Five or more	7,755	4,742	2,320	191,346	62.1	98.8	77.0	97.6
Six or more	7,069	5,428	1,516	192,150	56.6	99.2	82.3	97.3

NPV: negative predictive value; PPV: positive predictive value; Se: sensitivity: Sp: specificity

Diabetes-related MBS items

Except for some allied health service items which are specific for people with type 2 diabetes, most diabetes-related MBS items can be used by people with either type 1 or type 2 diabetes; thus, all diabetes-related MBS items were combined for these analyses. As mentioned above, for this analysis HbA1c tests were excluded from the diabetes-related MBS items as HbA1c tests from the pathology table are used in this analysis.

Among people with type 2 diabetes based on the standard definition, about 2 in 5 (40% sensitivity) had at least one diabetes-related MBS item recorded. Of this group, about 88% had type 2 diabetes, suggesting high PPV (Table 4.5). Although the assessed diabetes-related MBS items have very high PPV (> 88%), due to low sensitivity (< 40%), alone this diabetes marker does not appear to be suitable for identifying people with type 2 diabetes in this data set.

Table 4.5: Predictive power (%) of type 2 diabetes status based on number of diabetes-related MBS items (excluding MBS items specific for HbA1c tests)

Number of diabetes-related MBS items	With type 2 diabetes		Without type 2 diabetes		Se	Sp	PPV	NPV
	With MBS item	Without MBS item	With MBS item	Without MBS item				
One or more	4,958	7,539	672	192,994	39.7	99.7	88.1	96.2
Two or more	3,269	9,228	355	193,311	26.2	99.8	90.2	95.4
Three or more	2,274	10,223	230	193,436	18.2	99.9	90.8	95.0
Four or more	1,622	10,875	169	193,497	13.0	99.9	90.6	94.7
Five or more	1,165	11,332	118	193,548	9.3	99.9	90.8	94.5
Six or more	804	11,693	85	193,581	6.4	100.0	90.4	94.3

MBS: Medicare Benefits Schedule; NPV: negative predictive value; PPV: positive predictive value; Se: sensitivity: Sp: specificity

Combined data approaches

Combined approaches using inclusion criteria for diabetes prescriptions, MBS items and HbA1c tests were assessed. Type 2 diabetes definition algorithms where diabetes markers were combined as pairs or all 3 markers using ‘or’ as well as ‘and’ (as shown in Box 3.3) were compared to the standard definition of type 2 diabetes.

While there were variations in sensitivity and PPV for the different combinations of the 3 diabetes markers, very high specificity and NPV were observed for the type 2 definition algorithms.

Approach with 2 diabetes markers combined using ‘and’

The combination of at least one diabetes prescription where people with only insulin prescriptions are excluded, and at least one HbA1c test, had a 76% sensitivity and high PPV (82%) (Table 4.6). In the analysis with all diabetes prescriptions included, combination of at least one prescription and a minimum of 2 HbA1c tests showed reduced sensitivity (70%) but relatively similar PPV (83%) (Table S20). For both analyses, increasing the minimum number of both prescriptions and HbA1c improved the PPV, but sensitivity decreased.

Using at least one diabetes prescription where people with only insulin prescriptions are excluded and a minimum of one HbA1c test, each with a 6-month gap to another HbA1c test, showed very high precision (91%) and sensitivity of 61% (Table S21). Applying a maximum gap of 12 months instead of 6 months between HbA1c tests showed a 5-percentage point increase for sensitivity (from 61% to 66%) and a slight decrease for PPV (from 91% to 89%) (Table S22).

Table 4.6: Sensitivity and PPV (%) based on number of prescriptions for diabetes medicine (excluding people with only insulin prescriptions) and HbA1c tests

	Number of prescriptions (excluding people with only insulin)	Number of HbA1c tests					
		One or more	Two or more	Three or more	Four or more	Five or more	Six or more
Sensitivity							
	One or more	75.5	69.1	63.6	57.9	52.8	48.3
	Two or more	53.1	50.1	47.3	43.7	40.7	37.7
	Three or more	36.0	34.9	33.7	31.8	30.0	28.2
	Four or more	25.2	24.8	24.2	23.2	22.1	21.0
	Five or more	17.6	17.5	17.1	16.6	16.0	15.4
	Six or more	12.6	12.5	12.3	12.0	11.6	11.3
PPV							
	One or more	82.3	86.6	88.8	90.7	91.7	92.3
	Two or more	89.1	90.7	91.5	92.5	92.8	92.9
	Three or more	91.5	92.1	92.6	93.1	93.2	93.3
	Four or more	92.6	93.1	93.3	93.4	93.4	93.5
	Five or more	93.1	93.4	93.4	93.5	93.5	93.4
	Six or more	92.9	93.1	93.0	93.0	93.1	92.9

PPV: positive predictive value

Combination of diabetes prescriptions where people with only insulin were excluded and diabetes-related MBS items had very high PPV (≥ 90%) but the sensitivity was very low (< 34%) (Table S23). Similar findings were observed for the combination of HbA1c tests and diabetes-related MBS items (Table S24).

Approach with 2 diabetes markers combined using ‘or’

Findings for the combination of diabetes prescriptions (including/excluding people prescribed only insulin) or HbA1c tests using different minimum thresholds for records are shown in Table 4.7. A minimum of 6 records of either diabetes prescriptions or HbA1c tests had high PPV, nearly 82%, and sensitivity about 68%. Almost similar estimates were observed when people with only insulin prescriptions were excluded from the analysis. For both analyses, increasing the minimum number of records of either diabetes prescriptions or HbA1c tests from one to 6 led to an increase in precision, but sensitivity declined.

A minimum of 3 records of either diabetes prescriptions with people prescribed only insulin excluded or HbA1c tests each with a 6-month maximum threshold was associated with high PPV (86%) and sensitivity about 67% (Table S25).

Table 4.7: Predictive power (%) of type 2 diabetes status based on number of prescriptions for diabetes medicines or HbA1c tests

	Sensitivity	PPV
Number of prescriptions (any diabetes medicine) OR HbA1c tests		
One or more	97.0	21.6
Two or more	92.3	39.3
Three or more	86.1	55.8
Four or more	79.6	68.0
Five or more	73.5	76.5
Six or more	67.8	81.9
Number of prescriptions (excluding people with only insulin) OR HbA1c tests		
One or more	96.7	21.7
Two or more	91.5	39.4
Three or more	85.1	56.1
Four or more	78.4	68.6
Five or more	72.4	77.4
Six or more	66.8	82.9

PPV: positive predictive value

When diabetes prescriptions or diabetes-related MBS items were combined, 2 or more records of either of the 2 markers had 69% sensitivity and high PPV (82%) (Table 4.8). Excluding people with only insulin prescriptions revealed an increase in PPV (87%) but a decrease in sensitivity (66%) for a similar minimum number of records.

Table 4.8: Predictive power (%) of type 2 diabetes status based on number of prescriptions for diabetes medicines or diabetes-related MBS items (excluding MBS items for HbA1c)

	Sensitivity	PPV
Number of prescriptions (any diabetes medicine) OR diabetes related MBS items		
One or more	86.4	72.7
Two or more	68.7	82.2
Three or more	54.4	86.2
Four or more	43.2	87.7
Five or more	34.5	88.5
Six or more	27.6	88.8
Number of prescriptions (excluding people with only insulin) OR diabetes-related MBS items		
One or more	85.3	75.4
Two or more	66.2	86.6
Three or more	51.1	90.6
Four or more	39.7	91.9
Five or more	31.2	92.6
Six or more	24.6	92.9

MBS: Medicare Benefits Schedule; PPV: positive predictive value

Approach with 3 diabetes markers combined using ‘and’

Combining diabetes prescriptions (including/excluding only insulin prescription) and HbA1c tests and diabetes-related MBS items resulted in high PPV (> 86%), but very low sensitivity (< 35%).

