MEDICINEINSIGHT

Validation of the MedicineInsight database: completeness, generalisability and plausibility

December 2020 version 1.1

Independent, not-for-profit and evidence-based, NPS MedicineWise enables better decisions about medicines, medical tests and other health technologies.

Level 7/418A Elizabeth St
Surry Hills NSW 2010
PO box 1147
Strawberry Hills NSW 2012

P. 02 8217 8700
F. 02 9211 7578
info@nps.org.au
www.nps.org.au

Suggested citation

Busingye D, Myton R, Mina R, Thistlethwaite J, Belcher J, Chidwick K. MedicineInsight report: Validation of the MedicineInsight database: completeness, generalisability and plausibility. Sydney: NPS MedicineWise, 2020.

Acknowledgments

This report is funded by the Australian Government Department of Health. NPS MedicineWise was responsible for the design, analyses and publication of the results of the project. The Australian Government Department of Health had input into the design of the project but was not involved in the analysis and interpretation of the data.

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

Contents

EXECUTIVE SUMMARY

This report describes the completeness, generalisability and plausibility of selected items in the MedicineInsight data. Providing a more complete picture of the quality, strengths and limitations of MedicineInsight data will inform decisions made by NPS MedicineWise, researchers and policymakers about the context in which MedicineInsight data can be confidently used. This information will also guide NPS MedicineWise's strategy for future data quality improvement activities if significant data quality issues are identified. This will lead to a better data source for assessing issues such as quality use of medicines, vaccines and tests to inform decision-making by the Australian Government Department of Health, the Pharmaceutical Benefits Advisory Committee (PBAC), the Drug Utilisation Sub-Committee of the PBAC, the Therapeutic Goods Administration, the Medical Services Advisory Committee, policymakers and other relevant agencies.

Key findings

Study cohorts

- Eligible patients had at least one clinical encounter at a general practice that participates in the MedicineInsight program in the 2 years, ending 30 June in the financial year of interest (FY). 'Regular attenders' had three or more clinical encounters at a MedicineInsight practice in the 2 years, ending 30 June FY and 'infrequent attenders' had 1–2 clinical encounters at a MedicineInsight practice in the 2 years, ending 30 June FY.
- The number of all patients who were eligible for the study ranged from approximately 1.9 million (1.2 million regular attenders and 634,000 infrequent attenders) in 2010/11 to 3.2 million (1.9 million regular attenders and 1.2 million infrequent attenders) in 2019/20.
- The number of eligible general practices ranged from 379 in 2010/11 to 441 in 2019/20.

Completeness of selected variables

- The information in this report represents completeness of data recorded in fields accessible to MedicineInsight and may not indicate non-recording of data.

Data completeness rates over 10 years (FY 2010/11 to FY 2019/20)

- Data completeness rates for most of the variables from the Encounter (> 95%), Diagnosis (> 99%), Prescriptions (> 95%), Pathology (> 95%) and Medicare Benefits Schedule (MBS) billing (> 99%) tables were excellent across the 10-year study period.
- Over the 10-year study period substantial improvements in completeness rates were observed:
    - The proportion of encounters with a reason for encounter recorded improved markedly over the 10 years from 53.0% in 2010/11 to 71.2% in 2019/20.
    - The use of coded entries, rather than free text, in the reason for encounter field increased from 45.2% in 2010/11 to 74.5% in 2019/20.

- The use of coded entries in the Diagnosis (Medical History) table increased from 86.9% in 2010/11 to 90.5% in 2019/20.

- The completeness rates for the following important variables were consistently low across the 10-year study period:
  - Reason for prescription remained around 36% complete and reason for requested test at < 55%. While these variables are not fit for purpose when used alone, when used in combination with other variables, such as reason for encounter and diagnosis, their utility is much improved.
  - Visit type (eg, surgery consult, non-visit) was < 46% complete. However, among practices that use Best Practice software, visit type was 80.6% complete in 2019/20 compared with 24.8% among practices that use Medical Director software.
- As expected, completeness rates for most of the variables were greater for regular attenders compared with infrequent attenders.

Time constant variables assessed in 2020

- Most of the Patient variables assessed had excellent completeness rates (> 95%).
- Aboriginal or Torres Strait Islander status was recorded for 82.4% of regular attenders and 72.6% of infrequent attenders.
- Smoking status was recorded for 80.9% of regular attenders and 58.9% of infrequent attenders, while 'SMOKING_CEASED_DATE' was only recorded for 41.8% of ex-smokers (regular or infrequent attenders).
- The completeness rate for the 'YEAR_OF_DEATH' variable was less than 1%.
- Completeness rates were excellent for Site variables (> 98%) and good for Clinical user variables (> 87%).

Risk factors and measurements

- For regular patients attending MedicineInsight practices from 1 July 2018 to 30 June 2020, the following were found:
  - Good completeness rates for smoking status (88.8%; aged ≥ 18 years), HbA_{1c} in the past 2 years (86.8%; patients with diabetes only), and relatively good rates for blood pressure in the past 2 years (79.7%; aged ≥ 40 years) and estimated glomerular filtration rate in the past 2 years (72.6%; aged ≥ 40 years).
  - Suboptimal completeness rates for total cholesterol in the past 2 years (65.7%; aged ≥ 40 years) and alcohol use (54.7%; aged ≥ 18 years).
  - Poor completeness rates for body mass index (BMI) or height and weight in the past 2 years (38.4%), allergy (28.4%), cardiovascular disease (CVD) risk score 'ever' (16.5%; aged ≥ 40 years), bone mineral density 'ever' (5.3%; aged ≥ 50 years) and spirometry 'ever' (2.9%; patients with chronic obstructive pulmonary disease (COPD)/asthma only). Depending on the research project, these observations may not be fit for purpose as a representative sample of all patients.

- Greater completeness rates were observed for all the assessed risk factors and measurements in regular attenders compared to infrequent attenders, with the following marked differences observed:
    - Blood pressure was recorded in the past 2 years for 79.7% of regular attenders aged ≥ 40 years, 2.5 times that for infrequent attenders (31.6%).
    - Cholesterol was recorded in the past 2 years for 65.7% of regular attenders aged ≥ 40 years, 7.5 times that for infrequent attenders (8.7%).
    - HbA_{1c} was recorded in the past 2 years for 86.8% of regular attenders with type 2 or unspecified diabetes, six times that for infrequent attenders with type 2 or unspecified diabetes (14.3%).

Generalisability of MedicineInsight data

- This report includes data from 441 eligible general practices, representing 5.5% of general practices nationally.
- MedicineInsight has national coverage of general practices among all states and territories, remoteness categories and primary health networks (PHNs), except for Western Queensland PHN. MedicineInsight coverage across most states (NSW, VIC, QLD, WA) is representative, however practices from Tasmania (15.7% coverage) are over-represented and practices from South Australia (1.8% coverage) are under-represented.
- Practices in inner and outer regional areas are somewhat over-represented and major cities slightly under-represented.
- Coverage of PHNs fluctuates around the national estimate of 5.5%, ranging from 2.5% to 7.0% coverage in the majority of PHNs. Practices from the Hunter New England and Central Coast PHN in NSW (22.0%) and the Tasmania PHN (15.7%) are over-represented.
- Approximately 3.2 million patients attended one of the 441 eligible general practices from 1 July 2018 to 30 June 2020 – 1.9 million of these were regular attenders and 1.2 million were infrequent attenders.
- The demographic profile of MedicineInsight patients and MBS data on all Australian patients who have visited a GP at least once are similar in terms of age, gender and socio-economic status.
- MedicineInsight regular female patients are slightly over-represented (56.1%) compared with MBS data (52.3%).
- Infrequent attenders are over-represented for younger patients aged < 40 years (63.5%) and under-represented for patients aged ≥ 40 years (36.5%) compared with MBS data (50% < 40 years).
- A similar proportion of MedicineInsight patients were identified and recorded as Aboriginal or Torres Strait Islander as in the ABS 2016 national census (3.0% vs 2.8%).

Plausibility

Implausible values

- To assess the plausibility of MedicineInsight data, we determined whether the results for selected patient characteristics agreed with established plausible values from external gold-standard sources.
- The majority of MedicineInsight patients who had at least one record for each of the patient characteristics (height, weight, and BMI for adults; systolic and diastolic blood pressure and encounters per day for all patients) had plausible results with < 1% of patients outside the plausible value range.

External validity of condition prevalence estimates

- The regular patient prevalence estimates for most of the conditions (atrial fibrillation, chronic kidney disease, COPD, type 1 diabetes, migraine, myocardial infarction, rheumatoid arthritis and stroke) align with the 2017–18 Australian Bureau of Statistics National Health Survey, or are slightly higher (anxiety, asthma, CVD, type 2 diabetes, heart failure, low back pain, osteoarthritis and osteoporosis), demonstrating good external validity.
- The prevalence of all chronic and acute conditions was higher among regular attenders than infrequent attenders. As such, the infrequent patient prevalence estimates were substantially lower than the national estimates, however these findings have not been adjusted for age.
- Depending on the age group and the condition of interest, the infrequent attender patients may not be fit for purpose when considering prevalence estimates.
- Conditions were underestimated in comparison with national sources when the modified method (code or free text record in diagnosis/medical history only) was used compared with the current method (code or free text record in diagnosis/medical history, reason for encounter and reason for prescription).

External validity of prescriptions issued

- At Anatomical Therapeutic Chemical level 1, the proportions of total prescriptions ordered for regular attenders closely matched the proportions of prescriptions dispensed on the Pharmaceutical Benefits Scheme (PBS), demonstrating good external validity.
- Cardiovascular medicines accounted for 31.5% of total prescriptions prescribed to MedicineInsight regular attenders and 31.3% of prescriptions dispensed on the PBS.
- Medicines for the nervous system, which include the analgesics, were the next most common prescriptions, accounting for 23.0% of total MedicineInsight prescriptions and 21.8% of PBS prescriptions. Medicines for the alimentary tract and metabolic system accounted for 14.6% and 15.9% of total prescriptions for MedicineInsight and the PBS, respectively.
- The proportions of total prescriptions for infrequent attenders do not align with the proportions of PBS prescriptions. In most cases the proportions were lower among infrequent attenders, except for genitourinary system and sex hormones (which include contraceptives) and anti-infectives for

systemic use. This is not surprising as infrequent attenders are younger and healthier and often present to a GP for the management of acute conditions.

Preliminary analysis of duplicate patients (uniqueness plausibility)

- Based on a preliminary analysis of a representative sample of 958,641 (30.3%) patients from the 2019/20 financial year cohort who were included in the linkage pilot project, 3.8% were identified as duplicate patients, matched to either one other patient (3.6%; n = 34,674) or more than one other patient (0.2%; n = 1597).
- Of the 34,674 duplicate patients who were matched to only one other patient, 10% were identified as the same patient within a practice site and 90% were identified as the same patient between practice sites.

Recommendations

- Consider measures to improve data entry at the general practice level (including data quality feedback reports, training modules for practice staff, modification to the clinical information system by vendors, incentives), with a focus on the following key variables: reason for prescription, reason for encounter, Aboriginal and Torres Strait Islander status, smoking status and smoking ceased date (for ex-smokers), alcohol use, BMI, CVD risk score and allergy/adverse events.
- Further research is required to understand whether high completeness rates for some variables at the practice level could be used as a marker of better data quality overall.
- When identifying conditions, it is advisable to use information from the diagnosis/medical history, reason for encounter and reason for prescription fields – both coded and free text.
- Consider targeted recruitment of new practices to the MedicineInsight program to improve geographical under-representation among PHNs with less than 3.5% coverage (ie, > 2% lower coverage than the national MedicineInsight average), including Adelaide (1.9%), Central and Eastern Sydney (2.6%), Country SA (< 2.0%), East Melbourne (3.0%), North Sydney (3.1%), South Eastern Melbourne (2.5%), South Western Sydney (1.9%), Western NSW (< 2.0%), Western Queensland (0.0%) and Western Sydney (2.4%).
- Depending on the research question, consider using the regular attender cohort, while acknowledging the potential impact of selection bias and generalisability on study estimates.
- In the Australian setting, where patients can attend multiple general practices, when estimating condition prevalence, it is prudent to attempt to exclude temporary and visitor patients, whose medical history might not be recorded, to avoid underestimates. Acknowledge the potential to overestimate condition prevalence when using a regular attender cohort, as regular attenders are more likely to have complete records, be older, and have more chronic conditions than infrequent attenders. Infrequent attenders are generally younger and healthier patients who visit less frequently and are less likely to have a regular GP. Depending on the research question, consider age and sex standardisation of the regular attender cohort to the Australian population or MBS patient population.

- Conduct future validation studies focusing on other domains of data quality such as the accuracy of the information recorded in MedicineInsight.
- Consider linkage of MedicineInsight to other datasets to identify duplicate patients and to improve the capture of data with low completeness rates. Linkage can also help capture care or information that occurs outside the MedicineInsight practices, such as: the Australian Institute of Health and Welfare's national death index to improve death recording, PBS data to capture prescriptions from outside MedicineInsight practices (other GPs and specialists), MBS data to capture MBS billing and test orders outside MedicineInsight practices, and hospital data/registries to more accurately identify serious health outcomes and hospital episodes.

1. BACKGROUND

1.1. MedicineInsight program

MedicineInsight is a large-scale database containing de-identified electronic health records (EHRs) from almost 700 participating general practices across Australia. 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 postmarket surveillance of medicines.

MedicineInsight uses third-party data extraction tools – GeneRic Health Network Information Technology for the Enterprise (GRHANITE)[1] and Precedence Health Care's INCA[2] – 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 database 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 healthcare 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.

The data collected through MedicineInsight are used for quality improvement activities, including customised reports and encrypted patient lists (that can be re-identified by practice staff only) to support practices in identifying which patients might benefit from changes in management. In this way, MedicineInsight is used to identify opportunities for practices to enhance the clinical care of their patients. The data are also used by NPS MedicineWise and external organisations for research and evaluation, program design and policy development. MedicineInsight data are only used and shared consistent with the principles of public good, including contributing to improving health outcomes for Australians.

Further information is available online: https://www.nps.org.au/medicine-insight.

1.2. Assessment of the validity of MedicineInsight data

The extent to which the findings of analyses of MedicineInsight data are a true reflection of general practice activities and patient health – and are trusted by clinicians, policymakers and researchers – depends on the quality and completeness of the data. MedicineInsight reflects everyday health care provided to patients within a sample of practices across Australia. MedicineInsight data are real-world data entered into the CIS by practice staff for the purposes of providing clinical care and administrative

activities within the practice. There are limited standards for data entry: the medical classification and coding systems used to record conditions differ between CISs ('Pyefinch' coding is available in BP, while 'Docle' coding is available in MD); coding is not mandatory; and terms can also be entered as free text. This means that patient information could be selectively or inaccurately recorded.

NPS MedicineWise already works with practices to improve data quality in multiple ways. For example, when practice quality improvement reports are developed as part of the implementation of national therapeutic educational programs, the quality of the data is checked with sentinel practices to ensure there is correct identification of patients, medicines, tests, conditions and other relevant data elements. Previous research examining the validity of MedicineInsight flags for five medical conditions (anxiety, asthma, depression, osteoporosis and type 2 diabetes) found these measures were highly accurate when compared with gold-standard EHRs.[3]

However, providing a more complete picture of the quality, strengths and limitations of MedicineInsight data will inform decisions by NPS MedicineWise, researchers and policymakers about the context in which MedicineInsight data can be confidently used and guide NPS MedicineWise's strategy for future data quality improvements activities if significant data quality issues are identified. This will ultimately lead to a better data source for assessing issues such as the quality use of medicines, vaccines and tests to inform decision-making by the Australian Government Department of Health, the Pharmaceutical Benefits Advisory Committee (PBAC), the Drug Utilisation Sub-Committee of the PBAC (DUSC), the Therapeutic Goods Administration (TGA), the Medical Services Advisory Committee, policymakers and other relevant agencies.

1.3. Focus of the current study

The following categories of data quality[4,5] were assessed in this study:

1. **Completeness** – are data values present? Completeness is the degree to which all of the necessary data are included in the dataset, without reference to the accuracy of the data. Completeness measures the absence of data at a single moment in time, or when measured at multiple moments over time.[5] MedicineInsight data are dependent on the accuracy and completeness of data recorded in general practice CISs, in fields that can be extracted and are in a useable format. The completeness of data varies across practices and patients. For example, information about lifestyle risk factors, such as body mass index (BMI), may be selectively recorded for patients with particular health conditions but not recorded for other patients, or recorded in the progress notes which are not extracted.[6] Patients who attend a general practice infrequently, such as visitors or temporary patients, may have less complete records than patients who regularly attend a practice. For practices to be accredited, there are minimum standards for completeness of some key variables, such as allergies, BMI and smoking status, which could drive improvements in data quality. We might expect to see improvements in data quality over time as the use of EHRs becomes more embedded in general practice, and with the release of guidelines[7] for improving health record quality in primary care.