Approach with 3 diabetes markers combined using 'or'

A minimum of 6 records of either diabetes prescriptions or MBS items or HbA1c tests had high PPV (82%) and sensitivity of about 69% (Table 4.9). Similar results were observed when people with only insulin prescriptions were excluded.

Table 4.9: Predictive power (%) of type 2 diabetes status based on number of prescriptions for diabetes medicines or diabetes-related MBS items or HbA1c tests

	Sensitivity	PPV
Number of prescriptions (any diabetes medicine) OR HbA1c tests OR diabetes-related MBS items		
One or more	97.1	21.6
Two or more	92.4	39.3
Three or more	86.5	55.8
Four or more	80.2	68.1
Five or more	74.4	76.5
Six or more	69.3	81.9
Number of prescriptions (excluding people with only insulin) OR HbA1c tests OR diabetes-related MBS items		
One or more	96.8	21.6
Two or more	91.6	39.4
Three or more	85.5	56.1
Four or more	79.1	68.6
Five or more	73.5	77.4
Six or more	68.3	82.8

MBS: Medicare Benefits Schedule; PPV: positive predictive value

Two or more records of either diabetes prescriptions where those with only insulin are excluded or MBS items each separated by 6 months or HbA1c tests each separated by 6 months had sensitivity of 78% and high PPV (80%) (Table S26).

A minimum of 2 records of either diabetes prescriptions or MBS items or HbA1c tests had a sensitivity of 63% and PPV of about 83% when a maximum gap of 6 months was applied for each marker (Table S27).

Potential algorithms for type 2 diabetes case definition

Table 4.10 shows the performance characteristics for potential algorithms for type 2 diabetes case definitions. The specificity (> 99%) and NPV (> 97%) estimates for all the suggested algorithms were very high.

The algorithm with the highest precision was a minimum of one diabetes prescription where people with only insulin prescriptions are excluded and at least one HbA1c test, each with a gap of 6 months to another HbA1c test (PPV 91% and sensitivity 61%). Using this algorithm, the probability that a person meeting this definition has type 2 diabetes is 91% and the probability that an individual not meeting this definition has type 2 diabetes is 3%.

A minimum of one diabetes prescription excluding people with only insulin prescriptions and at least one HbA1c test, each with a 12-month gap to another HbA1c test, had the second highest precision (PPV 89% and sensitivity 66%). The positive probability for this definition was 90% and negative probability was 2%.

The definition with the highest sensitivity (76%) while maintaining precision above 80% was at least one diabetes prescription (excluding people with only insulin prescriptions) and at least one HbA1c test (positive probability 84%, negative probability 1%).

Table 4.10: Validation of type 2 diabetes definition algorithms against people identified with type 2 diabetes using MedicineInsight standard definition; all study population (prevalence 6.1%), 2010 to 2020

Algorithm for type 2 diabetes case identification	Performance characteristics (%)					
	Se	Sp	PPV	NPV	Pr+	Pr-
One or more diabetes prescriptions excluding people with only insulin prescriptions AND at least one HbA1c test each with 6 months gap of another HbA1c test	60.5	99.6	90.5	97.5	91.3	2.7
One or more diabetes prescriptions excluding people with only insulin prescriptions AND at least one HbA1c test each with 12 months gap of another HbA1c test	66.3	99.5	88.9	97.9	90.2	2.3
Five or more records of either diabetes prescriptions excluding people with only insulin prescriptions OR HbA1c tests each with 12 months gap of another HbA1c test	61.5	99.4	87.1	97.6	87.7	2.6
Three or more records of either diabetes prescriptions excluding people with only insulin prescriptions OR MBS items each with a 6-month threshold OR HbA1c tests each with a 6-month threshold	66.7	99.3	86.2	97.9	86.9	2.3
Two or more records of either diabetes prescriptions excluding people with only insulin prescriptions OR diabetes related MBS items	66.2	99.3	86.6	97.9	86.8	2.3
One or more diabetes prescriptions excluding people with only insulin prescriptions AND at least one HbA1c test	75.5	99.0	82.3	98.4	84.0	1.4

MBS: Medicare Benefits Schedule; NPV: negative predictive value; PPV: positive predictive value; Pr+: positive probability, probability of having diabetes with the algorithm; Pr-: negative probability, probability of having diabetes without the algorithm; Se: sensitivity; Sp: specificity.

Findings from validation of the algorithms for type 1 diabetes case definition

As mentioned previously, this analysis was limited to people aged under 35 to optimise statistical power. However, it is important to note that type 1 diabetes affects people of all ages.

Single diabetes marker approach

Prescriptions for diabetes medicines

Diabetes medicines are one of the main discriminative features for type 1 and type 2 diabetes because the main treatment for type 1 diabetes is insulin. While some people with type 2 diabetes are treated with insulin, typically all people with type 1 diabetes are treated with insulin although this can be supplemented with other glucose lowering medicines in some cases when clinically relevant (Holt et al. 2021).

Table 4.11: Predictive power (%) of type 1 diabetes status based on number of prescriptions for individual diabetes medicines

Number of prescriptions	With type 1 diabetes		Without type 1 diabetes		Se	Sp	PPV	NPV	FIR
	With prescription	Without prescription	With prescription	Without prescription					
Insulin only (without other diabetes medicines)									
One or more	266	72	45	84,170	78.7	100.0	85.5	99.9	14.5
Two or more	225	113	22	84,193	66.6	100.0	91.1	99.9	8.9
Three or more	101	237	8	84,207	29.9	100.0	92.7	99.7	7.3
Four or more	59	279	6	84,209	17.5	100.0	90.8	99.7	9.2
Five or more	36	302	<5	n.p.	10.7	100.0	90.0	99.6	10.0
Six or more	24	314	<5	n.p.	7.1	100.0	96.0	99.6	4.0
Insulin (with/without other diabetes medicines)									
One or more	292	46	84	84,131	86.4	99.9	77.7	100.0	22.3
Two or more	249	89	46	84,169	73.7	100.0	84.4	99.9	15.6
Three or more	118	220	10	84,205	34.9	100.0	92.2	99.7	7.8
Four or more	68	270	8	84,207	20.1	100.0	89.5	99.7	10.5
Five or more	43	295	5	84,210	12.7	100.0	89.6	99.7	10.4
Six or more	30	308	<5	n.p.	8.9	100.0	96.8	99.6	3.2

FIR: false identification rate; NPV: negative predictive value; n.p.: not published; PPV: positive predictive value; Se: sensitivity: Sp: specificity

The analyses of diabetes prescriptions for people with type 1 diabetes focused on record of insulin prescriptions with or without other diabetes medicines and record of only insulin prescriptions (without other diabetes medicines). Consistent with treatment for type 1 diabetes, none of the people with type 1 diabetes aged under 35 had record of a prescription for thiazolidinediones, combinations of oral blood glucose lowering drugs, alpha glucosidase inhibitors, and dipeptidyl peptidase 4 inhibitors. Very few people (< 5) had sulfonylureas,

glucagon-like peptide-1 analogues and sodium-glucose co-transporter 2 inhibitors recorded, and 27 had at least one prescription for metformin.

The risk of misclassification of type 1 diabetes was lower for people with a minimum of one prescription for only insulin compared with those who had at least one prescription for insulin with/without other diabetes medicines (FIR about 15% and 22%, respectively) (Table 4.11). The risk of misclassification decreased with increase in the minimum number of insulin prescriptions.

Findings show that one or more prescriptions for only insulin (without other diabetes medicines) had 79% sensitivity and high PPV (86%). When the minimum number of prescriptions for only insulin is increased to 2, sensitivity decreases to 67%, but very high PPV (91%) is observed. Similar to previous studies, combining age under 35 and record of only insulin prescriptions appears to be a good indicator of type 1 diabetes.

While 2 or more prescriptions for insulin (with or without other diabetes medicines) was associated with high PPV (84%) and sensitivity of 74%, use of this algorithm to identify people with type 1 diabetes poses a risk of misclassification since some people with type 2 diabetes can be prescribed insulin in conjunction with other diabetes medicines.

HbA1c tests and diabetes-related MBS items

The HbA1c test is recommended for monitoring blood glucose in people with type 1 and type 2 diabetes. HbA1c tests alone are not good for differentiating between type 1 and type 2 diabetes even in people under 35, as demonstrated by low sensitivity and PPV estimates (Table S28).

Similarly, assessment of diabetes-related MBS items as an individual marker for identifying type 1 diabetes did not perform well in this data set (Table S29). Since only 119 people under 35 had at least one diabetes-related MBS item recorded (49 with type 1 diabetes and 70 without type 1 diabetes), the very few records may have reduced the predictive power for this diabetes marker.

Combined data approaches

Approach with 2 diabetes markers combined using ‘and’

For people aged under 35, the combination of at least one insulin prescription and at least one HbA1c test had 79% PPV and 66% sensitivity (Table 4.12). Increasing the minimum number of insulin prescriptions to 2 while keeping the minimum number of HbA1c test as one improved PPV by 5 percentage points, but sensitivity decreased. In the analysis with prescriptions for only insulin, combination of at least one prescription and at least one HbA1c test had a sensitivity of about 60% and a high PPV of 88% (Table S30).

Findings for the combination of insulin prescriptions and diabetes-related MBS items, as well as the combination of HbA1c tests and diabetes-related MBS items, are not presented. This is because of the small number of people with the 2 markers recorded (54 for insulin and diabetes-related MBS items; 91 for HbA1c tests and diabetes-related MBS items) among people under 35. The small numbers could affect the statistical robustness of the findings for these 2 analyses.

Table 4.12: Sensitivity and PPV (%) based on number of prescriptions for insulin (with/without other diabetes medicines) and HbA1c tests

	Number insulin prescriptions (with/without other diabetes medicines)	Number of HbA1c tests					
		One or more	Two or more	Three or more	Four or more	Five or more	Six or more
Sensitivity							
	One or more	66.0	51.5	38.8	31.1	25.7	21.9
	Two or more	57.4	45.9	34.9	27.8	22.8	20.4
	Three or more	30.5	25.4	20.1	16.9	14.8	13.0
	Four or more	18.1	15.7	12.4	10.7	9.2	8.3
	Five or more	12.1	11.2	8.6	7.4	6.8	6.5
	Six or more	8.6	8.0	6.2	5.3	5.3	5.0
PPV							
	One or more	78.8	83.3	85.6	87.5	85.3	85.1
	Two or more	84.4	87.1	90.1	91.3	89.5	89.6
	Three or more	92.0	91.5	91.9	90.5	89.3	89.8
	Four or more	88.4	88.3	87.5	85.7	83.8	84.9
	Five or more	89.1	90.5	90.6	89.3	88.5	91.7
	Six or more	96.7	96.4	100.0	100.0	100.0	100.0

PPV: positive predictive value

Approach with 2 diabetes markers combined using ‘or’

For people aged under 35, a minimum of 2 records of either insulin prescriptions or diabetes-related MBS items had high PPV (81%) and sensitivity of 76% (Table 4.13).

Table 4.13: Predictive power (%) of type 1 diabetes status based on number of insulin prescriptions or diabetes-related MBS items (excluding MBS items for HbA1c)

	Sensitivity	PPV
Number of insulin prescriptions (with/without other diabetes medicines) OR diabetes-related MBS items		
One or more	87.3	66.9
Two or more	75.7	81.3
Three or more	39.6	90.5
Four or more	25.2	91.4
Five or more	16.9	91.9
Six or more	12.1	93.2
Number of prescriptions for only insulin (without other diabetes medicines) OR diabetes-related MBS items		
One or more	80.8	70.5
Two or more	69.5	88.0
Three or more	34.9	92.2
Four or more	21.9	92.5
Five or more	14.5	92.5
Six or more	10.1	94.4

MBS: Medicare Benefits Schedule; PPV: positive predictive value

When prescriptions for only insulin or MBS items were assessed, PPV increased by 7 percentage points (88%), but there was a trade-off, with sensitivity decreasing (70%) for the same minimum number of records.

A minimum of 2 records of either prescriptions for only insulin or diabetes-related MBS items with a 6- or 12-month gap to another MBS item had very high PPV (91%) and sensitivity of 67% (Table 4.14).

Table 4.14: Predictive power (%) of type 1 diabetes status based on number of prescriptions for insulin only (without other diabetes medicines) or diabetes-related MBS items (excluding MBS items for HbA1c) with 6, 12, 18 and 24 months gap to another diabetes-related MBS item

Number of prescriptions for only insulin OR MBS items each with a gap of:	Sensitivity	PPV
6 months to another MBS item		
One or more	79.0	84.8
Two or more	66.6	90.7
Three or more	30.5	92.8
Four or more	17.8	90.9
Five or more	11.0	90.2
Six or more	7.4	92.6
12 months to another MBS item		
One or more	79.0	84.8
Two or more	66.6	90.7
Three or more	30.8	92.9
Four or more	18.6	91.3
Five or more	11.2	90.5
Six or more	7.4	92.6
18 months to another MBS item		
One or more	79.9	83.6
Two or more	67.8	90.5
Three or more	31.4	93.0
Four or more	19.2	91.6
Five or more	12.1	91.1
Six or more	8.9	93.8
24 months to another MBS item		
One or more	79.9	83.1
Two or more	67.8	90.5
Three or more	32.0	93.1
Four or more	19.2	91.6
Five or more	12.4	91.3
Six or more	8.9	93.8

MBS: Medicare Benefits Schedule; PPV: positive predictive value

Approach with 3 diabetes markers combined

The definition combining insulin prescriptions and HbA1c tests and diabetes-related MBS items did not perform well in identifying type 1 diabetes in people aged under 35. It is possible that the small number of people with all 3 markers recorded (46 for type 1 diabetes and 8 for those without type 1 diabetes) limited the predictive power of this analysis, and findings are not statistically robust.

Similarly, the combination of insulin prescriptions or MBS items or HbA1c tests did not result in suitable estimates for PPV and sensitivity.

Potential algorithms for type 1 diabetes case definition

Potential algorithms for type 1 diabetes case definition for people aged under 35 in MedicineInsight are shown in Table 4.15. Both specificity and NPV estimates for the presented algorithms were nearly perfect (100%).

A minimum of 2 prescriptions for insulin only (without other diabetes medicines) during the study period had the highest precision (91%) with sensitivity of 67%. Using this algorithm, the

probability that a person not meeting this definition has type 1 diabetes is very minimal (0.1%). The positive probability is not reported because the positive likelihood ratio (used to calculate positive probability) cannot be calculated when specificity is 100% (see Box 2.1 for positive likelihood ratio formula).

The definition combining prescriptions for insulin only (without other diabetes medicines) with variation of the maximum gap (6, 12, 18 or 24 months) between MBS items performed well. A minimum of 2 records of either prescriptions for insulin only or MBS items each with a 6, 12,18 or 24-month gap to another MBS item had a PPV of 91%. These algorithms had sensitivity of 67% for definitions with maximum gaps of 6 and 12 months between MBS items, and 68% for definitions with maximum gaps of 18 and 24 months.

The definition with the highest sensitivity (79%) while maintaining precision above 85% was having at least one prescription for insulin only (without other diabetes medicines) at any time in the study period (negative probability 0.1%).