2. **Generalisability.** Generalisability is about understanding the scope and coverage of a dataset in terms of its relevance to the target population which it aims to describe.[4] MedicineInsight is drawn from a non-random sample of practices across Australia. Therefore, when used for research, information on the generalisability or representativeness of the results compared to the Australian patient population is important.[4] This is achieved by comparing the distribution of factors such as age, sex, socioeconomic status, geography and remoteness of patients in the MedicineInsight dataset with a national reference standard. Generalisability is particularly important when measuring prevalence and incidence but is not as important when measuring associations between variables, such as assessing the risk of heart attacks among patients taking anti-inflammatory medicines.
3. **Plausibility.** Plausibility seeks to determine if observed data values or distribution densities agree with 'common' knowledge (verification) or estimates from external sources that are deemed to be trusted or relative gold standards (validation).[5] This data quality criterion is also referred to as coherence.[4] Improving our understanding of the plausibility of data recorded in MedicineInsight will help provide an understanding of the contexts in which we can have confidence in the validity of estimates made from the data. Additionally, this work will further inform rules to be applied to improve data quality prior to conducting analyses.
4. **Duplicate patients (uniqueness plausibility)**. Understanding uniqueness in a dataset is about determining if subjects or observations appear multiple times in settings where they should not be duplicated or cannot be distinguished within a database (verification), or when compared with an external reference (validation).[5] Historically MedicineInsight has been unable to link patients across different practice sites and, consequently, there is potential for duplication of patient information within a practice and across practices in cases where patients attend multiple MedicineInsight practice sites.

1.4. Ethics approval for MedicineInsight

In December 2017, NPS MedicineWise was granted ethics approval for the standard operations and uses of the MedicineInsight database by NPS MedicineWise. This program approval was given by the Royal Australian College of General Practitioners (RACGP) National Research and Evaluation Ethics Committee (NREEC 17-017). This project is considered low risk and falls under this umbrella ethics approval.

The use of MedicineInsight data for the purposes of this report was approved on 20 July 2020 by the independent Data Governance Committee (2020–019).

2. AIMS AND METHODS

2.1. Aims

Aims of this study:

1. Describe the completeness of a series of key indicators related to patient demographics, patient risk factors, condition recording, prescriptions, pathology test results and billing data over time.
2. Assess for systematic differences in the type of patients who have more complete data by comparing completeness according to patient factors such as attendance at the practice, age and presence of chronic health conditions.
3. Quantify the frequency of practices with ≥ 50%, 80%, 90% and 99% completeness for selected variables to inform whether restricting to high completeness practices is feasible for some projects.
4. Document the generalisability of the patient sample in the MedicineInsight dataset by comparing observed age, sex, socioeconomic status, geographical location and remoteness of patients in the MedicineInsight database with relevant Australian national estimates.
5. Document the plausibility of observations from MedicineInsight by comparing observed condition prevalence and medicine utilisation estimates with Australian national estimates.
6. Present a preliminary analysis of the number of duplicate patients existing within the MedicineInsight dataset, based on the privacy preserving (bloom filter) record linkage proof-of-concept project.

2.2. Study design

This was a descriptive analysis of data from patients attending MedicineInsight general practices.

2.3. Study period

The study period for Aims 1–3 on the completeness of MedicineInsight data covered a 10-year period from 1 July 2010 to 30 June 2020 inclusive (Table 1).

The study period for Aims 4–6 (representativeness and plausibility) was based on the most recent 2 years of data (1 July 2018 to 30 June 2020), except where otherwise stated. Variables without historical data were assessed at the date of the data download on 31 October 2020.

Historical records outside of the study period were included when identifying patient demographics, and diagnoses.

TABLE 1: 10-YEAR STUDY PERIOD

Patient cohort (at least one clinical encounter in this period)	Study period	Financial Year (FY)
1 July 2009 – 30 June 2011	1 July 2010 – 30 June 2011	FY1
1 July 2010 – 30 June 2012	1 July 2011 – 30 June 2012	FY2
1 July 2011 – 30 June 2013	1 July 2012 – 30 June 2013	FY3
1 July 2012 – 30 June 2014	1 July 2013 – 30 June 2014	FY4
1 July 2013 – 30 June 2015	1 July 2014 – 30 June 2015	FY5
1 July 2014 – 30 June 2016	1 July 2015 – 30 June 2016	FY6
1 July 2015 – 30 June 2017	1 July 2016 – 30 June 2017	FY7
1 July 2016 – 30 June 2018	1 July 2017 – 30 June 2018	FY8
1 July 2017 – 30 June 2019	1 July 2018 – 30 June 2019	FY9
1 July 2018 – 30 June 2020	1 July 2019 – 30 June 2020	FY10

2.4. Study cohort

General practice sites

A general practice site describes one or more practices that share the same general practice database, either because they are operating within a common administrative system (eg, the same corporate entity) or in the same geographical area. De-identified patient data were obtained from 369 general practice sites, comprising 441 general practices, which met the standard data quality criteria in the MedicineInsight October 2020 data download. The majority (> 90%) of MedicineInsight practice sites are single general practices.

These standard data quality criteria were applied:

- the site had been established for at least 2 years, and
- had no significant interruptions of longer than 2 months in the 2 years prior to download of their practice data, and
- met the minimum threshold of clinical activity of at least 50 patients in the last 2 years.

Patient population

The study population for ascertaining completeness (**completeness population**; Aims 1–3) were patients who met the following inclusion criteria:

- visited a practice site that contributes data to MedicineInsight and met specific MedicineInsight data quality requirements
- had valid (non-missing, plausible) information for age (0–112 years) and sex (male, female, intersex/indeterminate) for each year of study, and
- had at least one clinical encounter (defined under section 2.5) in the 2 years, ending 30 June FY (for each financial year [FY] of interest from 1 July 2010 to 30 June 2020). For example, the patient cohort for FY10 are those who had at least one clinical encounter from 1 July 2018 to 30 June 2020 (Table 1).

For each financial year of the study, three sub-populations were included to help understand the impact of frequency of patient attendance at the general practice on completeness of the data:

1. **All patients** – with at least one clinical encounter in the 2 years, ending 30 June FY.
2. **Regular attenders** – with at least three clinical encounters in the 2 years, ending 30 June FY (a sub-set of all patients). This cohort aligns with the RACGP's definition of an 'active' patient.
3. **Infrequent attenders** – with 1–2 clinical encounters in the 2 years, ending 30 June FY (a sub-set of all patients).

The patient cohort for FY10 was used for ascertaining generalisability (**generalisability and plausibility population**; Aims 4–6). These patients had at least one clinical encounter between 1 July 2018 and 30 June 2020 (all patients), including patients with at least three clinical encounters (regular attenders) and those with 1–2 clinical encounters (infrequent attenders) to help understand the impact of frequency of patient attendance at the general practice on generalisability.

2.5. Definitions

Clinical encounters

A clinical encounter, or any professional exchange between a patient and a healthcare professional (GP or nurse), was defined as all those encounters at the practice site that are: a) not identified as administrator entries nor encounters that have been transferred/imported from another practice; b) are not identified by predefined 'administration-type' terms found in the 'reason for encounter' field such as 'administrative reasons', 'forms', and 'recall'; and c) are identified as being with a doctor or a nurse.

Completeness of selected variables (Aims 1-3)

A list of key variables in the MedicineInsight data to be assessed for completeness (Table 2) was selected in consultation with representatives from DUSC and the TGA.

The completeness of these variables was assessed in each financial year over a 10-year period from 1 July 2010 to 30 June 2020. The completeness of variables in the 'Patient' table were calculated as a proportion of all patients. The completeness of variables in the 'Site' and 'Practice recruitment' tables were calculated as a proportion of all general practices, and the completeness of variables from the 'Clinical user' table were calculated as a proportion of all clinical users. Denominators for completeness calculations for the encounter, diagnosis, pathology results header, pathology results detailed (atomised), investigations requested, prescription and Medicare Benefits Schedule (MBS) billing tables were the total number of encounters, diagnoses, pathology tests, prescriptions and MBS items, respectively, recorded for the study participants in the year of interest.

To be considered complete, the variable of interest should not be missing and should not have one of the CIS default entries recorded, which indicate the record is missing eg, '99999' or '01JAN1900'. Variables with 'not stated or inadequately described' entries were also considered as missing. The completeness definitions used for each variable are presented in Appendix A Table A1. Details of the MedicineInsight variables are available in the data dictionary, which can be provided upon request.

Each financial year that was assessed included a different patient cohort. For each financial year the '**all patients'** cohort included patients who had at least one clinical encounter in the previous 2 years and was stratified by **regular attenders'** and '**infrequent attenders'** sub-cohorts (see section 2.4). For some variables, such as sex and smoking status, only the most recent record in the CIS appears in the MedicineInsight database. Historical dated records of changes made to smoking status will be available in future releases from the MedicineInsight Data Warehouse. The completeness of these variables was described as at the date of the data download on 31 October 2020, rather than each financial year (Table 2).

TABLE 2: VARIABLES ASSESSED FOR COMPLETENESS AND REPORTING PERIOD

Data table [MedicineInsight table name]	Description of data table	Variable name	Time period for reporting
Patient *[EMI_PATIENT]*	Patient-specific information.	VALID_PATIENT_FLAG[a] CIS_PATIENT_STATUS_NAME CLIN_REGULAR_ATTENDER_FLAG[a] GENDER_NAME PATIENT_AGE[a] YEAR_OF_BIRTH YEAR_OF_DEATH DECEASED_INDICATOR[a] PATIENT_POSTCODE PATIENT_CITY PHN_CODE[a] ASGS_RA_NAME_2011 (remoteness)[a] ASGS_RA_NAME_2016 (remoteness)[a] SMOKING_STATUS_NAME SMOKING_CEASED_DATE ATSI_NAME PENSION_CODE	31 October 2020 (data download date)
Encounter *[EMI_ENCOUNTER]*	Information about recorded patient encounters including both clinical and administrative encounters	PROVIDER_ID VISIT_DATE VISIT_TYPE ENCOUNTER_REASON IS_CLINICAL[a] PROVIDER_IS_DR_OR_NURSE[a]	Each FY
Encounter reason *[EMI_ENCOUNTER_REASON]*	Reason for patient encounter	VISIT_DATETIME ENCOUNTER_REASON_CODE ENCOUNTER_REASON	Each FY
Diagnosis *[EMI_DIAGNOSIS]*	Patient diagnosis	PROVIDER_ID ADMIN_FLAG[a] DIAGNOSIS_REASON DIAGNOSIS_STATUS_ACTIVE_FLAG[a] DIAGNOSIS_TYPE CREATED_DATETIME DIAGNOSIS_DATE CIS_CODED_STATUS DIAGNOSIS_REASON_CODE DIFFERENTIAL_FLAG	Each FY

Data table [MedicineInsight table name]	Description of data table	Variable name	Time period for reporting
Pathology results header *[EMI_PATHOLOGY]*	General information regarding results (eg, pathology, radiology etc) received. Includes results from requests made by the practice, or from external providers who have copied results to the practice	RESULT_NAME COLLECTION_DATE REPORT_DATE IMPORT_DATE COMPLETION_FLAG NORMAL_FLAG	Each FY
Requested investigations *[EMI_REQUESTED_TEST]*	Details of any investigations requested through the CIS eg, pathology, radiology, ECG etc. (Does not contain any test results.)	REQUESTED_TESTS TEST_REASON REQUEST_DATE BILLING COPIES	Each FY
Pathology results detail *[EMI_PATHOLOGY_RESULT_ ATOM]*	Details of results for specific investigations, whether ordered individually or as a group. Includes results from requests made by the practice or from external providers who have copied results to the practice	PATHOLOGY_RESULT_ID RESULT_DATE DATA_TYPE LOINC_CODE RESULT_NAME RESULT_VALUE UNITS NORMAL_RANGE ABNORMAL_FLAG RECORD_STATUS CREATED_DATETIME UPDATED_DATETIME	Each FY
Medicine history *[EMI_PRESCRIPTION]*	Current and past history of medicines for a patient	PROVIDER_ID FIRST_DATE LAST_DATE MEDICINE_ACTIVE_INGREDIENT MEDICINE_NAME PRODUCT_NAME ATC_CODE[a] DOSE STRENGTH FORM ROUTE QUANTITY	Each FY

Data table [MedicineInsight table name]	Description of data table	Variable name	Time period for reporting
		FREQUENCY INSTRUCTIONS REPEAT_INTERVAL REPEATS PBS_STATUS RESTRICTION_CODE AUTHORITY_INDICATION PREVIOUS_AUTHORITY REASON REASON_CODE IS_CURRENT[a] RX_STATUS_LIMITED_MEDICATION RECORD_STATUS	
Prescription issued *[EMI_SCRIPT_ITEM]*	Each prescription printed from the CIS	SCRIPT_DATE MEDICINE_ACTIVE_INGREDIENT MEDICINE_NAME ATC_CODE[a] DOSE ROUTE STRENGTH FREQUENCY QUANTITY INSTRUCTIONS PRN REGULATION_24 REPEATS REPEAT_INTERVAL RESTRICTION_CODE AUTHORITY_INDICATION	Each FY
MBS billing *[EMI_BILLING_SERVICE]*	Description of MBS codes billed to the patient	ITEM_NUMBER SERVICE_PATIENT_COUNT SERVICE_RECORD_STATUS CREATED_DATETIME SERVICE_DATETIME UPDATED_DATETIME VISIT_DATETIME	Each FY
Site *[EMI_SITE]*	Descriptors of practice sites	STATE PHN_CODE[a] MULTI_PRACTICE_FLAG[a] ASGS_RA_CODE_2016 (remoteness)[a]	31 October 2020
Practice recruitment *[EMI_PRACTICE_RECRUITMENT]*	Details of recruited practices and participation status	PRACTICE_CATEGORY[a] PRACTICE_NUMBER_OF_GP[a] PRACTICE_POSTCODE PRACTICE_STATE PRACTICE_SUBURB	31 October 2020
Clinical user *[EMI_CLINICAL _USER]*	Clinical users are those with a login account created on the clinical system eg, practice staff, nurses,	CLINICAL_USER_TYPE_NAME DOCTOR_INDICATOR[a] NURSE_INDICATOR[a]	31 October 2020

Data table [MedicineInsight table name]	Description of data table	Variable name	Time period for reporting
	doctors, nurse practitioners		

[a] Derived by NPS MedicineWise

Completeness was assessed for selected patient risk factors and measurements/tests of interest in relevant sub-populations based on age, gender or clinical characteristics, as detailed in Table 3. The definitions used in assessing completeness for these selected patient characteristics are provided in Appendix A Table A2. For the condition algorithms that have been developed by NPS MedicineWise, completeness of recording was explored based on the fields where this information was recorded. For a selection of chronic conditions (Table 3) we described the proportion of patients with the condition ever recorded, stratified by:

- whether the condition was ever recorded (coded or as free text) in the diagnosis/medical history field or in the reason for encounter or reason for prescription fields (current method)
- whether the condition was ever recorded (coded or as free text) in the diagnosis/medical history field only (modified method).

Selected acute conditions recorded in the last financial year of study were also assessed (Table 3). Relevant terms for the included conditions are shown in Appendix A Table A3.

TABLE 3: SPECIFIC PATIENT CHARACTERISTICS ASSESSED FOR COMPLETENESS

Characteristic	Type	Additional notes	Time period for reporting
Risk factors/measurements	BMI	Either BMI or a height and weight available By sex and 10-year age groups By selected chronic conditions	Recorded in the last 2 years (1 July 2018 to 30 June 2020)
	Smoking status	By sex and 10-year age groups By selected chronic conditions	31 October 2020
	Current alcohol status	By sex and 10-year age groups By selected chronic conditions	31 October 2020
	BMD	50+ years By sex By selected chronic conditions	Ever recorded up to 30 June 2020
	CVD risk score	40+ years By sex By selected chronic conditions	Ever recorded up to 30 June 2020
	Blood pressure	40+ years By sex By selected chronic conditions	Recorded in the last 2 years (1 July 2018 to 30 June 2020)
	Spirometry	Patients with COPD and/or asthma only	Ever recorded up to 30 June 2020
	eGFR	40+ years By sex	Recorded in the last 2 years (1 July 2018 to 30 June 2020)

Characteristic	Type	Additional notes	Time period for reporting
	Total cholesterol	40+ years By sex	Recorded in the last 2 years (1 July 2018 to 30 June 2020)
	HbA_{1c}	By diabetes	Recorded in the last 2 years (1 July 2018 to 30 June 2020)
	Allergy/adverse events		Ever recorded up to 30 June 2020
Condition algorithms (based on the conditions detail table)	Chronic conditions: • Hypertension • Myocardial infarction • Stroke • Venous thromboembolism • Diabetes type 2/unspecified • Osteoporosis • Depression • Anxiety • Dementia • COPD • Asthma	Stratification by: Condition flags based on a) the current algorithm which includes coded and free text diagnosis/medical history, reason for encounter and reason for prescription with modified versions b) coded and free text diagnosis/medical history only	Ever recorded up to 30 June 2020
	Acute conditions: • URTI • LRTI • UTI • Otitis media		Recorded in the last FY (1 July 2019 to 30 June 2020)

BMD = bone mineral density, BMI = Body Mass Index, COPD = chronic obstructive pulmonary disease, CVD = cardiovascular disease, eGFR = estimated glomerular filtration rate, FY = financial year, URTI = upper respiratory tract infection.