Table 4.15: Validation of type 1 diabetes definition algorithms against people identified with type 1 diabetes using MedicineInsight standard definition; study population aged under 35 (prevalence 0.4%), 2010 to 2020

Algorithm for type 1 diabetes case identification	Performance characteristics (%)					
	Se	Sp	PPV	NPV	Pr+	Pr-
Two or more prescriptions for insulin only (without other diabetes medicines)	66.6	100.0	91.1	99.9	n.a.	0.1
Two or more records of either prescriptions for insulin only (without other diabetes medicines) OR MBS items each with a 6-month or 12-month gap of another MBS item	66.6	100.0	90.7	99.9	n.a.	0.1
Two or more records of either prescriptions for insulin only (without other diabetes medicines) OR MBS items each with an 18-month or 24-month gap of another MBS item	67.8	100.0	90.5	99.9	n.a.	0.1
Two or more records of either prescriptions for insulin only (without other diabetes medicines) OR diabetes-related MBS items	69.5	100.0	88.0	99.9	n.a.	0.1
One or more prescriptions for insulin only (without other diabetes medicines)	78.7	100.0	85.5	99.9	n.a.	0.1

MBS: Medicare Benefits Schedule; n.a.: not applicable; NPV: negative predictive value; PPV: positive predictive value; Pr+: positive probability, probability of having diabetes with the algorithm; Pr-: negative probability, probability of having diabetes without the algorithm; Se: sensitivity; Sp: specificity. Positive probability is missing because it cannot be calculated when specificity is 100%.

Potential algorithms for type 1 diabetes case definitions for people of all ages are shown in Table S31. While most of the algorithms are similar to those for people aged under 35, the sensitivity and PPV estimates are lower. For example, the algorithm with at least 2 prescriptions for insulin only (without other diabetes medicines) in the study period had PPV of 72% and sensitivity of 58% for people of all ages, while the same algorithm had estimates of 91% and 67% for PPV and sensitivity, respectively, for people under 35.

5 Recording of diabetes markers before and after diagnosis of diabetes

This chapter looks at recording of the markers of diabetes status, prescriptions, HbA1c tests and MBS items before (pre-diagnosis) and on or after the date of the diagnosis of diabetes (post-diagnosis) using MedicineInsight data. Understanding pre- and post-diagnosis recording of these diabetes markers will help inform whether these markers can be used to improve the accuracy of identifying people with diabetes in Australia.

Methods

A cohort of people newly diagnosed with diabetes in the 2-year period from 1 September 2017 to 31 August 2019 was identified based on the MedicineInsight standard definition (see Figure 2.1 for further details on cohort selection). At least one year look forward period (September 2017 to August 2020) from the date of diagnosis was considered reasonable to allow sufficient time for diabetes prescriptions and health service use (HbA1c tests and diabetes-related MBS items) to be recorded for post-diagnosis assessment. Pre-diagnosis assessment included records before the diagnosis date.

The earliest date on which diabetes was recorded in one of the 3 diagnosis fields – diagnosis, reason for visit or reason for prescriptions – was defined as the first diagnosis date.

Diabetes pre-diagnosis period was defined as the time prior to each individual's diabetes diagnosis date, with the earliest date of records set at 1 January 2010 up to the day before the diagnosis date.

The diabetes post-diagnosis period was defined as the period from the date of diagnosis up to the end of the study, 31 August 2020.

To minimise misclassification of people with pre-existing diabetes as incident diabetes, people who had at least one visit at the practice in the 12 months before the diagnosis date were included to allow sufficient opportunity for recording of diabetes diagnosis, prescriptions or health service use.

The proportion of people with at least one record for each diabetes marker recorded pre-diagnosis and post-diagnosis period was assessed. As some people might have markers recorded both in the pre- and post-diagnosis period, we also assessed the proportion of records of each diabetes marker recorded in the 2 periods.

Findings

Approximately 1,900 people were identified as being newly diagnosed with diabetes (incident population) in the 2-year study period, 910 during 2017–18 and 970 in 2018–19. Appendix B shows the characteristics of people newly diagnosed with diabetes. Of the incident population, about 91% had at least one HbA1c test recorded, 75% had at least one prescription for any diabetes medicine and only 21% had a diabetes-related MBS item recorded (Table 5.1).

People with diabetes markers recorded pre- and post-diagnosis

HbA1c tests

Among the diabetes incident population who had at least one HbA1c test recorded, about 69% had at least one test recorded before the diagnosis date (pre-diagnosis period) and about 91% had a test recorded on or after the diagnosis date (post-diagnosis period) (Table 5.1).

To account for the HbA1c tests done in the lead-up to confirming a diabetes diagnosis, tests recorded 21 days prior to the date of diagnosis were considered as post-diagnosis. For this sensitivity analysis, just over 45% of those who had at least one HbA1c test recorded had an HbA1c test recorded pre-diagnosis (24 percentage points decline) and a test was recorded post-diagnosis for 96% (5 percentage points increase). It is important to note that HbA1c tests can be performed for monitoring blood glucose in the management of other conditions, which may reflect the pre-diagnosis recording in the diabetes incident population.

Diabetes-related MBS items

MBS billing data are incomplete in MedicineInsight because for one-third of the MedicineInsight practices, the practice billing and clinical information systems are not compatible. This is reflected in the results, with only 1 in 5 people in the incident cohort having at least one diabetes-related MBS item recorded.

Findings indicate that among the diabetes incident population who had at least one diabetes-related MBS item recorded, 91% and 14% had diabetes-related MBS items in the post-diagnosis and pre-diagnosis period, respectively (Table 5.1). As most of the people with a diabetes-related MBS item during the pre-diagnosis period had MBS items for diabetes cycle of care (n=43) and/or allied health care services (n=19), this suggests that these few individuals are likely not to be newly diagnosed but prevalent cases, since these MBS items are specific for people with existing diabetes.

Prescriptions for diabetes medicines

Of the over 1,400 diabetes incident population who had at least one diabetes prescription recorded, 24% had at least one diabetes prescription recorded pre-diagnosis and about 92% had a diabetes prescription recorded during the post-diagnosis period (Table 5.1). Although the majority of people with diabetes prescriptions in the pre-diagnosis period had at least one prescription for metformin (with/without other diabetes medicines) which is used in the treatment of other conditions, a number of individuals had prescriptions for medicines specific for diabetes including combinations of oral blood glucose lowering drugs, insulin, sulfonylureas, DPP4 inhibitors and SGLT2 inhibitors. This also indicates possible misclassification of people with prevalent diabetes as incident diabetes.

Table 5.1: Proportion of the diabetes incident cohort who had a minimum of one diabetes marker and proportion of those with at least one HbA1c test, diabetes-related MBS item or prescriptions who had markers recorded pre-diagnosis or post-diagnosis period

	Incident cohort (N=1,879)		Number of people [a]		Per cent of incident cohort with at least one record of the marker [a]	
	Number with at least one record of the marker	% of incident population	Pre-diagnosis	Post-diagnosis	Pre-diagnosis	Post-diagnosis
HbA1c tests	**1,718**	**91.4**	**1,177**	**1,556**	**68.5**	**90.6**
HbA1c test (adjusted) [b]	1,718	91.4	778	1,654	45.3	96.3
All diabetes-related MBS items [c]	**401**	**21.3**	**56**	**365**	**14.0**	**91.0**
Diabetes cycle of care	330	17.6	43	305	13.0	92.4
Allied Health Services	90	4.8	19	77	21.1	85.6
MBS HbA1c test	14	0.7	<5	14	n.p.	100.0
Eye examination	<5	n.p.	0	<5	0.0	n.p.
All diabetes medicine prescriptions [c]	**1,416**	**75.4**	**340**	**1,300**	**24.0**	**91.8**
Metformin	1,175	62.5	261	1,091	22.2	92.9
Combinations of oral blood glucose lowering drugs	317	16.9	111	301	35.0	95.0
Sulfonylureas	246	13.1	91	231	37.0	93.9
Insulin	235	12.5	96	215	40.9	91.5
Sodium-glucose co-transporter 2 inhibitors	136	7.2	47	135	34.6	99.3
Dipeptidyl peptidase 4 inhibitors	133	7.1	54	130	40.6	97.7
Glucagon-like peptide-1 analogues	87	4.6	36	86	41.4	98.9
Thiazolidinediones	<5	n.p.	<5	<5	n.p.	n.p.
Alpha glucosidase inhibitors	<5	n.p.	<5	<5	n.p.	n.p.

MBS: Medicare Benefits Schedule; n.p.: not published

[a] Some people have markers recorded both pre-and post-diagnosis, thus the total per cent might exceed 100%.

[b] Post diagnosis includes HbA1c tests recorded 21 days prior to date of diagnosis to account for HbA1c tests that may have been performed in the lead-up to diagnosis confirmation.

[c] An individual may have more than one of the MBS items or diabetes medicine classes recorded.

About 1 in 4 people (25%) with incident diabetes did not have a prescription for any diabetes medicine recorded. This suggests that some people may have obtained their medicines from another general practice or from other health care settings such as specialists or hospitals. However, this could also mean that some people who are newly diagnosed with type 2 diabetes use non-pharmacological interventions to manage their diabetes. Table S32 shows evidence to support this theory as nearly 9 in 10 people (88%) with incident diabetes who did not have a prescription for any diabetes medicines had type 2 diabetes.

Pre- and post-diagnosis records for diabetes markers

Table 5.2 shows the number of records for each diabetes marker recorded in the pre- and post-diagnosis period. Most records for diabetes-related MBS items (86%), diabetes prescriptions (83%) and HbA1c tests (66%) were recorded in the post-diagnosis period.

Table 5.2: Proportion of all records for each diabetes marker for the diabetes incident cohort recorded pre- or post-diagnosis period

	Number of records			Per cent of total records		
	Pre-diagnosis	Post-diagnosis	Total	Pre-diagnosis	Post-diagnosis	Total
HbA1c tests	**2644**	**5018**	**7662**	**34.5**	**65.5**	**100.0**
All diabetes-related MBS items	**108**	**677**	**785**	**13.8**	**86.2**	**100.0**
Diabetes cycle of care	66	413	479	13.8	86.2	100.0
Allied Health Services	41	215	256	16.0	84.0	100.0
MBS HbA1c test	< 5	> 45	< 50	< 3	> 95	100.0
Eye examination	0	< 5	< 5	0.0	100.0	100.0
All diabetes prescriptions	**590**	**2772**	**3362**	**17.5**	**82.5**	**100.0**
Metformin	280	1367	1647	17.0	83.0	100.0
Combinations of oral blood glucose lowering drugs	64	380	444	14.4	85.6	100.0
Sulfonylureas	71	243	314	22.6	77.4	100.0
Insulin	108	399	507	21.3	78.7	100.0
Sodium-glucose co-transporter 2 inhibitors	25	146	171	14.6	85.4	100.0
Dipeptidyl peptidase 4 inhibitors	23	137	160	14.4	85.6	100.0
Glucagon-like peptide-1 analogues	14	96	110	12.7	87.3	100.0
Thiazolidinediones	< 5	< 5	< 5	n.p.	n.p.	n.p.
Alpha glucosidase inhibitors	< 5	< 5	5	n.p.	n.p.	n.p.

n.p.: not published

Findings indicate that the majority of the diabetes incident population had the diabetes markers recorded in the post-diagnosis period. However, some people identified as having incident diabetes also had HbA1c tests, diabetes-specific MBS items and prescriptions recorded before the date of diagnosis. A number of reasons could explain the observed pre-diagnosis records for these diabetes markers, including:

- The diabetes diagnosis date used may not be accurate, particularly for people who joined the MedicineInsight practice after their original date of diagnosis whose first record of diabetes is regarded as the diagnosis date. This could lead to potential misclassification of prevalent diabetes as incident diabetes.

- Some diabetes medicines like metformin are prescribed for pre-diabetes or other conditions and these would appear as pre-diagnosis records.
- Pre-diagnosis HbA1c tests could have been performed for blood glucose monitoring in the management of other conditions.
- Pre-diagnosis HbA1c tests could have been performed prior to the date of diagnosis as a lead-up to making a diagnosis. When this was accounted for, there was a reduction in the proportion of people with HbA1c tests recorded in the pre-diagnosis period and an increase in the proportion of those with HbA1c tests recorded post-diagnosis.

Nevertheless, the majority of records for the diabetes markers were recorded during the post-diagnosis period, indicating that these markers can be used to improve the accuracy of identifying people with established diabetes, although potential misclassification cannot be ruled out. This highlights that case definitions that combine diabetes markers may provide better capture of the diabetes population while minimising misclassification.

Conclusion

Findings from this analysis show that approaches using a combination of diabetes markers provide better case definitions for identifying people with diagnosed diabetes with very high precision (≥ 90%) and acceptable sensitivity (> 60%). The specificity and NPV estimates for all the suggested algorithms are very high (> 97%). High specificity indicates that a greater proportion of people identified as not having diabetes according to the MedicineInsight standard definition were identified as not having diabetes by the case definition algorithms using diabetes markers. High NPV implies that most people identified as not having diabetes by the case definition algorithms did not have diabetes according to the standard definition.

The diabetes case definition with a minimum of one diabetes prescription and at least one HbA1c test each with a 6-month gap to another HbA1c test during the study period had the highest precision (96%) and sensitivity of 61%. The probability that a person meeting this definition has diabetes is 96% and the probability that a person not meeting this definition has diabetes is 3%.

The findings show that there are no definitive case definition algorithms, based on diabetes markers, for differentiating between type 1 and type 2 diabetes when all age groups are considered. However, when the analysis was limited to people aged under 35, a minimum of 2 prescriptions for insulin only (without other diabetes medicines) had the highest precision for identifying type 1 diabetes (91%) with sensitivity of 67%. The probability that a person not meeting this definition has type 1 diabetes is very low at 0.1%. This case definition that uses age and prescription of insulin only to identify people with type 1 has been consistently used by other researchers and could be utilised in data sets where prescriptions and demographic data are available. As type 1 diabetes affects people of all ages, this definition would not pick up people with type 1 diabetes aged 35 and over. However, these definitions are intended to supplement existing data sources in identifying people with diagnosed diabetes and type 1 diabetes is well captured in the National Diabetes Services Scheme (NDSS) registrant data.

Using the study cohort with all age groups, a minimum of one diabetes prescription with exclusion of people prescribed insulin only and at least one HbA1c test each with a 6-month gap to another HbA1c test had the highest precision (91%) of identifying type 2 diabetes and 61% sensitivity. The probability that a person meeting this definition has type 2 diabetes is 91% and the probability that a person not meeting this definition has type 2 diabetes is 3%.

We explored whether markers of diabetes status can reliably be used to identify established diabetes by assessing pre-diagnosis and post-diagnosis recording of these markers. Our findings show that although pre-diagnosis records for each marker were found, most records were recorded in the post-diagnosis period, indicating that these markers can be used to identify people with established diabetes, although potential misclassification cannot be ruled out. This underscores the importance of using definition algorithms that combine diabetes markers for better identification of diabetes.

A number of factors might confound generalisability of the findings of this study including:

- differences in the diabetes markers in the data used and administrative data sets such as PBS and MBS

- incompleteness of the MBS billing data in MedicineInsight which could have affected the predictive power of this marker as well as affecting the performance of algorithms using this diabetes marker
- the MedicineInsight standard definition that was used as the reference standard may have limitations and this could have affected the validity estimates
- performance characteristics like PPV and NPV depend on the prevalence of the condition in the study population, and these would vary in a setting where the prevalence of diabetes is different from this study population.