Practice-level completeness rates of ≥ 50%, ≥ 80%, ≥ 90% and ≥ 99% were further assessed for selected variables including Aboriginal and/or Torres Strait Islander status, MBS billing item, BMI and smoking status for patients in FY10.

Generalisability of practice and patient characteristics (Aim 4)

The representativeness of general practices in the MedicineInsight dataset was assessed by comparing geographical location (state/territory and PHN) and remoteness of practices in the MedicineInsight dataset with Australian national estimates.

The representativeness of patients in the MedicineInsight dataset was assessed by comparing sociodemographic characteristics (age group, gender, Aboriginal and Torres Strait Islander status, state/territory, rurality and socioeconomic status) of patients in the MedicineInsight dataset with Australian national estimates from the MBS data or other national sources (MBS data are provided by the Department of Health). Stratification by geographical location was presented where possible. Results were presented separately for all patients, regular attenders and infrequent attenders.

Plausibility of MedicineInsight data (Aim 5)

To assess plausibility of MedicineInsight data, we assessed key patient characteristics (Table 4) in the MedicineInsight database based on established plausible values and those that have previously been used for the MedicineInsight data portal (Appendix A Table A5). The references used for the plausible values are shown in Table 4.

The external validity of estimates for conditions and medicine utilisation in the MedicineInsight dataset was further assessed in relation to Australian national estimates. Patient prevalence estimates for selected chronic and acute conditions recorded in MedicineInsight were compared with population estimates from the Australian Bureau of Statistics (ABS) National Health Survey (NHS)[8] and when not available, estimates from existing literature were used. Further information about definitions used for selected conditions is available in Appendix A. MedicineInsight prescribing estimates reported at the Anatomical Therapeutic Chemical (ATC) level 1 were compared to the Pharmaceutical Benefits Scheme (PBS) medication dispensing estimates.[9] Details about MedicineInsight prescriptions data are provided in Appendix A.

Results for plausibility analyses were presented separately for all patients, regular attenders and infrequent attenders.

TABLE 4: PLAUSIBLE VALUES FOR SELECTED PATIENT CHARACTERISTICS

Characteristic [units]	Plausible values	
	Minimum	Maximum
Height (adult, ≥ 18 years) [cm][10]	112	251
Weight (adult) [kg][10]	25	610
BMI (adult) [kg/m^2][11]	12	70
Systolic blood pressure [mm Hg][10,12]	50	250
Diastolic blood pressure [mm Hg][10,12]	30	140
Clinical encounters per patient per day	1	3

Preliminary analysis of duplicate patients (Aim 6)

NPS MedicineWise commissioned a proof-of-concept project with the Centre for Data Linkage at Curtin University to provide privacy preserving (bloom filter) record linkage services for 150 general practices that contribute data to MedicineInsight. This proof-of-concept study was conducted using data from patients in the March 2020 data download. The output of this project was the creation of a single linkage map, identifying the same individuals within and between the included practice sites. We used the results of this pilot project, merged with the patients included in this report, to present a preliminary analysis describing the proportion of unique and duplicate patients identified, among the patient cohort for FY10 who had at least one clinical encounter from 1 July 2018 to 30 June 2020.

The proportion of patients who were matched to at least one other patient or more than one other patient was analysed. We also assessed the proportion of patients duplicated within and between practice sites. As a preliminary assessment of the quality of the privacy preserving linkage we measured the concordance (or agreement) of a selection of patient characteristics (year of birth,

gender, postcode and hypertension diagnosis) between patients who were matched to at least one other patient.

Sociodemographic characteristics

Sociodemographic characteristics included in the study are defined in Table 5.

TABLE 5: SOCIODEMOGRAPHIC DEFINITIONS

Characteristic	Definition
Age	Age was calculated on 1 July of each year based on the patient's date of birth (defined as 1 July in the patient's year of birth) and presented as 10-year age groups. Valid age was defined as 0–112 years.
Gender	As recorded in the CIS (male, female or intersex/indeterminate).
Aboriginal and Torres Strait Islander status	As recorded in the CIS.
State or territory in Australia	State was assigned based on each patient's postcode of residence. If patient postcode was missing, the practice postcode was used as a proxy.
Rurality/remoteness	Rurality was assigned based on mapping of each patient's postcode of residence using the ABS mapping of postcode 2016 to the ASGS Remoteness Areas 2016 data.[13]
Socioeconomic status (SEIFA)	SEIFA was assigned based on mapping of each patient's postcode of residence using the ABS mapping of postcode 2016 to the IRSAD.[14]

ABS = Australian Bureau of Statistics, ASGS = Australian Statistical Geography Standard, CIS = clinical information system, IRSAD = Index of Relative Socioeconomic Advantage and Disadvantage, SEIFA = Socio-Economic Indexes for Areas

2.6. Data analysis and reporting

Analysis of the data was conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA). Measures included are descriptive statistics, frequencies and proportions as appropriate. To indicate the reliability of the estimates of proportions, 95% confidence intervals were included as needed. Robust errors were used to adjust for clustering by practice site when calculating confidence intervals.

If a particular result was only reported in 1–4 patients or practices, this result was reported as < 5 in order to preserve the privacy of individuals and practices (with the exception of missing variables).

3. STUDY COHORTS

- The number of all patients who were eligible for the study ranged from about 1.9 million (1.2 million regular attenders and 634,000 infrequent attenders) in FY1 to 3.2 million (1.9 million regular attenders and 1.2 million infrequent attenders) in FY9 and FY10.
- The number of eligible general practices ranged from 379 in FY1 to 441 in FY9 and FY10.

3.1. Patient cohorts and general practices

The patient cohorts for each financial year are presented in Table 6. The number of eligible general practices ranged from 379 in FY1 to 441 in FY9 and FY10 (Table 6). The number of all patients who were eligible for the study ranged from approximately 1.9 million (1.2 million regular attenders and 634,000 infrequent attenders) in FY1 to 3.2 million (1.9 million regular attenders and 1.2 million infrequent attenders) in FY9 and FY10. Figure 1 shows the selection flowchart for the patient cohorts in FY10. The sociodemographic characteristics of the 'all patients', 'regular attenders' and 'infrequent attenders' cohorts for each financial year are presented in Appendix B.

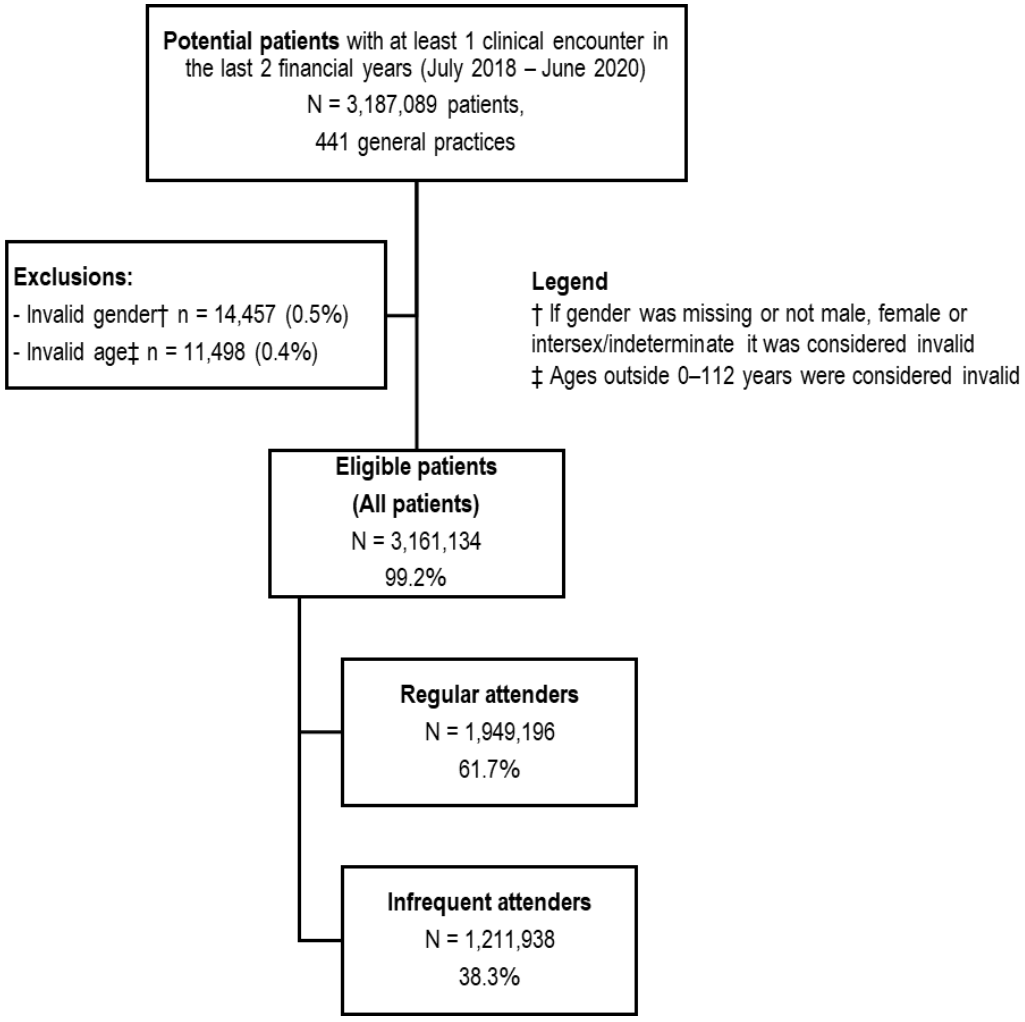

FIGURE 1: STUDY SELECTION FLOWCHART FOR THE FY10 2019/20 COHORT

The proportion of patients excluded from the study due to invalid (missing) gender increased over the 10-year study, from 0.1% of the total cohort in FY1 to 0.5% in FY10 (Table 6). The proportion of patients excluded from the study due to invalid age remained stable at 0.4% to 0.5%.

TABLE 6: PATIENT COHORT SELECTION CRITERIA FOR EACH FINANCIAL YEAR OF STUDY

Category	FY1, 10/11	FY2, 11/12	FY3, 12/13	FY4, 13/14	FY5, 14/15	FY6, 15/16	FY7, 16/17	FY8, 17/18	FY9, 18/19	FY10, 19/20
Potential patients with at least one clinical encounter in the last 2 FYs	1,864,822	2,330,053	2,484,102	2,612,458	2,729,651	2,870,073	3,006,362	3,116,698	3,194,932	3,187,089
Invalid gender (excluded; % of potential patients)	1,407 (0.1%)	2,558 (0.1%)	2,689 (0.1%)	2,842 (0.1%)	3,202 (0.1%)	3,674 (0.1%)	5,413 (0.2%)	7,748 (0.2%)	9,115 (0.3%)	14,457 (0.5%)
Invalid age (excluded; % of potential patients)	9,242 (0.5%)	10,492 (0.5%)	10,644 (0.4%)	10,975 (0.4%)	11,439 (0.4%)	12,648 (0.4%)	12,456 (0.4%)	12,585 (0.4%)	12,741 (0.4%)	11,498 (0.4%)
Eligible patients (All patients cohort)	1,854,173	2,317,003	2,470,769	2,598,641	2,715,010	2,853,751	2,988,493	3,096,365	3,173,076	3,161,134
Regular attenders cohort (% of eligible patients)	1,220,576 (65.8%)	1,375,787 (59.4%)	1,483,697 (60.1%)	1,572,564 (60.5%)	1,649,365 (60.7%)	1,734,080 (60.8%)	1,810,354 (60.6%)	1,875,033 (60.6%)	1,930,420 (60.8%)	1,949,196 (61.7%)
Infrequent attenders cohort (% of eligible patients)	633,597 (34.2%)	941,216 (40.6%)	987,072 (39.9%)	1,026,077 (39.5%)	1,065,645 (39.3%)	1,119,671 (39.2%)	1,178,139 (39.4%)	1,221,332 (39.4%)	1,242,656 (39.2%)	1,211,938 (38.3%)
General practices	379	389	398	406	418	428	437	440	441	441

Eligible patients had at least one clinical encounter in the previous 2 years, ending 30 June FY; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.
FY = financial year.

4. DATA COMPLETENESS

- Completeness rates for most of the variables from the Encounter (> 95%), Diagnosis (> 99%), Prescriptions (> 95%), Pathology (> 95%) and MBS billing (> 99%) tables were excellent across the 10-year study period.
- Over the 10-year study period the following improvements in completeness rates were observed:
  - The proportion of encounters with a reason for encounter recorded improved markedly over the 10 years, from 53.0% in FY 2010/11 to 71.2% in FY 2019/20.
  - The use of coded entries in the reason for encounter field, rather than free text, increased from 45.2% of reason for encounter records in FY 2010/11 to 74.5% in FY 2019/20.
  - The use of coded entries in the Diagnosis (Medical History) table increased from 86.9% of recorded diagnoses in FY 2010/11 to 90.5% in FY 2019/20.
- The completeness rates for the following important variables were consistently low across the 10-year study period:
  - Reason for prescription remained around 36% complete, reason for requested test (< 55%) and visit type (eg, surgery consult, non-visit) (< 46%).
- As expected, completeness rates for most of the variables were greater in regular attenders compared with infrequent attenders.
- The majority of the Patient variables assessed had excellent recording rates (> 95%) except for 'YEAR_OF_DEATH' where the recording rates were less than 1%.
- Aboriginal and Torres Strait Islander status was recorded for 82.4% of regular attenders and 72.6% of infrequent attenders.
- Smoking status was recorded for 80.9% of regular attenders and 58.9% of infrequent attenders, while smoking ceased date was only recorded for 41.8% of all ex-smokers (regular or infrequent attenders).
- Completeness rates were excellent for Site variables (> 98%) and good for Clinical user variables (> 87%).
- Among regular attenders, good recording rates were observed for smoking status (88.8%; aged ≥ 18 years), HbA_{1c} (86.8%; diabetes patients only) and relatively good rates for blood pressure (79.7%; aged ≥ 40 years) and estimated glomerular filtration rate (eGFR; 72.6%; aged ≥ 40 years).
- Completeness rates for total cholesterol (65.7%; aged ≥ 40 years) and alcohol use (54.7%; aged ≥ 18 years) were sub-optimal in regular attenders.
- Among regular attenders, poor recording rates were observed for BMI or height and weight (38.4%), allergy (28.4%), cardiovascular disease (CVD) risk score (16.5%; aged ≥ 40 years), bone mineral density (5.3%; aged ≥ 50 years) and spirometry (2.9%; chronic obstructive pulmonary disease [COPD]/asthma patients only).
- Completeness rates for all the assessed risk factors and measurements were greater in regular compared with infrequent attenders, with the following marked differences observed:
  - Blood pressure was recorded in the past 2 years for 79.7% of regular attenders aged ≥ 40 years, 2.5 times that for infrequent attenders (31.6%).
  - Cholesterol was recorded in the past 2 years for 65.7% of regular attenders aged ≥ 40 years, 7.5 times that for infrequent attenders (8.7%).
  - HbA_{1c} was recorded in the past 2 years for 86.8% of regular attenders with type 2 or unspecified diabetes, 6 times that for infrequent attenders (14.3%).
- Condition prevalence estimates were lower in infrequent attenders compared with regular attenders.
- Condition prevalence estimates were lower in comparison with national sources when the modified method (code or free text record in diagnosis/medical history only) was used compared to the current method (code or free text record in diagnosis/medical history, reason for encounter and reason for prescription).

4.1. Study questions

- What are the completeness rates for a series of selected variables related to patient demographics, patient risk factors, condition recording, prescriptions, pathology test results and billing data over time?
- Are there systematic differences in recording of the data in relation to patient factors such as attendance at the practice, age, gender and presence of chronic health conditions?
- How many practice sites have completeness rates of ≥ 50%, ≥ 80%, ≥ 90% and ≥ 99% for selected variables including Aboriginal and/or Torres Strait Islander status, BMI, smoking status and MBS billing item.

4.2. Completeness of selected variables

Time constant variables assessed in 2020

Table 7 shows completeness rates for selected variables for which only the most recent record in the CIS appears in the MedicineInsight database and there are no historical records of changes made. The completeness of these variables was described for 3,161,134 patients in FY10, 369 general practice sites (representing 441 individual general practices) and 150,379 clinical users as at the date of the data download, 31 October 2020, rather than each financial year. The completeness of variables in the 'Site', and 'Practice recruitment' tables were assessed as a proportion of all practices and the completeness of variables from the 'Clinical user' table was assessed as a proportion of all clinical users. Variables that are derived by NPS MedicineWise (as indicated in the tables) usually have a 100% completeness rate.

Most patient variables assessed had high (> 80%) recording rates. The recording rates for 'YEAR_OF_DEATH' were less than 1% (among all patient cohorts), with a separate validation study in progress to assess the accuracy of death recording in MedicineInsight. Mostly, recording rates were higher in regular attenders compared with infrequent attenders. Aboriginal and Torres Strait Islander status was recorded for 82.4% of regular attenders and 72.6% of infrequent attenders. Smoking status was recorded for 80.9% of regular attenders and 58.9% of infrequent attenders, while smoking ceased date was only recorded for 41.8% of all ex-smokers (regular or infrequent attenders) (Table 7).

High completeness rates were observed for Site variables (> 98%) and Clinical user variables (> 87%) (Table 7). Among the variables assessed from the Practice recruitment table, the recording rate for 'PRACTICE_CATEGORY' was less than 50% while just over two-thirds (67.8%) of the practices had 'PRACTICE_NUMBER_OF_GP' recorded. These two variables are derived by NPS MedicineWise from information requested from practices when they join the MedicineInsight program. The MedicineInsight Practice Agreement includes a space to record the type of practice (eg, private, large corporate, community health service, Aboriginal health service) and the number of GPs at the practice, however these fields are not mandatory and are often left blank.

TABLE 7: COMPLETENESS OF SELECTED VARIABLES ASSESSED AT 31 OCTOBER 2020, STRATIFIED BY PATIENT COHORT

Selected variables	% completeness of variables		
	All patients	Regular attenders	Infrequent attenders
Patient variables	**N = 3,161,134**	**N = 1,949,196**	**N = 1,211,938**
VALID_PATIENT_FLAG[a]	100.0%	100.0%	100.0%
CIS_PATIENT_STATUS_NAME	100.0%	100.0%	100.0%
CLIN_REGULAR_ATTENDER_FLAG[a]	100.0%	100.0%	100.0%
GENDER_NAME[b]	100.0%	100.0%	100.0%
PATIENT_AGE[a,b]	100.0%	100.0%	100.0%
YEAR_OF_BIRTH	100.0%	100.0%	100.0%
YEAR_OF_DEATH	0.6%	0.8%	0.2%
DECEASED_INDICATOR[a]	100.0%	100.0%	100.0%
PATIENT_POSTCODE	99.7%	99.9%	99.4%
PATIENT_CITY	100.0%	100.0%	100.0%
PHN_CODE[a]	100.0%	100.0%	100.0%
ASGS_RA_NAME_2011[a]	99.3%	99.5%	98.9%
ASGS_RA_NAME_2016[a]	99.3%	99.5%	98.9%
SMOKING_STATUS_NAME	72.5%	80.9%	58.9%
SMOKING_CEASED_DATE (ex-smokers only)	41.8%	41.5%	43.0%
ATSI_NAME	78.6%	82.4%	72.6%
PENSION_CODE	80.4%	81.4%	78.8%
Site variables	**All practice sites (N = 369)**		
STATE	100.0%		
PHN_CODE[a]	100.0%		
MULTI_PRACTICE_FLAG[a]	100.0%		
ASGS_RA_CODE_2016 (remoteness)[a]	99.7%		
MBS_BILLING_ITEM	98.1%		
Practice recruitment variables	**All general practices (N = 441)**		
PRACTICE_CATEGORY[c]	45.1%		
PRACTICE_NUMBER_OF_GP[c]	67.8%		
PRACTICE_POSTCODE	100.0%		
PRACTICE_STATE	100.0%		
PRACTICE_SUBURB	100.0%		
Clinical user variables	**All clinical users (N = 150,379)**		
CLINICAL_USER_TYPE_NAME	87.1%		
DOCTOR_INDICATOR[a]	99.6%		
NURSE_INDICATOR[a]	99.6%		

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] Derived by NPS MedicineWise

[b] Variable used in cohort selection.

[c] Derived by NPS MedicineWise from information collected from practices when they join the MedicineInsight program.

Completeness of key general practice activities in the last financial year, 1 July 2019 to 30 June 2020

Key general practice activities among patients, such as having at least one encounter, diagnosis, prescription, or test result recorded within a year were assessed (see Appendix A, Table A2). These activities are not expected to be 100% complete as, for example, not all patients will need a prescription or a test within a year. Completeness rates for these key activities in the last year were

greater in regular attenders compared to infrequent attenders (Figure 2 and Table 8). Among regular attenders, more than 80% had at least one encounter, encounter reason and MBS billing service recorded in the period 1 July 2019 to 30 June 2020; 70.3% had at least one issued prescription; 61.8% a pathology result; and 54.1% a recorded diagnosis. Only 51.2% of the infrequent attenders had an encounter recorded in the same time period, only 20.6% had a diagnosis and 21.8% had a prescription (21.8%) recorded.

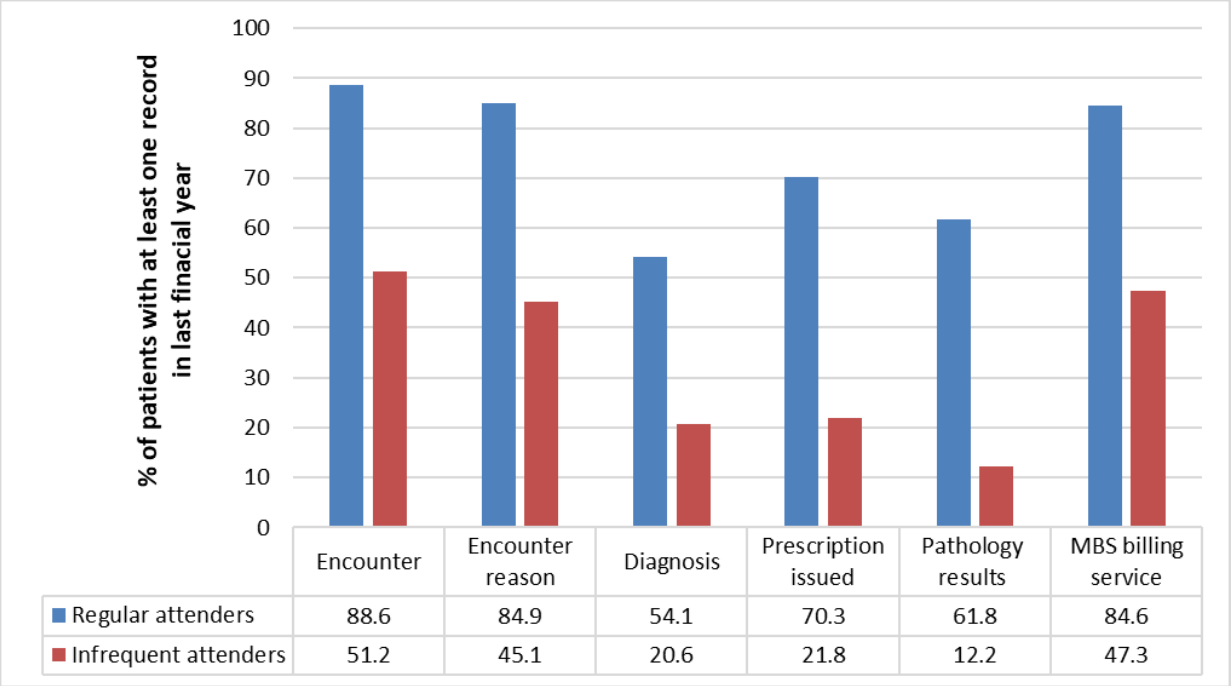

FIGURE 2: KEY GENERAL PRACTICE ACTIVITIES RECORDED IN FY10 (1 JULY 2019 TO 30 JUNE 2020) IN REGULAR AND INFREQUENT ATTENDERS

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

TABLE 8: KEY GENERAL PRACTICE ACTIVITIES RECORDED IN THE LAST FINANCIAL YEAR (FY10), STRATIFIED BY PATIENT COHORT

Key variable	% of patients in FY10 with at least one record during 1 July 2019 to 30 June 2020		
	All patients N = 3,161,134	Regular attenders N = 1,949,196	Infrequent attenders N = 1,211,938
Has record in last FY for:			
Encounter	74.3%	88.6%	51.2%
Encounter reason	69.6%	84.9%	45.1%
Diagnosis	41.3%	54.1%	20.6%
Prescription issued	51.7%	70.3%	21.8%
Pathology results	42.8%	61.8%	12.2%
MBS billing service	70.3%	84.6%	47.3%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

Completeness of variables over 10 years (1 July 2010 to 30 June 2020)

Completeness rates for selected variables from various MedicineInsight data tables assessed each year, from 1 July 2010 to 30 June 2020, are presented in Tables 9–11. The denominators used for each data table and financial year are presented in Appendix C. Completeness rates for most of the

variables from the Encounter (> 95%), Diagnosis (> 99%), Prescriptions (> 95%), Pathology (> 95%) and MBS billing (> 99%) tables were excellent across the 10-year study period (Table 9). Completeness rates for most variables were greater among regular attenders (Table 10) than infrequent attenders (Table 11).

Recording rates for most of the encounter variables assessed were high (> 95%) except for 'ENCOUNTER_REASON' and 'VISIT_TYPE' which both increased substantially across the 10-year study period (Figures 3 and 4). The proportion of encounters with an 'ENCOUNTER_REASON' recorded improved over the 10 years from 53.0% in the first year to 71.2% in the final year (Table 9). Interestingly, the recording rates for 'ENCOUNTER_REASON' were greater among infrequent attenders than regular attenders (Figure 3). A potential explanation is that infrequent attenders might be more likely than regular attenders to present for a new problem that the GP records as the reason for encounter, whereas regular attenders are more likely to have chronic diseases managed and repeat prescriptions ordered, which might not be recorded in the reason for encounter at every visit.

Of the encounters with an 'ENCOUNTER_REASON' recorded (ie, all those included in the 'Encounter reason' table) 45.2% had a code (Docle or Pyefinch) recorded in the first year and this improved to 74.5% in the final year (Table 9). In other words, 54.8% of the ''ENCOUNTER_REASON' records were free text in 2010/11 and by 2019/20 only 25.5% were free text records. The use of codes is much higher in the Diagnosis (Medical History) table with 86.9% of 2010/11 records having a 'DIAGNOSIS_REASON_CODE', increasing to 90.5% in 2019/20 (Table 9).

At the beginning of the study period the 'VISIT_TYPE' field, which allows staff to record information on the encounter setting (eg, surgery consultation, practice admin) was only available in the MD CIS. As such this variable was only 28.6% complete overall in the first year, 60.8% complete for MD practices, and totally missing for BP practices (Figure 4). By the final year, the recording of 'VISIT_TYPE' in MD had increased to 80.6%. In 2016 this field became available in the BP CIS and completeness rates for BP practices increased from 5.0% in FY6 to 24.8% in FY10. While this field isn't mandatory, some practices may set the default VISIT_TYPE to 'surgery consultation' which will impact the accuracy of the data in this field.

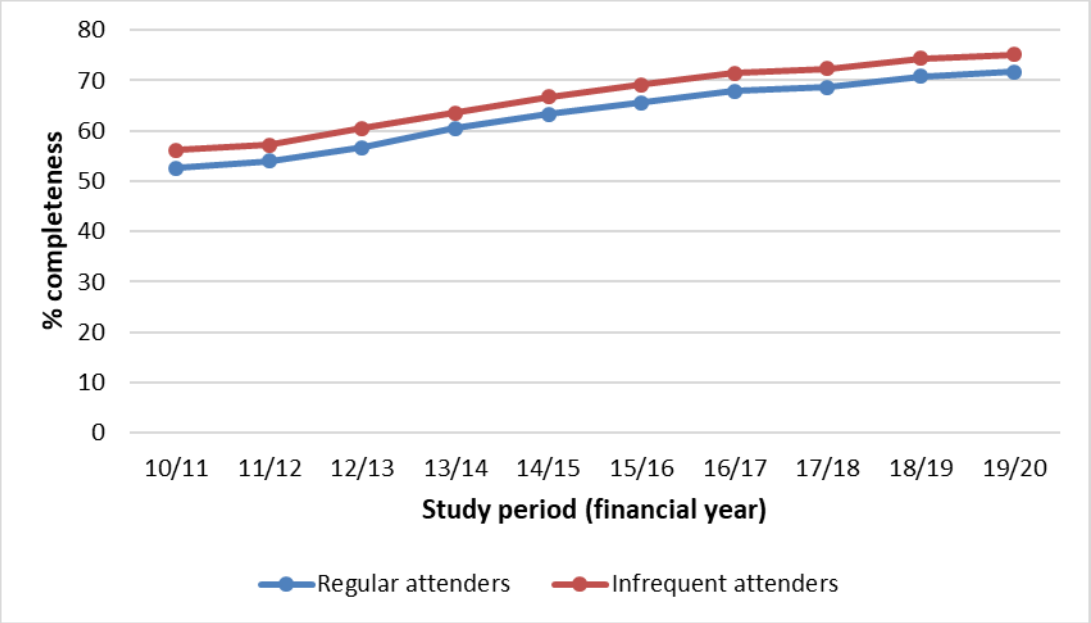

FIGURE 3: COMPLETENESS RATES FOR 'ENCOUNTER_REASON' IN REGULAR AND INFREQUENT ATTENDERS ACROSS THE 10-YEAR STUDY PERIOD (1 JULY 2010 TO 30 JUNE 2020)

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June FY.

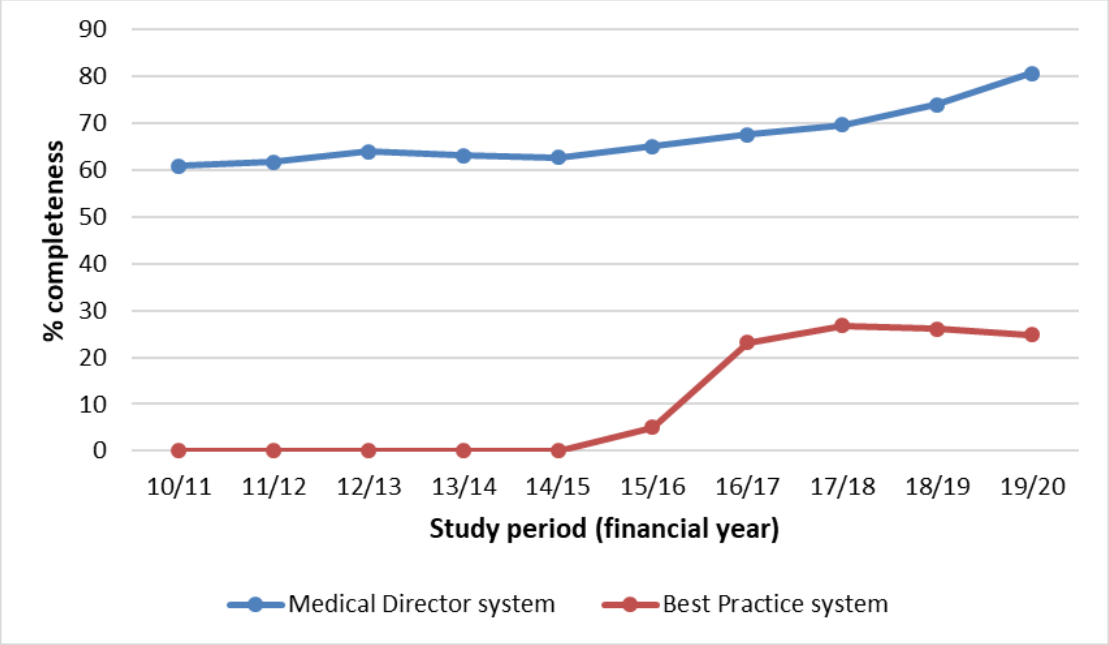

FIGURE 4: COMPLETENESS RATES FOR 'VISIT_TYPE' IN ALL PATIENTS BY MEDICAL DIRECTOR AND BEST PRACTICE CLINICAL INFORMATION SYSTEMS ACROSS THE 10-YEAR STUDY PERIOD (1 JULY 2010 TO 30 JUNE 2020)

TABLE 9: COMPLETENESS RATES FOR SELECTED VARIABLES ASSESSED EACH FINANCIAL YEAR FOR THE 'ALL PATIENTS' COHORT (1 JULY 2010 TO 30 JUNE 2020)