Despite the above limitations, the case definition algorithms in this report add to a body of evidence using similar markers and can help understand criteria for identifying diabetes in administrative data sets with similar diabetes markers. The findings can also be tailored for use in linked data collections like the National Health Data Hub (NHDH), or Kidney and Diabetes Data Integration (KADDI), analysis assets which consist of PBS and MBS administrative data in addition to other health data. These findings can help supplement existing data sources, such as the NDSS, in identifying people with diagnosed diabetes to enable better estimation of its prevalence and further monitoring of the diabetes population. This is vital for implementing policies for prevention and control as well as proper resource allocation.

Recommendations

- The suggested algorithms are those where high precision is suitable as opposed to high sensitivity. Algorithms that maximise precision but with acceptable sensitivity were preferred. The suggested algorithms for identifying diagnosed diabetes and its types have good agreement with the MedicineInsight standard definitions. However, these algorithms might not be suitable for studies where high sensitivity is required.
- The suggested case definitions for type 1 diabetes are limited to people aged under 35. Further studies are required to determine case definitions for type 1 diabetes among people of all age groups.
- These definitions are intended to supplement the existing data sources, such as the NDSS, in identifying people with diagnosed diabetes, particularly in linked data collections.

Appendix A: Definitions

Table A1: Diabetes definition

Condition	Terms included in definition
Type 1 diabetes	Includes: diabetes mellitus (iddm or type I or type 1), iddm, insulin dependent diabetes mellitus, juvenile onset diabetes
Type 2 or non-specified diabetes	Includes: diabetes, diabetes (controlled or cortisone induced or unstable), diabetes mellitus, diabetes mellitus (niddm, or type ii or type 2 or type 3c), latent autoimmune diabetes of adults, niddm, non insulin dependent diabetes mellitus, pancreatogenic diabetes, t2dm, t11, tii
Diabetes mellitus	Includes: type 1, type 2 or non-specified diabetes terms

Table A2: ATC codes used to identify diabetes medicines

Diabetes medicine class	ATC code
Insulin and analogues	A10A
Biguanides - metformin	A10BA
Sulfonylureas	A10BB
Combinations of oral blood glucose lowering drugs	A10BD
Alpha glucosidase inhibitors	A10BF
Thiazolidinediones	A10BG
Dipeptidyl peptidase 4 inhibitors	A10BH
Glucagon-like peptide-1 analogues	A10BJ
Sodium-glucose co-transporter 2 inhibitors	A10BK

ATC: Anatomical Therapeutic Chemical Classification

Table A3: Diabetes-related Medicare Benefits Schedule item numbers included in the study

Item	Item number
Category 1 - Professional attendances	
General practitioner (GP) completion of annual diabetes cycle of care for patients with established diabetes mellitus.	2517, 2518, 2521, 2522, 2525, 2620, 2622, 2624, 2631,2633, 2635
Medical practitioner (non-GP) completion of the annual diabetes cycle of care for patients with established diabetes mellitus.	259, 260, 261, 262, 263, 264
Professional attendance of more than 15 minutes duration, being the first in a course of attention involving the examination of the eyes, with the instillation of a mydriatic, of a patient with diabetes mellitus requiring comprehensive reassessment.	10915
Category 2 - Diagnostic procedures and investigations	
Assessment of visual acuity and bilateral retinal photography with a non-mydriatic retinal camera, including analysis and reporting of the images for initial or repeat assessment for presence or absence of diabetic retinopathy, in a patient with medically diagnosed diabetes. The patient is of Aboriginal and Torres Strait Islander descent in addition to other criteria.	12325
Assessment of visual acuity and bilateral retinal photography with a non-mydriatic retinal camera, including analysis and reporting of the images for initial or repeat assessment for presence or absence of diabetic retinopathy, in a patient with medically diagnosed diabetes.	12326
Category 6 - Pathology services[a]	
Quantitation of fructosamine performed in the management of established diabetes.	66557
Category 8 - Miscellaneous services	
Diabetes education service - assessment for group services.	81100
Diabetes education service - group service.	81105
Exercise physiology service - assessment for group services.	81110
Exercise physiology service - group service.	81115
Dietetics service - assessment for group services.	81120
Dietetics service - group service.	81125
Diabetes education service.	10951
Diabetes education health service provided to a person who is of Aboriginal or Torres Strait Islander descent by an eligible diabetes educator.	81305

[a] MBS items for HbA1c tests were not included among diabetes-related MBS items to minimise duplication as HbA1c tests from the pathology data were assessed.

Appendix B: Sociodemographic characteristics of study populations

Table B1: Sociodemographic characteristics of study populations

Sociodemographic characteristic	Baseline population		Diabetes cohort		Type 2 diabetes cohort		Population aged < 35		Type 1 diabetes and aged < 35 years cohort		Diabetes incident cohort	
	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
Total number	206,163		13,939		12,497		84,553		338		1,879	
Mean age (SD, years)	42 (24)		65 (16)		66 (14)		18 (10)		23 (8)		61 (16)	
Age group (years)												
0–4	12,919	6.3	n.a.	n.a.	n.a.	n.a.	12,919	15.3	<5	n.p.	n.a.	n.a.
5–9	11,918	5.8	n.a.	n.a.	n.a.	n.a.	11,918	14.1	<15	n.p	n.a.	n.a.
10–14	9,659	4.7	n.a.	n.a.	n.a.	n.a.	9,659	11.4	37	10.9	n.a.	n.a.
15–19 (or < 20 for diabetes cohorts)	10,006	4.9	155	1.1	34	0.3	10,006	11.8	63	18.6	30	1.6
20–24	12,046	5.8	101	0.7	27	0.2	12,046	14.2	72	21.3	17	0.9
25–29	13,557	6.6	127	0.9	55	0.4	13,557	16.0	70	20.7	21	1.1
30–34	14,428	7.0	216	1.5	131	1.0	14,428	17.1	78	23.1	40	2.1
35–39	14,691	7.1	356	2.6	270	2.2	n.a.	n.a.	n.a.	n.a.	69	3.7
40–44	13,112	6.4	447	3.2	380	3.0	n.a.	n.a.	n.a.	n.a.	99	5.3
45–49	13,653	6.6	768	5.5	646	5.2	n.a.	n.a.	n.a.	n.a.	141	7.5
50–54	13,055	6.3	1,002	7.2	909	7.3	n.a.	n.a.	n.a.	n.a.	193	10.3
55–59	12,975	6.3	1,301	9.3	1,203	9.6	n.a.	n.a.	n.a.	n.a.	190	10.1
60–64	12,711	6.2	1,608	11.5	1,474	11.8	n.a.	n.a.	n.a.	n.a.	246	13.1
65–69	11,489	5.6	1,822	13.1	1,714	13.7	n.a.	n.a.	n.a.	n.a.	248	13.2
70–74	10,807	5.2	1,991	14.3	1,873	15.0	n.a.	n.a.	n.a.	n.a.	243	12.9
75–79	7,620	3.7	1,586	11.4	1,479	11.8	n.a.	n.a.	n.a.	n.a.	160	8.5
80–84	5,378	2.6	1,260	9.0	1,180	9.4	n.a.	n.a.	n.a.	n.a.	88	4.7
85+	6,119	3.0	1,198	8.6	1,122	9.0	n.a.	n.a.	n.a.	n.a.	94	5.0
Not recorded	20	0.0	<5	n.p.	0	n.a.	20	0.0	0	n.a.		
Sex												
Female	114,491	55.5	6,310	45.3	5,658	45.3	46,735	55.3	169	50.0	833	44.3
Male	91,184	44.2	7,612	54.6	6,826	54.6	37,547	44.4	167	49.4	1,043	55.5
Not stated/inadequately described	488	0.2	17	0.1	13	0.1	271	0.3	<5	n.p.	<5	n.p.
Indigenous status												

Sociodemographic characteristic	Baseline population		Diabetes cohort		Type 2 diabetes cohort		Population aged < 35		Type 1 diabetes and aged < 35 years cohort		Diabetes incident cohort	
	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
First Nations people	6,840	3.3	584	4.2	539	4.3	4,311	5.1	13	3.8	88	4.7
Non-Indigenous people	163,534	79.3	11,503	82.5	10,307	82.5	64,746	76.6	272	80.5	1,550	82.5
Not stated/inadequately described	35,789	17.4	1,852	13.3	1,651	13.2	15,496	18.3	53	15.7	241	12.8
SEIFA IRSD quintile												
Group 1 (most disadvantaged)	37,780	18.3	3,453	24.8	3,137	25.1	14,494	17.1	60	17.8	437	23.3
Group 2	38,661	18.8	2,892	20.7	2,587	20.7	15,483	18.3	73	21.6	394	21.0
Group 3	47,004	22.8	3,307	23.7	3,004	24.0	19,797	23.4	72	21.3	433	23.0
Group 4	38,917	18.9	2,159	15.5	1,910	15.3	16,805	19.9	71	21.0	327	17.4
Group 5 (least disadvantaged)	42,564	20.6	2,050	14.7	1,791	14.3	17,482	20.7	61	18.0	283	15.1
Not recorded	1,237	0.6	78	0.6	68	0.5	492	0.6	<5	n.p.	5	0.3
Remoteness areas												
Major cities	124,495	60.4	7,536	54.1	6,722	53.8	53,625	63.4	188	55.6	1,041	55.4
Inner regional	52,213	25.3	4,046	29.0	3,640	29.1	19,943	23.6	106	31.4	503	26.8
Outer regional	25,610	12.4	2,062	14.8	1,866	14.9	9,594	11.3	38	11.2	298	15.9
Remote and very remote	2,611	1.3	217	1.6	201	1.6	899	1.1	5	1.5	32	1.7
Not recorded	1,234	0.6	78	0.6	68	0.5	492	0.6	<5	n.p.	5	0.3

IRSD: Index of Relative Socio-Economic Advantage and Disadvantage; n.a: not applicable; n.p: not published; SD: Standard deviation

Acknowledgements

This report was prepared by Doreen Busingye and James Pearce of the Australian Institute of Health and Welfare (AIHW). Valuable input and guidance were provided by Louise Gates, Heidi Dietz, Miriam Lum On, Justin Harvey, Jenna Haddin, Allan Pollock and Ashleigh Bennett.

The authors gratefully acknowledge the contributions of the AIHW's Diabetes Expert Advisory Group, chaired by Jonathan Shaw. Members at the time of preparation of the report were Wendy Davis, Mark Harris, Renza Scibilia, Glynis Ross, Maria Craig, Sophia Zoungas and Justine Cain. Contributions from the AIHW's Primary Healthcare Advisory Committee members are also gratefully acknowledged.

We are grateful to the general practices and general practitioners who participate in the MedicineInsight Program and the patients whose de-identified data make this work possible.

Abbreviations

ABS	Australian Bureau of Statistics
ACSQHC	Australian Commission on Safety and Quality in Health Care
AIHW	Australian Institute of Health and Welfare
APEG	Australasian Paediatric Endocrine Group
ATC	Anatomical Therapeutic Chemical Classification
BP	Best Practice
CIS	clinical information system
EHR	electronic health record
FIR	false identification rate
GP	general practitioner
GRHANITE	GeneRic Health Network Information Technology for the Enterprise
HbA1c	glycated haemoglobin
ICD	International Classification of Diseases
IDDM	insulin dependent diabetes mellitus
IRSD	Index of Relative Socio-Economic Advantage and Disadvantage
KADDI	Kidney and Diabetes Data Integration
LOINC	Logical Observation Identifiers Names and Codes
MBS	Medicare Benefits Schedule
MD	MedicalDirector 3
NDSS	National Diabetes Services Scheme
NHDH	National Health Data Hub
NHS	National Health Survey
NIDDM	non-insulin dependent diabetes mellitus
NIHSI	National Integrated Health Services Information
NPS	National Prescribing Service
NPV	negative predictive value
PBS	Pharmaceutical Benefits Scheme
PLIDA	Person Level Integrated Data Asset
PPV	positive predictive value
RACGP	Royal Australian College of General Practitioners

RPBS	Repatriation Pharmaceutical Benefits Scheme
SD	standard deviation
SEIFA	Socio-Economic Indexes for Areas

Symbols

Symbol	Definition
n.a.	not available, not applicable
n.p.	not published
− (minus)	negative
+ (plus)	positive
<	less than
>	greater than
≥	greater than or equal to

Glossary

Aboriginal or Torres Strait Islander (First Nations people): A person of Aboriginal and/or Torres Strait Islander descent who identifies as an Aboriginal and/or Torres Strait Islander.

Clinical information system (CIS): A generic term to describe one of several Australian national general practice software programs used by GPs to store patient/consultation/ prescription data of which Best Practice and Medical Director are two examples).

Diabetes (diabetes mellitus): A chronic condition in which the body cannot properly use its main energy source, the sugar glucose. This is due to a relative or absolute deficiency in insulin, a hormone that is produced by the pancreas and helps glucose enter the body's cells from the bloodstream and then be processed by them. Diabetes is marked by an abnormal build-up of glucose in the blood, and it can have serious short- and long-term effects. For the 3 main types of diabetes see **type 1 diabetes, type 2 diabetes** and **gestational diabetes.**

Diagnosed diabetes: Includes cases where a diagnosis is certified by a doctor, nurse or credentialled diabetes educator or an individual self-reports having been told they have diabetes by a doctor or nurse.

False identification rate (FIR, or false discovery rate): This indicates the proportion of false positives among those who test positive (that is, the proportion of people misclassified as having the disease).

General practice: Includes fully-qualified general practitioners (GPs). Physicians in training are normally excluded.

General practitioner (GP): A medical practitioner who provides primary comprehensive and continuing care to patients and their families in the community.

Gestational diabetes: A form of diabetes when higher than optimal blood glucose is first diagnosed during pregnancy (gestation). It may disappear after pregnancy but signals a high risk of diabetes occurring later on.

Glycated haemoglobin (HbA1c): The main biomarker used to assess long-term glucose control in people living with diabetes. Haemoglobin is a protein in red blood cells which can bind with sugar to form HbA1c. It is directly related to blood glucose levels and strongly related with the development of long-term diabetes complications.

Negative likelihood ratio (LR-): Shows how many times a test result is less likely to be found in diseased, compared with non-diseased, people.

Negative predictive value: The probability of not having the disease when the test result is negative or the proportion of people with a negative test result who truly do not have the disease.

Negative probability (Pr-): The probability of having a disease given a negative test for that disease.

Other diabetes: A name for less common diabetes resulting from a range of different health conditions or circumstances.

Prevalence: The number or proportion (of cases, instances, and so forth) in a population at a given time. For example, in relation to cancer, refers to the number of people alive who had been diagnosed with cancer in a prescribed period (usually 1, 5, 10 or 26 years).

Positive likelihood ratio (LR+): Shows how many times a test result is more likely to be found in diseased, compared with non-diseased, people.

Positive predictive value (precision): The probability of disease in a person with a positive test result or the proportion of people with a positive test for the disease who truly have the disease.

Positive probability (Pr+): The probability of having a disease given a positive test for that disease.

Remoteness: A system which classifies geographical locations into groups (Major cities, Inner regional, Outer regional, Remote, Very remote) according to distance from major population centres and services. In this analysis, remoteness is based on Accessibility/Remoteness Index of Australia (ARIA) and defined as Remoteness Areas by the Australian Statistical Geographical Standard (ASGS) (in each Census year). Remoteness is a geographical concept and does not take account of accessibility which is influenced by factors such as the socioeconomic status or mobility of a population.

Sensitivity: The proportion of people with the disease who have a positive test for the disease.