Selected variables	% completeness of selected variables for each financial year among all patients									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
Encounter										
PROVIDER_ID	99.3%	99.1%	99.2%	99.2%	99.3%	99.2%	99.2%	99.1%	99.4%	98.3%
VISIT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
VISIT_TYPE	28.6%	28.5%	28.7%	28.0%	27.8%	31.0%	41.7%	44.1%	44.8%	45.4%
VISIT_TYPE_BP	0.0%	0.0%	0.0%	0.0%	0.0%	5.0%	23.2%	26.8%	26.0%	24.8%
VISIT_TYPE_MD	60.8%	61.7%	63.9%	63.1%	62.7%	65.0%	67.6%	69.7%	73.9%	80.6%
ENCOUNTER_REASON	53.0%	54.3%	57.0%	60.7%	63.6%	65.9%	68.1%	69.0%	71.0%	71.9%
IS_CLINICAL[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
PROVIDER_IS_DR_OR_NURSE[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Encounter reason										
VISIT_DATETIME[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
ENCOUNTER_REASON	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
ENCOUNTER_REASON_CODE	45.2%	56.2%	64.1%	68.4%	70.3%	71.7%	71.6%	72.9%	73.7%	74.5%
Diagnosis										
ADMIN_FLAG[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_REASON	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_REASON_CODE	86.9%	87.3%	87.6%	88.0%	88.2%	88.2%	88.8%	89.7%	90.0%	90.5%
DIAGNOSIS_STATUS_ACTIVE_FLAG	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_TYPE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CREATED_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CIS_CODED_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIFFERENTIAL_FLAG	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Pathology results header										
RESULT_NAME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
COLLECTION_DATE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Selected variables	% completeness of selected variables for each financial year among all patients									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
REPORT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
IMPORT_DATE	94.0%	96.2%	99.6%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
COMPLETION_FLAG	97.7%	97.9%	98.3%	98.1%	97.9%	98.3%	98.3%	98.4%	98.4%	98.6%
NORMAL_FLAG	71.3%	71.6%	71.8%	72.1%	71.4%	72.7%	73.7%	74.4%	74.8%	75.9%
Pathology results detail										
PATHOLOGY_RESULT_ID	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RESULT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DATA_TYPE	96.3%	97.7%	98.4%	99.0%	99.4%	99.5%	99.6%	99.8%	99.9%	100.0%
LOINC_CODE	97.5%	97.7%	97.6%	97.8%	97.7%	97.6%	97.3%	97.1%	97.3%	97.3%
RESULT_NAME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RESULT_VALUE	100.0%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%	99.7%	99.2%
UNITS	94.0%	93.6%	93.3%	92.7%	92.7%	92.3%	92.6%	94.4%	94.7%	94.8%
NORMAL_RANGE	89.0%	90.1%	90.8%	90.4%	90.0%	89.3%	89.9%	92.0%	92.6%	92.2%
ABNORMAL_FLAG	29.5%	31.6%	31.7%	31.9%	31.3%	30.3%	30.5%	30.3%	29.4%	28.2%
RECORD_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CREATED_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
UPDATED_DATETIME	62.1%	54.5%	48.4%	46.4%	45.7%	44.0%	42.6%	41.2%	40.0%	38.2%
Requested investigations										
REQUESTED_TESTS	99.4%	99.5%	99.6%	99.8%	99.8%	99.8%	99.8%	99.8%	99.8%	99.7%
TEST_REASON	49.5%	51.3%	52.0%	52.6%	53.0%	52.5%	54.0%	54.6%	54.1%	54.5%
REQUEST_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
BILLING	28.4%	29.6%	29.4%	30.5%	30.5%	30.7%	31.8%	32.4%	32.8%	33.7%
COPIES	0.4%	0.5%	0.6%	0.7%	0.8%	0.8%	0.8%	0.9%	0.9%	0.9%
Medicine history										
CREATED_BY	94.0%	96.6%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
FIRST_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
LAST_DATE	94.4%	94.4%	94.5%	94.6%	94.6%	94.4%	94.1%	94.0%	93.9%	93.4%
MEDICINE_ACTIVE_INGREDIENT	99.2%	99.1%	99.1%	99.1%	99.1%	99.1%	99.1%	99.1%	99.2%	99.2%

Selected variables	% completeness of selected variables for each financial year among all patients									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
MEDICINE_NAME	99.6%	99.5%	99.5%	99.5%	99.5%	99.5%	99.5%	99.5%	99.5%	99.4%
PRODUCT_NAME	51.7%	50.5%	49.6%	48.8%	48.3%	47.0%	46.2%	45.0%	44.1%	42.2%
ATC_CODE[b]	93.7%	93.6%	93.4%	93.2%	93.1%	93.1%	93.0%	93.0%	93.1%	93.1%
DOSE	89.1%	88.0%	87.9%	87.9%	87.5%	87.0%	86.9%	86.2%	85.8%	85.4%
STRENGTH	96.4%	96.4%	96.1%	96.0%	95.9%	95.6%	95.6%	95.4%	95.6%	95.7%
FORM	73.8%	73.6%	73.5%	74.1%	74.0%	73.8%	74.3%	74.5%	74.0%	73.3%
ROUTE	94.0%	96.6%	99.8%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	100.0%
QUANTITY	99.5%	99.5%	99.4%	99.4%	99.4%	99.4%	99.4%	99.4%	99.4%	99.4%
FREQUENCY	93.4%	93.5%	93.6%	93.8%	93.9%	94.1%	93.9%	94.0%	94.1%	94.3%
INSTRUCTIONS	36.5%	37.6%	38.7%	39.6%	40.8%	41.4%	41.9%	42.4%	43.3%	43.1%
REPEAT_INTERVAL	54.5%	53.0%	49.9%	50.7%	51.2%	52.5%	53.2%	54.4%	55.4%	57.2%
REPEATS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
PBS_STATUS	98.4%	98.5%	98.7%	98.8%	98.9%	99.0%	99.1%	99.2%	99.2%	99.3%
RESTRICTION_CODE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
AUTHORITY_INDICATION	7.0%	7.0%	7.0%	7.5%	7.7%	7.6%	8.2%	8.4%	9.6%	12.2%
PREVIOUS_AUTHORITY	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REASON	35.7%	36.6%	36.9%	38.9%	38.9%	38.1%	38.3%	36.6%	35.3%	36.0%
REASON_CODE	58.0%	60.2%	63.0%	64.9%	65.8%	66.8%	67.5%	67.6%	67.8%	69.8%
IS_CURRENT[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RX_STATUS_LIMITED_MEDICATION	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RECORD_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Prescription issued										
SCRIPT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
MEDICINE_ACTIVE_INGREDIENT	99.0%	99.1%	98.9%	98.9%	98.8%	98.7%	98.6%	98.7%	99.0%	99.3%
MEDICINE_NAME	99.6%	99.8%	99.7%	99.7%	99.7%	99.7%	99.7%	99.7%	99.7%	99.8%
ATC_CODE[b]	94.0%	94.0%	93.7%	93.6%	93.3%	93.3%	93.2%	93.2%	93.3%	93.5%
DOSE	87.1%	86.9%	87.9%	88.5%	88.5%	88.4%	88.3%	88.2%	87.9%	87.7%
ROUTE	92.9%	95.4%	98.8%	99.1%	99.2%	99.4%	99.5%	99.7%	99.7%	99.9%

Selected variables	% completeness of selected variables for each financial year among all patients									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
STRENGTH	96.9%	97.5%	97.9%	98.0%	97.9%	97.8%	97.8%	97.6%	97.6%	97.8%
FREQUENCY	94.2%	94.4%	94.5%	94.7%	94.9%	95.0%	95.0%	95.1%	95.2%	95.4%
QUANTITY	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
INSTRUCTIONS	35.4%	35.5%	35.3%	35.5%	36.2%	36.7%	37.0%	37.0%	37.6%	37.1%
PRN	50.9%	51.7%	52.5%	53.0%	53.3%	54.1%	55.2%	56.1%	57.0%	58.9%
REGULATION_24	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REPEATS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REPEAT_INTERVAL	52.2%	53.0%	53.7%	54.2%	54.4%	55.1%	56.1%	56.9%	57.8%	59.6%
RESTRICTION_CODE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
AUTHORITY_INDICATION_CODE_AX	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
MBS billing										
ITEM_NUMBER	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
SERVICE_RECORD_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
SERVICE_PATIENT_COUNT	97.0%	97.3%	97.7%	98.0%	98.2%	98.3%	98.4%	98.5%	98.5%	98.6%
SERVICE_DATETIME[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CREATED_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
UPDATED_DATETIME	100.0%	100.0%	100.0%	99.9%	100.0%	99.9%	99.3%	98.7%	97.8%	96.2%
VISIT_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June FY; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] Date variable used in selection of relevant records for the study period of interest

[b] Derived by NPS MedicineWise

Completeness rates for most variables from the Encounter reason and Diagnosis tables were greater among regular attenders (Table 10) than infrequent attenders (Table 11). The recording rates for the assessed Encounter reason and Diagnosis variables were excellent (100%) except for 'ENCOUNTER_REASON_CODE' and 'DIAGNOSIS_REASON_CODE' which increased across the study period as discussed above (Figures 5 and 6).

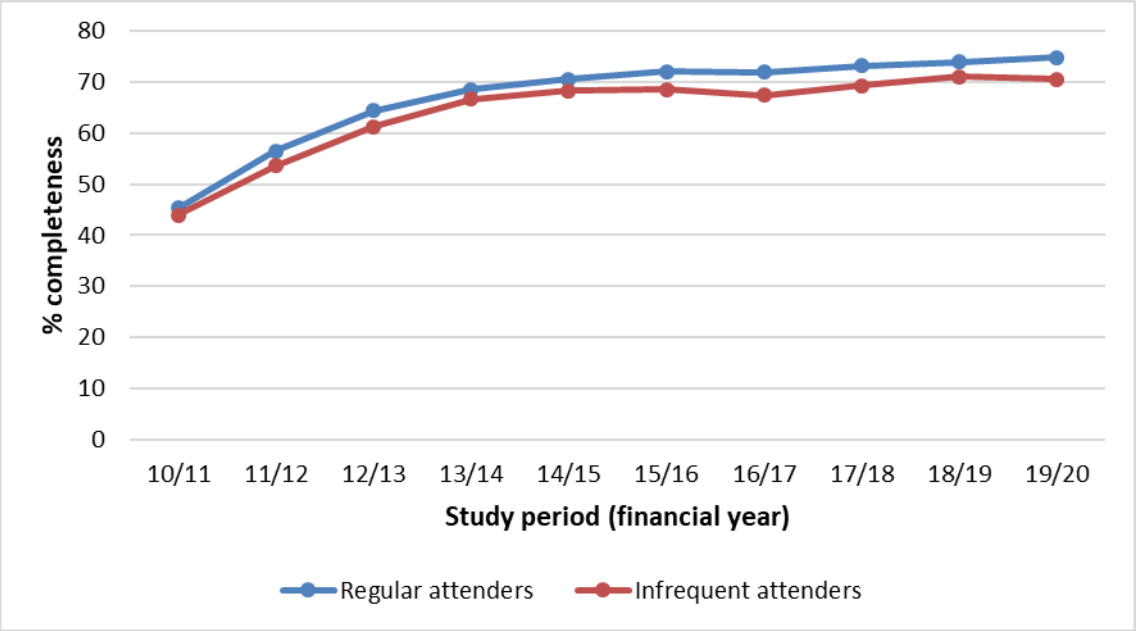

FIGURE 5: COMPLETENESS RATES FOR 'ENCOUNTER_REASON_CODE' IN REGULAR AND INFREQUENT ATTENDERS ACROSS THE 10-YEAR STUDY PERIOD (1 JULY 2010 TO 30 JUNE 2020)

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June FY.

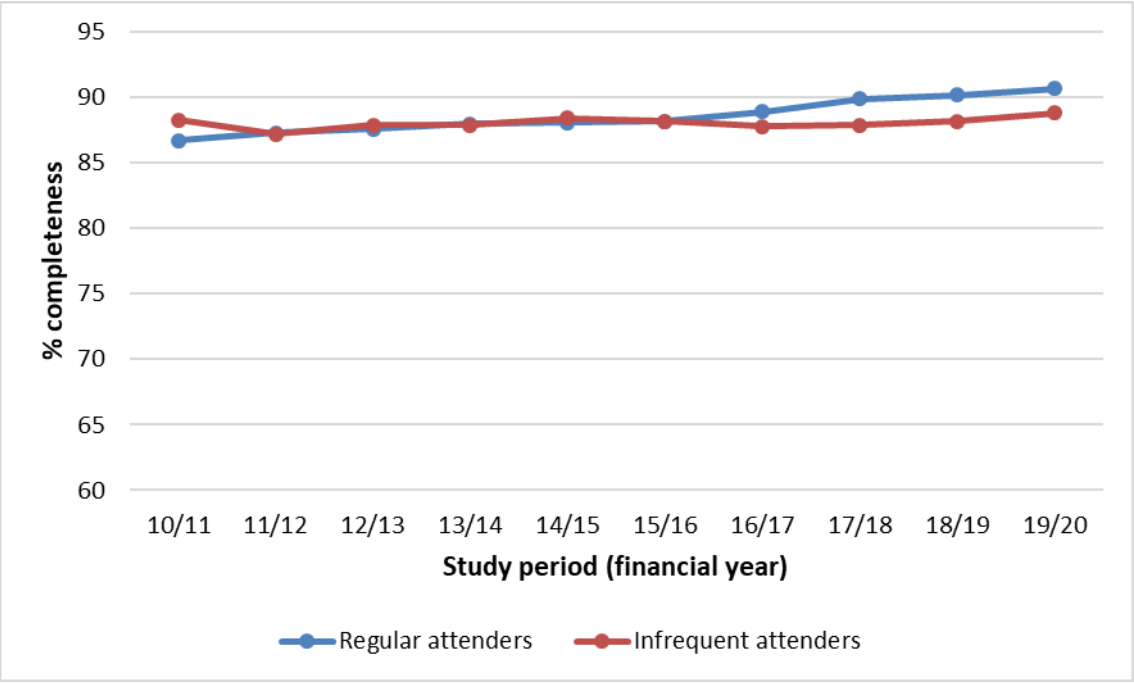

FIGURE 6: COMPLETENESS RATES FOR 'DIAGNOSIS_REASON_CODE' IN REGULAR AND INFREQUENT ATTENDERS ACROSS THE 10-YEAR STUDY PERIOD (1 JULY 2010 TO 30 JUNE 2020)

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June FY.

The completeness rates for most of the important variables in the three pathology tables – Requested investigations, Pathology results header and Pathology results detail – such as 'RESULT_NAME', 'RESULT_VALUE', 'LOINC_CODE', 'REQUESTED_TESTS' and 'UNITS' were high (> 95%) (Tables 9–11) with the exception of 'TEST_REASON' (Requested investigations table), which had relatively low recording rates for both regular and infrequent attenders across the study period (Figure 7). The 'NORMAL_FLAG' in the Pathology results header table, was 71–76% complete over the 10 years and the 'ABNORMAL_FLAG' in the Pathology results detail table was around 30% complete over the 10 years (Table 9). Based on these findings, it appears that the 'NORMAL_FLAG' is missing when the result is abnormal and conversely the 'ABNORMAL_FLAG' is missing when the result is normal.

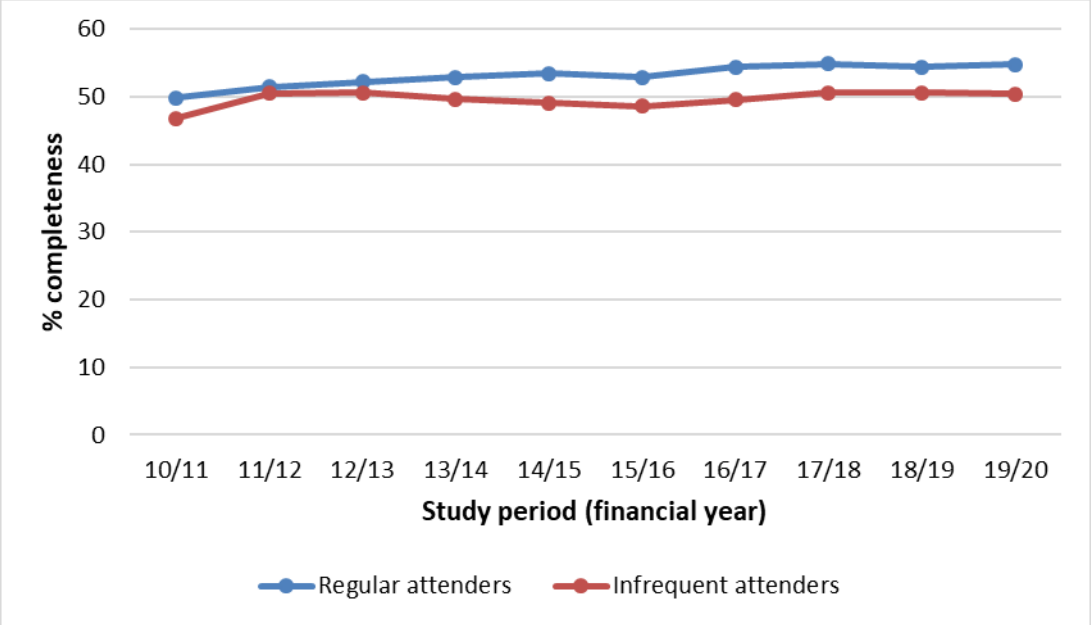

FIGURE 7: COMPLETENESS RATES FOR 'TEST_REASON' IN REGULAR AND INFREQUENT ATTENDERS ACROSS THE 10-YEAR STUDY PERIOD (1 JULY 2010 TO 30 JUNE 2020)

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June FY.