Socioeconomic areas: An indication of how 'well off' a person or group is. Socioeconomic areas are reported using the Australian Bureau of Statistics' Socio-Economic Indexes for Areas (SEIFA), whereby areas are classified on the basis of social and economic information (such as low income, low educational attainment, high levels of public sector housing, high unemployment and jobs in relatively unskilled occupations) collected in the Census of Population and Housing. Socio-Economic Indexes for Areas are divided into 5 groups, from the most disadvantaged (worst off) to the least disadvantaged (best off). Note that this index refers to the average disadvantage of all people living in an area, not to the level of disadvantage of a specific individual.

Specificity: The proportion of people without the disease who have a negative test.

Type 1 diabetes: A lifelong autoimmune disease that can be diagnosed at any age. The exact cause is unknown, but it is believed to be the result of an interaction of genetic and environmental factors.

Type 2 diabetes: The most common form of diabetes, it is a condition in which the body becomes resistant to the normal effects of insulin and gradually loses the capacity to produce enough insulin in the pancreas. The condition has strong genetic and family-related (non-modifiable) risk factors and is also often associated with modifiable risk factors.

References

AIHW (Australian Institute of Health and Welfare) (2009) *Diabetes prevalence in Australia: an assessment of national data sources*, AIHW, Australian Government, accessed 17 October 2023.

AIHW (2023) *Diabetes: Australian facts*, AIHW website, accessed 17 October 2023.

Busingye D, Gianacas C, Pollack A, Chidwick C, Merrifield A, Norman S, Mullin B, Hayhurst R, Blogg S, Havard A and Stocks N (2019) 'Data Resource Profile: MedicineInsight, an Australian national primary health care database', *International Journal of Epidemiology*, 48(6):1741–1741h, doi:10.1093/ije/dyz147.

Chen G, Khan N, Walker R and Quan H (2010) 'Validating ICD coding algorithms for diabetes mellitus from administrative data', *Diabetes Research and Clinical Practice*, 89(2):189–195, doi:10.1016/j.diabres.2010.03.007.

Havard A, Manski-Nankervis JA, Thistlethwaite J, Daniels BD, Myton R, Tu K and Chidwick K (2021) 'Validity of algorithms for identifying five chronic conditions in MedicineInsight, an Australian national general practice database', *BMC Health Services Research*, 21(1):551, doi:10.1186/s12913-021-06593-z.

Hirsch AG and Scheck McAlearney A (2014) 'Measuring diabetes care performance using electronic health record data: the impact of diabetes definitions on performance measure outcomes', *American Journal of Medical Quality*, 29(4):292–299, doi:10.1177/1062860613500808.

Holt RIG, DeVries JH, Hess-Fischl A, Hirsch IB, Kirkman MS, Klupa T, Ludwig B, Nørgaard K, Pettus J, Renard E, Skyler J.S, Snoek FJ, Weinstock RS and Peters AL (2021) 'The management of type 1 diabetes in adults: a consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD)', *Diabetes Care*, 44(11):2589–2625, doi:10.2337/dci21-0043.

Ke C, Stukel TA, Luk A, Shah BR, Jha P, Lau E, Ma RCW, So WY, Kong AP, Chow E and Chan JCN (2020) 'Development and validation of algorithms to classify type 1 and 2 diabetes according to age at diagnosis using electronic health records', *BMC Medical Research Methodology*, 20(1):35, doi:10.1186/s12874-020-00921-3.

Klompas M, Eggleston E, McVetta J, Lazarus R, Li L and Platt R (2013) 'Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data', *Diabetes Care*, 36(4):914–921, doi:10.2337/dc12-0964.

Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A and Rahme E (2013) 'Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records', *PLoS ONE*, 8(10):e75256, doi:10.1371/journal.pone.0075256.

Lethebe BC, Williamson T, Garies S, McBrien K, Leduc C, Butalia S, Soos B, Shaw M and Drummond N (2019) 'Developing a case definition for type 1 diabetes mellitus in a primary care electronic medical record database: an exploratory study', *Canadian Medical Association Journal Open*, 7(2):E246–E251, doi:10.9778/cmajo.20180142.

Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL and Tu K (2018) 'Identifying diabetes cases from administrative data: a population-based validation study', *BMC Health Services Research*,18(1):316, doi:10.1186/s12913-018-3148-0.

Nakhla M, Simard M, Dube M, Larocque I, Plante C, Legault L, Huot C, Gagne N, Gagne J, Wafa S, Benchimol EI and Rahme E (2019) 'Identifying pediatric diabetes cases from health administrative data: a population-based validation study in Quebec, Canada', *Clinical Epidemiology*, 11:833–843, doi:10.2147/CLEP.S217969.

Ng E, Dasgupta K and Johnson JA (2008) 'An algorithm to differentiate diabetic respondents in the Canadian Community Health Survey', *Health Reports*, 19(1):71–79.

RACGP (The Royal Australian College of General Practitioners) (2020) *Management of type 2 diabetes: a handbook for general practice*, RACGP, East Melbourne, accessed 19 September 2022.

Rahimi A, Liaw ST, Taggart J, Ray P and Yu H (2014) 'Validating an ontology-based algorithm to identify patients with type 2 diabetes mellitus in electronic health records', *International Journal of Medical Informatics*, 83(10):768–778, doi:10.1016/j.ijmedinf.2014.06.002.

RCGP (Royal College of General Practitioners) (2011) *Coding, classification and diagnosis of diabetes*, RCGP, United Kingdom, accessed 23 October 2023.

Sharma M, Petersen I, Nazareth I and Coton SJ (2016) 'An algorithm for identification and classification of individuals with type 1 and type 2 diabetes mellitus in a large primary care database', *Clinical Epidemiology*, 8:373–380, doi:10.2147/CLEP.S113415.

Teltsch DY, Fazeli Farsani S, Swain RS, Kaspers S, Huse S, Cristaldi C, Nordstrom BL and Brodovicz KG (2019) 'Development and validation of algorithms to identify newly diagnosed type 1 and type 2 diabetes in pediatric population using electronic medical records and claims data', *Pharmacoepidemiology and Drug Safety*, 28(2):234–243, doi:10.1002/pds.4728.

van Walraven C, Bennett C and Forster AJ (2011) 'Administrative database research infrequently used validated diagnostic or procedural codes', *Journal of Clinical Epidemiology*, 64(10):1054–1059, doi:10.1016/j.jclinepi.2011.01.001.

Vanderloo SE, Johnson JA, Reimer K, McCrea P, Nuernberger K, Krueger H, Aydede SK, Collet JP and Amed S (2012) 'Validation of classification algorithms for childhood diabetes identified from administrative data', *Pediatric Diabetes*, 13(3):229–234, doi:10.1111/j.1399-5448.2011.00795.x.

Weisman A, Tu K, Young J, Kumar M, Austin PC, Jaakkimainen L, Lipscombe L, Aronson R and Booth GL (2020) 'Validation of a type 1 diabetes algorithm using electronic medical records and administrative healthcare data to study the population incidence and prevalence of type 1 diabetes in Ontario, Canada', *BMJ Open Diabetes Research & Care*, 8(1):e001224, doi:10.1136/bmjdrc-2020-001224.

Wells BJ, Lenoir KM, Wagenknecht LE, Mayer-Davis EJ, Lawrence JM, Dabelea D, Pihoker C, Saydah S, Casanova R, Turley C, Liese AD, Standiford D, Kahn MG, Hamman R and Divers J (2020) 'Detection of Diabetes Status and Type in Youth Using Electronic Health Records: The SEARCH for Diabetes in Youth Study', *Diabetes Care*, 43(10):2418–2425, doi:10.2337/dc20-0063.

Williamson T, Green ME, Birtwhistle R, Khan S, Garies S, Wong ST, Natarajan N, Manca D and Drummond N (2014) 'Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records', *Annals of Family Medicine*, 12(4):367–372, doi:10.1370/afm.1644.

List of tables

List of figures

aihw.gov.au

AIHW

Stronger evidence,
better decisions,
improved health and welfare