The recording rates for most of the prescription variables in the two tables – Medicine history and Prescription issued – were high (> 95%) except for variables containing similar information as that already recorded in other fields eg, 'PRODUCT_NAME' and 'INSTRUCTIONS', which had relatively low completeness rates (Tables 9–11). In the Medicine history table, the 'FIRST_DATE' was 100% complete whereas the 'LAST_DATE' was around 94% complete across the 10 years. This was not a surprise as the last date only populates when the prescription is issued (printed). The missing records probably relate to medicines that the GP is making a note of, but not issuing a prescription for, such as those prescribed by specialists or purchased over the counter.

Within the Prescription issued table completeness of 'MEDICINE_NAME' and 'MEDICINE_ACTIVE_INGREDIENT' was excellent across the 10-year period (> 98%) and an ATC code could be assigned to 93–94% of the records. 'DOSE' was available for around 87% of issued prescriptions each year and 'FREQUENCY' of administration hovered around 95% (Table 9). While completeness rates for 'INSTRUCTIONS' for use were low at around 35–37%, it is often possible to calculate the dosage from other variables, including dose, strength and frequency. The

AUTHORITY_INDICATION was complete for 13–19% of issued prescriptions (Table 9) and this most likely represents the proportion of issued scripts that are authority required prescriptions.

Poor completeness was observed for prescription 'REASON' and 'REASON_CODE', with recording rates for the latter variable increasing slightly across the study period (Figure 8). The proportion of issued prescriptions with a reason recorded remained around 36% across the 10-year study period. However, where a reason was recorded, there was some improvement in the use of coded entries (from a drop-down list) rather than free text, from 58.0% of the prescription reason records in 2010/11 to 68.9% in 2019/20.

Completeness rates for most variables in the MBS billing table were excellent (> 99%) for all the cohorts (Tables 9–11).

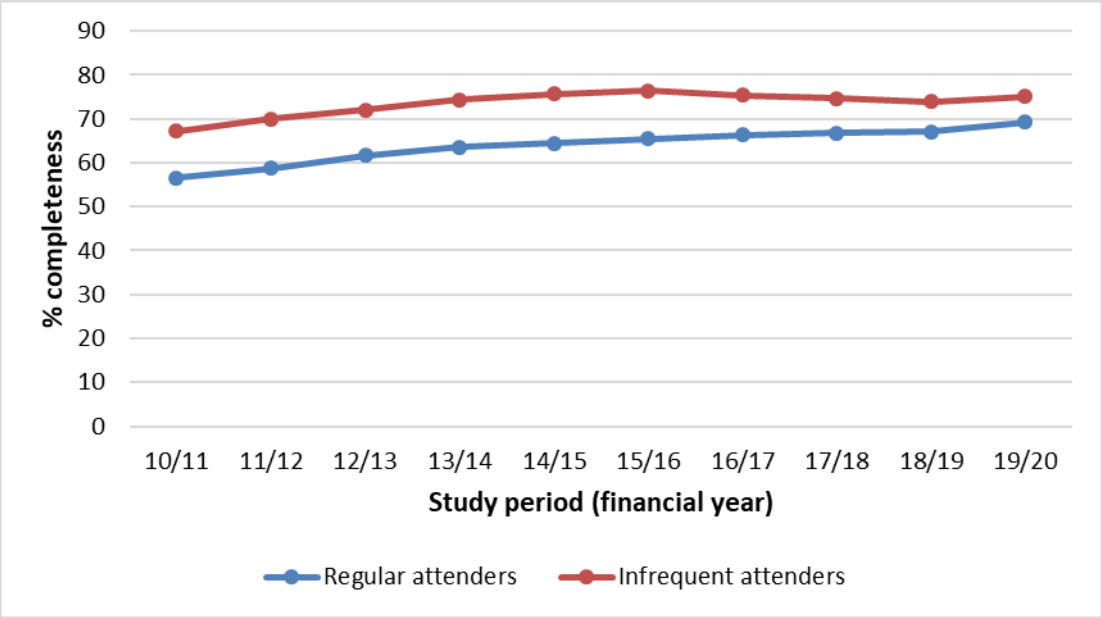

FIGURE 8: COMPLETENESS RATES FOR PRESCRIPTION 'REASON_CODE' IN REGULAR AND INFREQUENT ATTENDERS ACROSS THE 10-YEAR STUDY PERIOD (1 JULY 2010 TO 30 JUNE 2020)

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June FY.

TABLE 10: COMPLETENESS OF SELECTED VARIABLES ASSESSED EACH FINANCIAL YEAR FOR THE 'REGULAR ATTENDERS' COHORT (1 JULY 2010 TO 30 JUNE 2020)

Selected variables	% completeness of selected variables for each FY among regular attenders									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
Encounter										
PROVIDER_ID	99.4%	99.3%	99.4%	99.3%	99.4%	99.4%	99.4%	99.4%	99.5%	98.3%
VISIT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
VISIT_TYPE	28.3%	28.3%	28.7%	28.0%	27.6%	30.9%	41.8%	44.1%	44.8%	45.4%
VISIT_TYPE_BP	0.0%	0.0%	0.0%	0.0%	0.0%	5.2%	23.8%	27.3%	26.2%	24.9%
VISIT_TYPE_MD	58.7%	59.9%	62.5%	61.6%	60.8%	63.3%	66.3%	68.7%	73.4%	80.3%
ENCOUNTER_REASON	52.6%	54.0%	56.7%	60.5%	63.3%	65.6%	67.9%	68.7%	70.8%	71.7%
IS_CLINICAL[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
PROVIDER_IS_DR_OR_NURSE[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Encounter reason										
VISIT_DATETIME[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
ENCOUNTER_REASON_CODE	45.3%	56.5%	64.4%	68.6%	70.5%	72.0%	71.9%	73.2%	73.9%	74.8%
ENCOUNTER_REASON	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Diagnosis										
ADMIN_FLAG[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_REASON	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_STATUS_ACTIVE_FLAG[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_TYPE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CREATED_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CIS_CODED_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_REASON_CODE	86.7%	87.3%	87.6%	88.0%	88.1%	88.2%	88.9%	89.9%	90.2%	90.7%
DIFFERENTIAL_FLAG	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Pathology results header										
RESULT_NAME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
COLLECTION_DATE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REPORT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Selected variables	% completeness of selected variables for each FY among regular attenders									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
IMPORT_DATE	93.9%	96.1%	99.6%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
COMPLETION_FLAG	97.6%	97.9%	98.2%	98.1%	97.8%	98.2%	98.3%	98.4%	98.4%	98.6%
NORMAL_FLAG	70.5%	70.8%	71.2%	71.4%	70.8%	72.1%	73.4%	74.2%	74.7%	76.0%
Pathology results detail										
PATHOLOGY_RESULT_ID	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RESULT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DATA_TYPE	96.4%	97.8%	98.4%	99.1%	99.4%	99.6%	99.6%	99.8%	99.9%	100.0%
LOINC_CODE	97.6%	97.8%	97.7%	97.8%	97.7%	97.7%	97.3%	97.1%	97.4%	97.3%
RESULT_NAME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RESULT_VALUE	100.0%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%	99.7%	99.2%
UNITS	94.1%	93.6%	93.3%	92.7%	92.8%	92.3%	92.7%	94.4%	94.7%	94.8%
NORMAL_RANGE	89.1%	90.2%	90.9%	90.5%	90.1%	89.3%	89.9%	92.0%	92.6%	92.2%
ABNORMAL_FLAG	29.5%	31.5%	31.8%	31.9%	31.4%	30.4%	30.7%	30.5%	29.5%	28.3%
RECORD_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CREATED_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
UPDATED_DATETIME	63.4%	55.7%	49.3%	47.5%	46.7%	44.8%	43.2%	41.6%	40.2%	38.2%
Requested investigations										
REQUESTED_TESTS	99.4%	99.5%	99.6%	99.8%	99.8%	99.8%	99.8%	99.8%	99.8%	99.7%
TEST_REASON	49.8%	51.4%	52.2%	52.9%	53.4%	52.9%	54.4%	54.9%	54.4%	54.8%
REQUEST_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
BILLING	27.4%	28.5%	28.5%	29.6%	29.7%	30.0%	31.3%	32.2%	32.7%	33.8%
COPIES	0.4%	0.5%	0.6%	0.7%	0.8%	0.8%	0.8%	0.9%	0.9%	1.0%
Medicine history										
CREATED_BY	94.1%	96.5%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
FIRST_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
LAST_DATE	94.9%	94.8%	94.9%	95.0%	95.1%	94.9%	94.7%	94.6%	94.5%	94.0%
MEDICINE_ACTIVE_INGREDIENT	99.1%	99.1%	99.0%	99.1%	99.1%	99.1%	99.1%	99.1%	99.1%	99.2%
MEDICINE_NAME	99.6%	99.5%	99.4%	99.5%	99.5%	99.5%	99.5%	99.5%	99.5%	99.4%
PRODUCT_NAME	52.7%	51.5%	50.3%	49.8%	49.4%	48.0%	46.9%	45.5%	44.4%	42.3%

Selected variables	% completeness of selected variables for each FY among regular attenders									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
ATC_CODE[b]	93.6%	93.5%	93.3%	93.2%	93.0%	93.0%	92.9%	93.0%	93.0%	93.1%
DOSE	89.1%	88.0%	87.9%	88.0%	87.6%	87.1%	86.9%	86.3%	85.8%	85.4%
STRENGTH	96.4%	96.4%	96.2%	96.0%	95.9%	95.6%	95.7%	95.4%	95.6%	95.8%
FORM	73.9%	73.7%	73.3%	74.1%	74.2%	73.9%	74.3%	74.5%	74.1%	73.3%
ROUTE	94.1%	96.6%	99.8%	99.9%	99.8%	99.8%	99.9%	99.9%	99.9%	100.0%
QUANTITY	99.5%	99.5%	99.5%	99.5%	99.5%	99.5%	99.4%	99.4%	99.5%	99.4%
FREQUENCY	93.2%	93.3%	93.4%	93.6%	93.8%	93.9%	93.8%	93.9%	94.0%	94.3%
INSTRUCTIONS	36.7%	37.9%	38.9%	40.0%	41.3%	41.8%	42.2%	42.6%	43.4%	43.1%
REPEAT_INTERVAL	53.4%	52.0%	49.1%	49.7%	50.0%	51.4%	52.6%	54.0%	55.1%	57.1%
REPEATS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
PBS_STATUS	98.3%	98.4%	98.6%	98.7%	98.8%	98.9%	99.1%	99.1%	99.2%	99.2%
RESTRICTION_CODE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
AUTHORITY_INDICATION	7.2%	7.2%	7.2%	7.8%	8.0%	7.9%	8.5%	8.7%	9.8%	12.5%
PREVIOUS_AUTHORITY	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REASON	35.1%	36.0%	36.4%	38.6%	38.7%	38.0%	38.1%	36.2%	35.0%	35.8%
REASON_CODE	56.6%	58.8%	61.7%	63.5%	64.4%	65.5%	66.4%	66.7%	67.1%	69.3%
IS_CURRENT[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RX_STATUS_LIMITED_MEDICATION	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RECORD_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Prescription issued										
SCRIPT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
MEDICINE_ACTIVE_INGREDIENT	99.0%	99.1%	99.0%	99.0%	98.8%	98.8%	98.7%	98.8%	99.0%	99.3%
MEDICINE_NAME	99.6%	99.8%	99.7%	99.7%	99.7%	99.7%	99.7%	99.7%	99.7%	99.8%
ATC_CODE[b]	94.1%	94.2%	93.9%	93.8%	93.5%	93.5%	93.4%	93.4%	93.4%	93.6%
DOSE	87.4%	87.3%	88.1%	88.7%	88.7%	88.6%	88.5%	88.4%	88.1%	87.9%
ROUTE	92.8%	95.3%	98.8%	99.0%	99.2%	99.3%	99.5%	99.7%	99.7%	99.9%
STRENGTH	97.0%	97.6%	98.0%	98.1%	98.0%	97.9%	97.9%	97.8%	97.8%	97.9%
FREQUENCY	94.0%	94.2%	94.4%	94.6%	94.8%	94.9%	95.0%	95.1%	95.2%	95.4%
QUANTITY	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Selected variables	% completeness of selected variables for each FY among regular attenders									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
INSTRUCTIONS	35.3%	35.4%	35.1%	35.5%	36.3%	36.7%	36.8%	36.8%	37.2%	36.8%
PRN	49.8%	50.6%	51.6%	52.2%	52.4%	53.3%	54.7%	55.7%	56.9%	58.9%
REGULATION_24	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REPEATS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REPEAT_INTERVAL	51.1%	51.8%	52.9%	53.4%	53.5%	54.4%	55.7%	56.6%	57.7%	59.6%
RESTRICTION_CODE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
AUTHORITY_INDICATION_CODE_AX	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
MBS billing										
ITEM_NUMBER	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
SERVICE_RECORD_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
SERVICE_PATIENT_COUNT	97.3%	97.6%	97.9%	98.2%	98.3%	98.4%	98.6%	98.7%	98.7%	98.7%
SERVICE_DATETIME[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CREATED_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
UPDATED_DATETIME	100.0%	100.0%	100.0%	99.9%	100.0%	99.9%	99.4%	98.7%	97.8%	96.2%
VISIT_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June FY; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] Date variable used in selection of relevant records for the study period of interest

[b] Derived by NPS MedicineWise

TABLE 11: COMPLETENESS OF SELECTED VARIABLES ASSESSED EACH FINANCIAL YEAR FOR THE 'INFREQUENT ATTENDERS' COHORT (1 JULY 2010 TO 30 JUNE 2020)

Selected variables	% completeness of selected variables for each FY among infrequent attenders									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
Encounter										
PROVIDER_ID	98.2%	97.5%	97.9%	97.8%	98.0%	97.6%	97.6%	96.5%	97.6%	97.4%
VISIT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
VISIT_TYPE	31.4%	30.0%	29.2%	28.7%	30.0%	32.3%	40.6%	43.8%	45.5%	45.8%
VISIT_TYPE_BP	0.0%	0.0%	0.0%	0.0%	0.0%	3.5%	16.7%	21.5%	24.0%	23.3%
VISIT_TYPE_MD	87.3%	85.5%	83.1%	84.6%	90.2%	89.9%	85.3%	84.0%	80.8%	86.2%
ENCOUNTER_REASON	56.1%	57.2%	60.5%	63.5%	66.7%	69.2%	71.5%	72.3%	74.4%	75.1%
IS_CLINICAL[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
PROVIDER_IS_DR_OR_NURSE[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Encounter reason										
VISIT_DATETIME[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
ENCOUNTER_REASON_CODE	44.0%	53.6%	61.3%	66.6%	68.3%	68.5%	67.4%	69.3%	71.0%	70.5%
ENCOUNTER_REASON	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Diagnosis										
ADMIN_FLAG[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_REASON	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_STATUS_ACTIVE_FLAG[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_TYPE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CREATED_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CIS_CODED_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DIAGNOSIS_REASON_CODE	88.3%	87.2%	87.9%	87.9%	88.4%	88.2%	87.8%	87.9%	88.2%	88.8%
DIFFERENTIAL_FLAG	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Pathology results header										
RESULT_NAME	100.0%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
COLLECTION_DATE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%
REPORT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
IMPORT_DATE	95.3%	97.6%	99.7%	100.0%	99.9%	99.9%	99.9%	99.9%	100.0%	100.0%

Selected variables	% completeness of selected variables for each FY among infrequent attenders									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
COMPLETION_FLAG	98.5%	98.3%	98.9%	98.8%	98.7%	98.9%	98.9%	98.8%	98.6%	98.8%
NORMAL_FLAG	80.1%	80.3%	79.5%	80.1%	80.4%	80.5%	78.7%	77.5%	75.6%	75.4%
Pathology results detail										
PATHOLOGY_RESULT_ID	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RESULT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DATA_TYPE	95.1%	96.7%	98.1%	98.8%	99.1%	99.2%	99.4%	99.7%	99.9%	100.0%
LOINC_CODE	96.5%	97.0%	97.0%	97.1%	96.8%	96.9%	96.5%	96.8%	97.2%	97.2%
RESULT_NAME	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RESULT_VALUE	99.9%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%	99.7%	99.3%
UNITS	93.3%	92.8%	92.9%	92.3%	91.9%	91.8%	92.3%	94.0%	94.5%	94.3%
NORMAL_RANGE	87.6%	88.9%	90.4%	90.1%	89.5%	89.0%	89.9%	91.8%	92.4%	91.7%
ABNORMAL_FLAG	29.8%	32.3%	30.9%	32.4%	30.3%	29.3%	27.5%	27.0%	26.7%	26.5%
RECORD_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CREATED_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
UPDATED_DATETIME	42.6%	37.2%	35.6%	31.3%	30.1%	30.0%	32.2%	33.7%	35.5%	35.9%
Requested investigations										
REQUESTED_TESTS	99.6%	99.6%	99.7%	99.7%	99.7%	99.7%	99.7%	99.8%	99.8%	99.7%
TEST_REASON	46.8%	50.5%	50.6%	49.6%	49.1%	48.6%	49.5%	50.6%	50.6%	50.4%
REQUEST_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
BILLING	38.0%	40.1%	39.4%	40.5%	39.0%	38.7%	37.9%	35.9%	33.3%	31.8%
COPIES	0.3%	0.4%	0.6%	0.6%	0.6%	0.6%	0.6%	0.6%	0.6%	0.6%
Medicine history										
CREATED_BY	93.3%	96.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
FIRST_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
LAST_DATE	91.6%	91.4%	91.5%	91.2%	90.8%	90.2%	89.9%	88.9%	88.7%	87.3%
MEDICINE_ACTIVE_INGREDIENT	99.5%	99.4%	99.4%	99.4%	99.5%	99.4%	99.4%	99.4%	99.4%	99.3%
MEDICINE_NAME	99.7%	99.6%	99.6%	99.6%	99.6%	99.6%	99.6%	99.6%	99.6%	99.5%
PRODUCT_NAME	45.1%	43.8%	44.4%	42.1%	40.2%	39.3%	41.1%	41.3%	41.6%	40.9%
ATC_CODE[b]	94.5%	94.2%	94.1%	93.9%	93.8%	93.8%	93.6%	93.6%	93.6%	93.5%

Selected variables	% completeness of selected variables for each FY among infrequent attenders									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
DOSE	89.0%	87.6%	87.8%	87.4%	86.8%	86.6%	86.3%	85.6%	85.3%	84.8%
STRENGTH	96.3%	96.1%	95.7%	95.6%	95.5%	95.1%	95.0%	94.7%	94.8%	94.8%
FORM	72.6%	72.8%	74.5%	74.4%	72.8%	73.1%	74.8%	74.5%	73.7%	73.7%
ROUTE	93.2%	96.7%	99.8%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	100.0%
QUANTITY	99.1%	99.0%	98.7%	98.9%	98.8%	98.9%	98.8%	98.8%	98.8%	98.8%
FREQUENCY	95.1%	95.2%	94.9%	95.2%	95.2%	95.4%	94.6%	94.7%	94.5%	94.3%
INSTRUCTIONS	34.9%	35.5%	36.9%	37.0%	37.2%	38.1%	39.9%	40.6%	42.2%	42.3%
REPEAT_INTERVAL	61.9%	59.1%	55.2%	57.5%	59.5%	60.3%	58.5%	58.4%	58.0%	58.5%
REPEATS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
PBS_STATUS	99.3%	99.3%	99.3%	99.5%	99.5%	99.6%	99.6%	99.7%	99.6%	99.7%
RESTRICTION_CODE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
AUTHORITY_INDICATION	5.3%	5.6%	5.5%	5.5%	5.6%	5.6%	5.9%	6.4%	7.5%	9.1%
PREVIOUS_AUTHORITY	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REASON	39.3%	40.6%	40.6%	41.2%	40.1%	39.3%	40.1%	39.0%	37.9%	38.8%
REASON_CODE	67.2%	70.0%	72.0%	74.4%	75.7%	76.4%	75.4%	74.6%	73.9%	75.1%
IS_CURRENT[b]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RX_STATUS_LIMITED_MEDICATION	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RECORD_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Prescription issued										
SCRIPT_DATE[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
MEDICINE_ACTIVE_INGREDIENT	98.9%	98.5%	98.0%	98.6%	98.4%	98.1%	98.1%	97.3%	98.8%	99.0%
MEDICINE_NAME	99.8%	99.7%	99.8%	99.8%	99.8%	99.7%	99.7%	99.7%	99.7%	99.7%
ATC_CODE[b]	92.4%	91.8%	91.0%	91.3%	91.0%	90.8%	90.7%	89.5%	90.6%	90.7%
DOSE	83.6%	82.5%	85.0%	86.2%	86.5%	86.3%	85.9%	85.3%	84.8%	84.4%
ROUTE	93.9%	96.4%	99.4%	99.6%	99.6%	99.6%	99.7%	99.9%	99.9%	100.0%
STRENGTH	95.4%	96.0%	96.7%	96.7%	96.5%	96.1%	96.0%	95.5%	95.3%	95.4%
FREQUENCY	96.1%	96.2%	95.9%	95.9%	95.9%	96.0%	95.1%	95.0%	94.5%	94.3%
QUANTITY	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
INSTRUCTIONS	36.3%	36.9%	36.6%	35.7%	35.4%	36.7%	39.3%	41.3%	43.5%	43.8%

Selected variables	% completeness of selected variables for each FY among infrequent attenders									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
PRN	64.3%	64.9%	63.4%	64.1%	65.3%	65.2%	62.2%	61.0%	59.0%	58.8%
REGULATION_24	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REPEATS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
REPEAT_INTERVAL	65.8%	66.6%	64.6%	65.1%	66.1%	66.0%	63.1%	61.8%	59.7%	59.5%
RESTRICTION_CODE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
AUTHORITY_INDICATION_CODE_AX	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
MBS billing										
ITEM_NUMBER	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
SERVICE_RECORD_STATUS	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
SERVICE_PATIENT_COUNT	93.2%	94.5%	95.5%	96.2%	96.5%	96.4%	96.4%	96.2%	95.9%	95.9%
SERVICE_DATETIME[a]	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CREATED_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
UPDATED_DATETIME	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	99.0%	98.3%	97.6%	96.0%
VISIT_DATETIME	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June FY; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] Date variable used in selection of relevant records for the study period of interest

[b] Derived by NPS MedicineWise

4.3. Completeness of risk factors and measurements/tests

Recording rates of selected risk factors and measurements were assessed among the 1 July 2018 to 30 June 2020 patient cohort, comprising approximately 3.2 million 'all patients', of whom 1.9 million were regular attenders and 1.2 million infrequent attenders (see Table 6 and Figure 1 for further details on patient cohorts).

Body mass Index

BMI or height and weight (as defined in Table A2) was more frequently recorded for regular attenders than infrequent attenders (38.4% vs 12.6%) in MedicineInsight in the 24-month period from 1 July 2018 to 30 June 2020 (Table 12). Overall BMI recording rates in regular attenders were similar for males (38.5%) and females (38.4%), however, both men and women aged 70–89 years and children under 10 years had greater rates of BMI completeness than other age groups (Figure 9). BMI is selectively recorded; among adult regular patients recording of BMI was substantially higher overall (38.4%) for those patients with one of the selected chronic conditions, ranging from 48.2% in patients with cancer to 68.5% in patients with type 2 or unspecified diabetes (Table 12). This pattern was not observed for adult infrequent attenders, as BMI completeness rates for the selected chronic conditions did not vary substantially from the overall rate (12.6%). These findings highlight that among regular patients, BMI is selectively recorded, and excluding patients with missing data on BMI (who are generally younger and have less chronic conditions) from analyses would introduce significant selection bias. BMI is less likely to be recorded for patients with normal weight.[15]

According to the RACGP Redbook clinical guidelines for preventive activities in general practice,[16] BMI should be measured in adults every 2 years, and in children at times of child health surveillance or immunisation. Adults at increased risk (for example, with a history of CVD) and Aboriginal and Torres Strait Islander people should be assessed every 12 months, and adults with identified risk (those who are overweight and obese) should be assessed every 6 months. Our findings suggest that recording rates for BMI or height and weight are relatively low, even in regular attenders. It is possible that some GPs may record information on BMI or height and weight in different places within the CIS, for example in the progress notes (which are not available to MedicineInsight), and this can affect completeness rates in MedicineInsight data.

TABLE 12: PATIENTS IN FY10 WITH BMI OR HEIGHT AND WEIGHT RECORDED DURING 1 JULY 2018 TO 30 JUNE 2020, BY PATIENT CHARACTERISTICS

	% of patients in FY10 with at least one record of either BMI or height and weight		
	All Patients (N = 3,161,134)	Regular attenders (N = 1,949,196)	Infrequent attenders (N = 1,211,938)
Either BMI or height and weight available	28.5%	38.4%	12.6%
Sex			
Male	28.4%	38.5%	13.9%
Female	28.6%	38.4%	11.3%
Intersex or indeterminate	26.3%	39.0%	14.2%
Age group (years)			

	% of patients in FY10 with at least one record of either BMI or height and weight		
	All Patients (N = 3,161,134)	Regular attenders (N = 1,949,196)	Infrequent attenders (N = 1,211,938)
0–9	30.5%	42.1%	14.5%
10–19	19.1%	26.1%	10.9%
20–29	21.6%	29.9%	13.0%
30–39	24.2%	32.5%	13.4%
40–49	27.8%	36.7%	13.4%
50–59	30.6%	39.6%	12.9%
60–69	34.0%	43.2%	10.7%
70–79	42.4%	51.8%	8.5%
80–89	47.4%	56.2%	6.1%
90+	31.7%	38.3%	3.3%
Males stratified by age group (years)			
0–9	30.8%	42.7%	14.4%
10–19	18.1%	24.9%	10.9%
20–29	18.6%	24.2%	14.1%
30–39	22.6%	29.8%	15.3%
40–49	28.6%	37.9%	15.8%
50–59	32.2%	41.2%	15.5%
60–69	35.5%	44.3%	12.5%
70–79	42.7%	51.9%	9.3%
80–89	47.9%	56.7%	6.8%
90+	35.0%	42.2%	3.9%
Females stratified by age group (years)			
0–9	30.1%	41.5%	14.6%
10–19	20.0%	27.2%	11.0%
20–29	23.8%	33.1%	12.1%
30–39	25.4%	34.3%	11.5%
40–49	27.2%	35.8%	11.0%
50–59	29.3%	38.3%	10.6%
60–69	32.8%	42.2%	9.2%
70–79	42.1%	51.7%	7.9%
80–89	46.9%	55.7%	5.6%
90+	30.0%	36.2%	3.0%
Smoking status (all ages)			
Current smoker	30.9%	38.7%	16.5%
Ex-smoker	42.3%	49.1%	17.3%
Non-smoker	31.4%	39.2%	15.6%
Selected chronic conditions in adults (≥ 18 years)			
Cancer	43.9%	48.2%	10.0%
Cardiovascular disease	52.9%	57.2%	10.8%
Chronic kidney disease	60.8%	64.3%	10.2%
Chronic obstructive pulmonary disease	54.8%	59.3%	10.3%
Depression and/or anxiety	38.3%	43.8%	11.5%
Diabetes type 2/unspecified	61.7%	68.5%	14.5%
Dyslipidaemia	50.9%	55.5%	14.5%
Hypertension	49.3%	54.5%	13.7%
Osteoarthritis	52.9%	56.6%	11.2%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

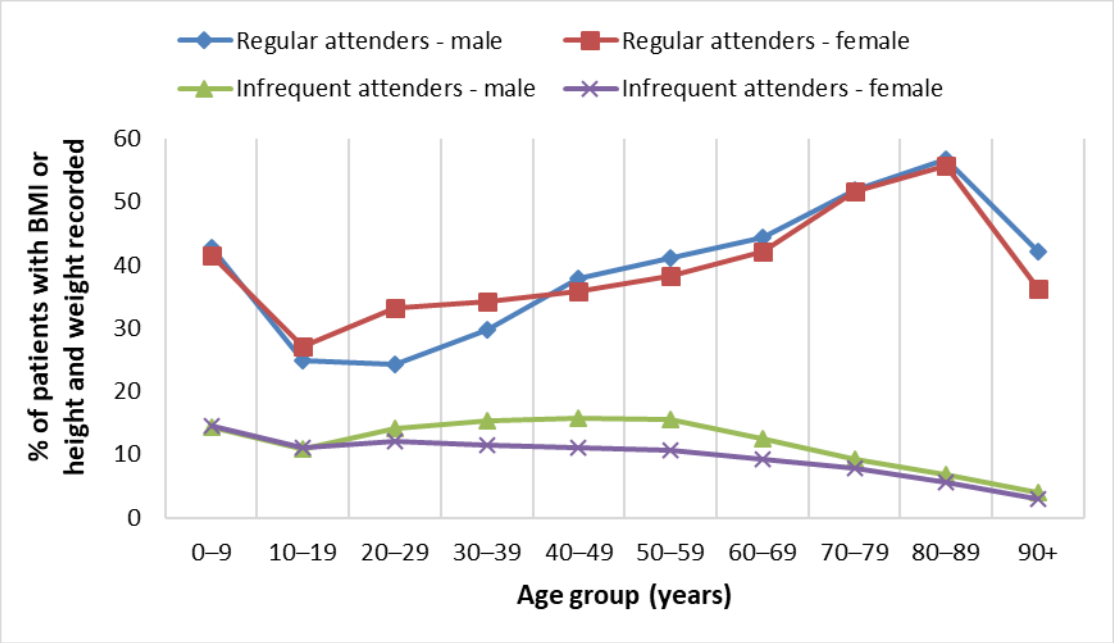

FIGURE 9: COMPLETENESS RATES OF BMI OR HEIGHT AND WEIGHT RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS BY PATIENT AGE AND SEX, 2018–2020

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

Smoking status

Data for smoking status are not yet available longitudinally in MedicineInsight and completeness rates are based on whether it was recorded in the October 2020 data download. Table 13 shows recording rates for smoking status in patients aged at least 18 years in FY10 at the data download date. Smoking status was recorded for the majority (79.9%) of patients aged 18 years or older in MedicineInsight. Recording rates of smoking status were greater in regular attenders at 88.8%, compared with infrequent attenders at 64.7%. Among regular attenders, the recording rates were similar between males (88.4%) and females (89.1%) and generally increased with age (Figure 10). Completeness rates for recording smoking status were greater in patients who had the selected chronic conditions (> 93%). Our findings indicate that most MedicineInsight patients who were regular attenders had smoking status recorded.

TABLE 13: PATIENTS AGED ≥ 18 YEARS IN FY10 WITH SMOKING STATUS RECORDED AS AT 31 OCTOBER 2020, BY PATIENT CHARACTERISTICS

	% of patients aged ≥ 18 years in FY10 with smoking status recorded		
	All Patients (N = 2,476,035)	Regular attenders (N = 1,564,247)	Infrequent attenders (N = 911,788)
Smoking status available for patients aged ≥ 18 years	79.9%	88.8%	64.7%
Sex			
Male	78.9%	88.4%	64.7%
Female	80.7%	89.1%	64.7%
Intersex or indeterminate	64.0%	76.1%	51.9%

	% of patients aged ≥ 18 years in FY10 with smoking status recorded		
	All Patients (N = 2,476,035)	Regular attenders (N = 1,564,247)	Infrequent attenders (N = 911,788)
Age group (years)			
18–19	66.9%	75.0%	57.2%
20–29	74.5%	84.4%	64.1%
30–39	78.9%	88.0%	67.1%
40–49	81.9%	90.0%	68.6%
50–59	83.3%	91.2%	67.4%
60–69	82.8%	91.1%	61.5%
70–79	83.3%	91.0%	55.9%
80–89	82.8%	89.5%	51.3%
90+	76.2%	82.7%	48.5%
Males stratified by age group (years)			
18–19	63.1%	70.9%	55.6%
20–29	71.1%	81.6%	62.7%
30–39	77.0%	86.8%	67.2%
40–49	81.0%	89.6%	69.1%
50–59	83.4%	91.5%	68.3%
60–69	83.2%	91.2%	62.3%
70–79	83.4%	90.9%	56.3%
80–89	83.2%	89.9%	52.1%
90+	77.9%	84.6%	49.1%
Females stratified by age group (years)			
18–19	69.9%	77.7%	58.7%
20–29	76.9%	86.1%	65.4%
30–39	80.3%	88.7%	67.1%
40–49	82.6%	90.3%	68.2%
50–59	83.2%	91.1%	66.7%
60–69	82.4%	91.0%	60.8%
70–79	83.3%	91.0%	55.5%
80–89	82.4%	89.2%	50.6%
90+	75.4%	81.6%	48.2%
Selected chronic conditions (≥ 18 years)			
Asthma	91.7%	93.9%	82.9%
Cancer	91.8%	93.3%	79.9%
Cardiovascular disease	93.5%	94.7%	81.4%
Chronic obstructive pulmonary disease	95.8%	96.8%	86.6%
Depression and/or anxiety	91.5%	93.2%	83.4%
Diabetes type 2/unspecified	92.1%	94.2%	77.3%
Dyslipidaemia	94.5%	95.5%	86.4%
Hypertension	92.8%	94.5%	81.3%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

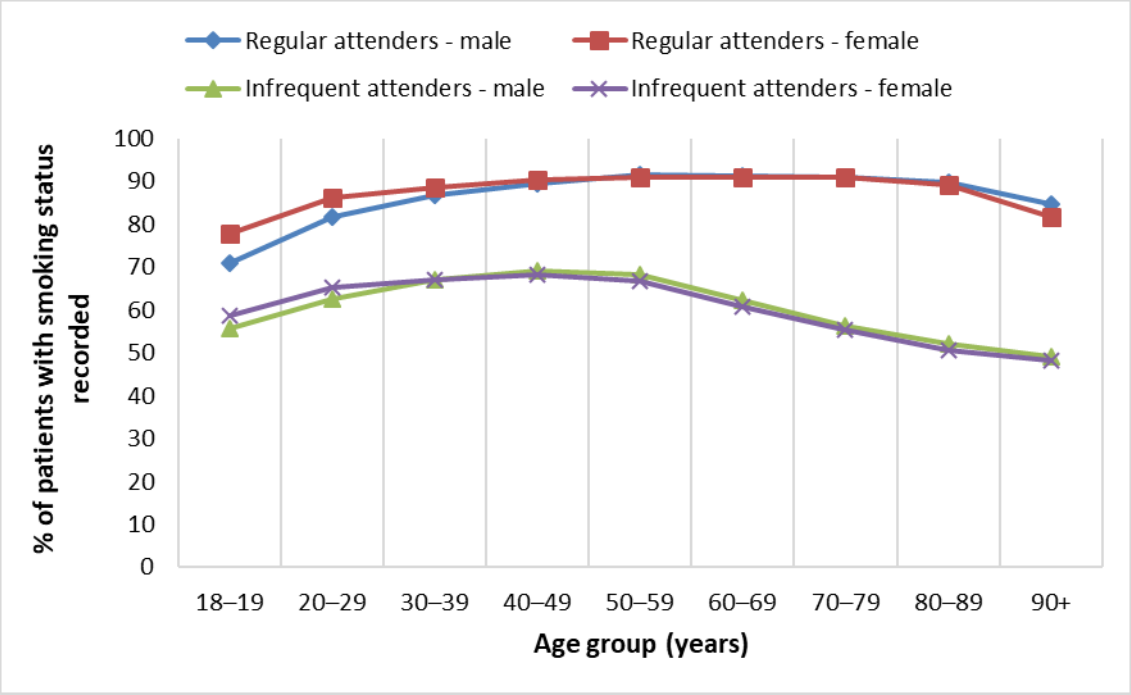

FIGURE 10: COMPLETENESS RATES OF SMOKING STATUS RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS BY PATIENT AGE AND SEX, 2020

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

Alcohol use status

Like smoking, alcohol use is not recorded longitudinally in MedicineInsight, and completeness rates are based on whether it was recorded in the October 2020 data download. Completeness rates of alcohol use in all patients aged 18 years or older were relatively low (49.8%). Recording rates of alcohol use were greater in regular attenders than infrequent attenders overall (54.7% vs 41.6%) and for all the patient characteristics assessed (Table 14 and Figure 11). Among regular attenders, recording rates of alcohol use were slightly greater in women (55.5%) than men (53.5%) and increased with age up to the 70–79 years age group for both genders. This may reflect a higher likelihood of recording risk factors for patients with comorbidities such as CVD, who are also more likely to be older. It may also partly reflect opportunity, as older people tend to visit their GPs more frequently than people in younger age groups. Completeness rates for recording alcohol use were greater in regular patients who had the selected chronic conditions, ranging from 55.4% in patients with asthma to 62.5% in those with cancer.

TABLE 14: PATIENTS AGED ≥ 18 YEARS IN FY10 WITH ALCOHOL STATUS RECORDED AS AT 31 OCTOBER 2020, BY PATIENT CHARACTERISTICS

	% of patients aged ≥ 18 years in FY10 with alcohol status recorded		
	All Patients (N = 2,476,035)	Regular attenders (N = 1,564,247)	Infrequent attenders (N = 911,788)
Current alcohol status available for patients aged ≥ 18 years	49.8%	54.7%	41.6%
Sex			
Male	48.4%	53.5%	40.8%
Female	51.0%	55.5%	42.3%

	% of patients aged ≥ 18 years in FY10 with alcohol status recorded		
	All Patients (N = 2,476,035)	Regular attenders (N = 1,564,247)	Infrequent attenders (N = 911,788)
Intersex or indeterminate	54.5%	56.0%	53.0%
Age group (years)			
18–19	45.9%	50.8%	40.0%
20–29	46.7%	52.0%	41.3%
30–39	48.8%	53.9%	42.1%
40–49	50.5%	54.9%	43.4%
50–59	50.7%	54.8%	42.7%
60–69	51.1%	55.4%	39.9%
70–79	53.6%	57.8%	38.7%
80–89	53.4%	57.0%	36.7%
90+	50.5%	53.7%	36.7%
Males stratified by age group (years)			
18–19	43.4%	48.4%	38.6%
20–29	43.8%	49.0%	39.6%
30–39	46.5%	51.7%	41.4%
40–49	49.1%	53.7%	42.8%
50–59	49.7%	53.8%	42.1%
60–69	50.5%	54.6%	39.6%
70–79	53.2%	57.3%	38.2%
80–89	53.6%	57.1%	37.1%
90+	50.4%	53.6%	36.9%
Females stratified by age group (years)			
18–19	47.8%	52.3%	41.4%
20–29	48.8%	53.7%	42.7%
30–39	50.5%	55.3%	42.8%
40–49	51.7%	55.8%	44.1%
50–59	51.5%	55.5%	43.3%
60–69	51.6%	56.1%	40.2%
70–79	54.0%	58.2%	39.2%
80–89	53.3%	56.8%	36.4%
90+	50.5%	53.7%	36.7%
Selected chronic conditions (≥ 18 years)			
Asthma	54.9%	55.4%	52.9%
Cancer	62.2%	62.5%	59.6%
Cardiovascular disease	58.3%	58.7%	54.8%
Chronic obstructive pulmonary disease	55.8%	56.1%	52.9%
Depression and/or anxiety	57.8%	58.4%	55.2%
Diabetes type 2/unspecified	55.1%	55.7%	50.4%
Dyslipidaemia	57.8%	58.1%	54.9%
Hypertension	57.2%	57.8%	53.2%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

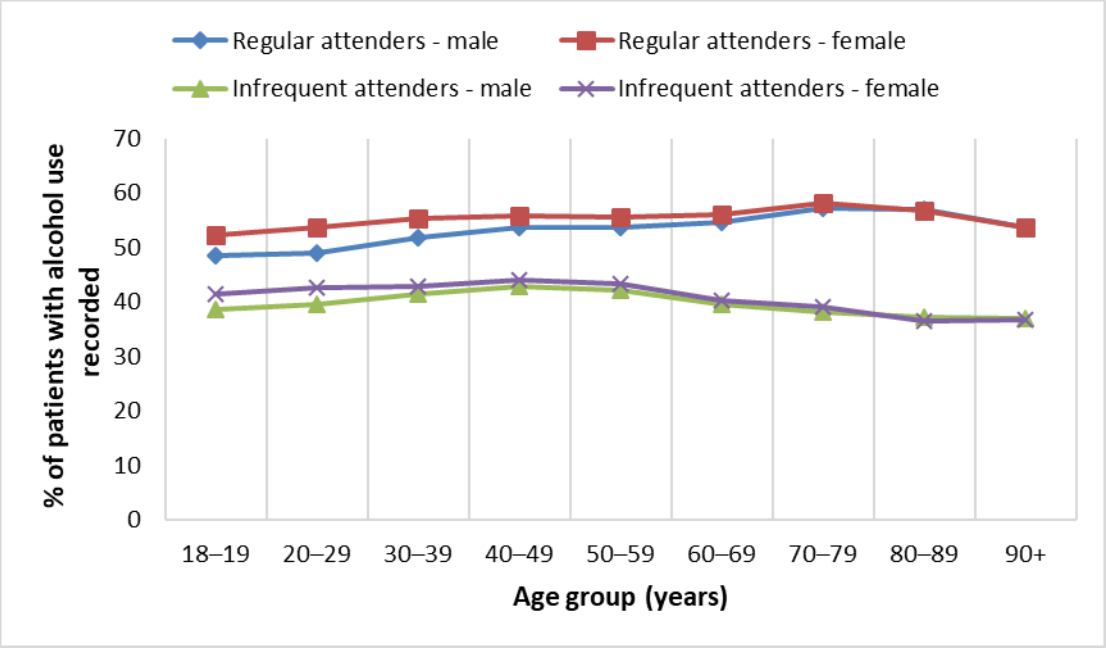

FIGURE 11: COMPLETENESS RATES OF ALCOHOL USE RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS BY PATIENT AGE AND SEX, 2020

Regular attenders are patients who had at least 3 clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

Blood Pressure

In the 24-month period from 1 July 2018 to 30 June 2020, blood pressure was recorded for 79.7% of regular attenders aged ≥ 40 years, 2.5 times that for infrequent attenders (31.6%) (Table 15). Overall completeness rates for blood pressure in regular attenders were similar for males (80.5%) and females (79.1%) and increased with age for both genders peaking in the 70–79 years age group (Figure 12). Patients aged ≥ 40 years with chronic conditions had greater rates of blood pressure recording, ranging from 84.1% in regular attenders with a diagnosis of depression/anxiety to 93.0% in those with chronic kidney disease (CKD). As high blood pressure is a common risk factor for chronic conditions such as CVD, diabetes and CKD, this may reflect higher rates of recording risk factors among patients with these conditions.

TABLE 15: PATIENTS AGED ≥ 40 YEARS IN FY10 WITH BLOOD PRESSURE RECORDED DURING 1 JULY 2018 TO 30 JUNE 2020, BY PATIENT CHARACTERISTICS

	% of patients aged ≥ 40 years in FY10 with at least one blood pressure record		
	All Patients (N = 1,464,585)	Regular attenders (N = 1,020,629)	Infrequent attenders (N = 443,956)
Patients aged ≥ 40 years with blood pressure available	65.1%	79.7%	31.6%
Sex			
Male	65.9%	80.5%	34.4%
Female	64.5%	79.1%	29.1%
Intersex or indeterminate	44.9%	62.0%	27.1%
Age group (years)			
40–49	55.2%	70.4%	30.6%

	% of patients aged ≥ 40 years in FY10 with at least one blood pressure record		
	All Patients (N = 1,464,585)	Regular attenders (N = 1,020,629)	Infrequent attenders (N = 443,956)
50–59	63.2%	78.0%	33.7%
60–69	69.5%	83.7%	33.1%
70–79	75.1%	87.6%	30.1%
80–89	75.6%	86.0%	26.9%
90+	61.8%	72.3%	16.8%
Males stratified by age group (years)			
40–49	54.9%	70.3%	33.6%
50–59	64.7%	79.3%	37.5%
60–69	71.1%	84.7%	35.8%
70–79	75.3%	87.4%	31.2%
80–89	75.7%	86.2%	26.8%
90+	63.9%	74.7%	17.7%
Females stratified by age group (years)			
40–49	55.5%	70.5%	27.6%
50–59	61.9%	77.0%	30.3%
60–69	68.2%	83.0%	30.9%
70–79	74.9%	87.7%	29.1%
80–89	75.6%	85.9%	27.0%
90+	60.7%	70.9%	16.4%
Selected chronic conditions (≥ 40 years)			
Cardiovascular disease	86.0%	91.1%	36.3%
Chronic kidney disease	89.1%	93.0%	29.3%
Depression and/or anxiety	77.2%	84.1%	33.0%
Diabetes type 2/unspecified	85.2%	91.9%	37.7%
Dyslipidaemia[a]	85.8%	91.0%	42.7%
Hypertension[a]	87.2%	92.8%	47.8%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] The MedicineInsight algorithms for identifying patients with dyslipidaemia and hypertension are based on recorded diagnoses, not test results.

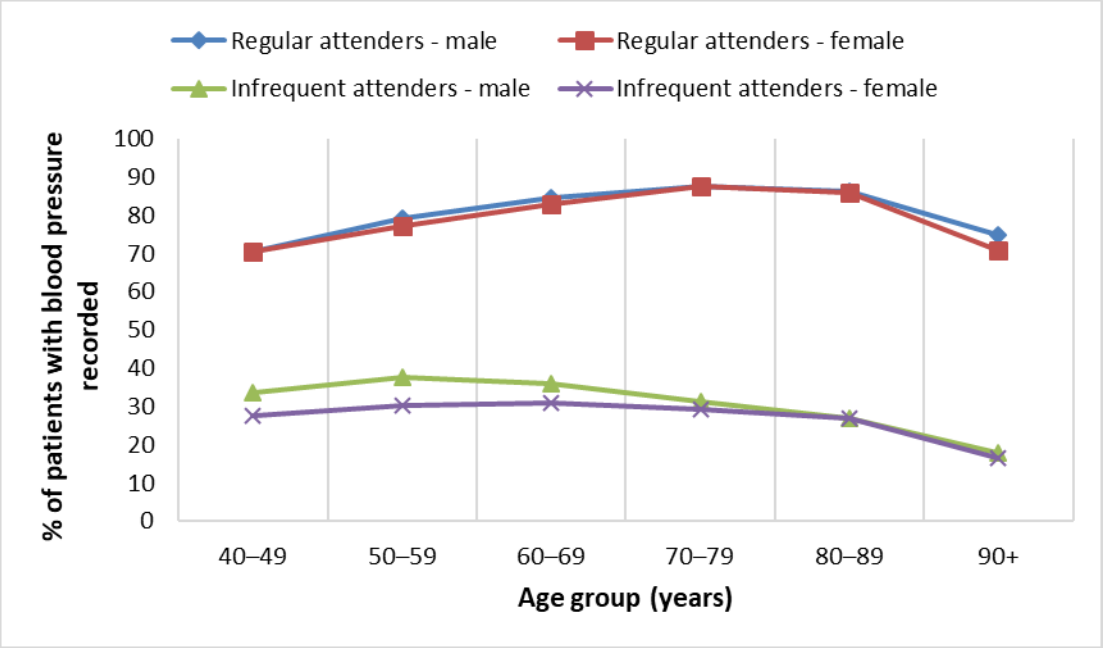

FIGURE 12: COMPLETENESS RATES OF BLOOD PRESSURE RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS BY PATIENT AGE AND SEX, 2018–2020

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

Total cholesterol

Total cholesterol (atomised results) was recorded for 48.4% of all MedicineInsight patients aged ≥ 40 years in the 24-month period from 1 July 2018 to 30 June 2020, and recording rates were greater in regular attenders than infrequent attenders (65.7% vs 8.7%) (Table 16). Completeness rates for total cholesterol in regular attenders were slightly higher for males (67.4%) than females (64.4%) and increased with age up to the 70–79 years age group for both genders (Figure 13). Recording rates for total cholesterol among regular attenders aged ≥ 40 years were high among patients with chronic conditions, ranging from 77.8% in those with hypertension to 85.1% in those with type 2 or unspecified diabetes.

TABLE 16: PATIENTS AGED ≥ 40 YEARS IN FY10 WITH TOTAL CHOLESTEROL RECORDED DURING 1 JULY 2018 TO 30 JUNE 2020, BY PATIENT CHARACTERISTICS

	% of patients aged ≥ 40 years in FY10 with at least one total cholesterol record		
	All Patients (N = 1,464,585)	Regular attenders (N = 1,020,629)	Infrequent attenders (N = 443,956)
Patients aged ≥ 40 years with total cholesterol available	48.4%	65.7%	8.7%
Sex			
Male	49.2%	67.4%	9.9%
Female	47.8%	64.4%	7.6%
Intersex or indeterminate	26.5%	48.0%	4.2%
Age group (years)			
40–49	35.7%	52.5%	8.3%
50–59	46.5%	64.7%	10.3%
60–69	55.1%	73.1%	9.1%

	% of patients aged ≥ 40 years in FY10 with at least one total cholesterol record		
	All Patients (N = 1,464,585)	Regular attenders (N = 1,020,629)	Infrequent attenders (N = 443,956)
70–79	61.3%	76.5%	6.4%
80–89	57.5%	68.6%	5.4%
90+	39.7%	47.6%	5.5%
Males stratified by age group (years)			
40–49	34.4%	52.8%	9.2%
50–59	47.3%	66.4%	12.0%
60–69	57.2%	75.1%	10.6%
70–79	62.2%	77.4%	7.2%
80–89	58.4%	69.7%	5.7%
90+	42.5%	51.0%	5.6%
Females stratified by age group (years)			
40–49	36.7%	52.4%	7.4%
50–59	45.7%	63.4%	8.9%
60–69	53.3%	71.3%	7.8%
70–79	60.4%	75.7%	5.7%
80–89	56.8%	67.7%	5.1%
90+	38.3%	45.8%	5.5%
Selected chronic conditions (≥ 40 years)			
Cardiovascular disease	72.2%	78.4%	11.7%
Chronic kidney disease	75.2%	79.3%	13.3%
Diabetes type 2/unspecified	76.2%	85.1%	13.0%
Dyslipidaemia[a]	76.2%	83.2%	19.4%
Hypertension[a]	69.7%	77.8%	12.7%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] The MedicineInsight algorithms for identifying patients with dyslipidaemia and hypertension are based on recorded diagnoses, not test results.

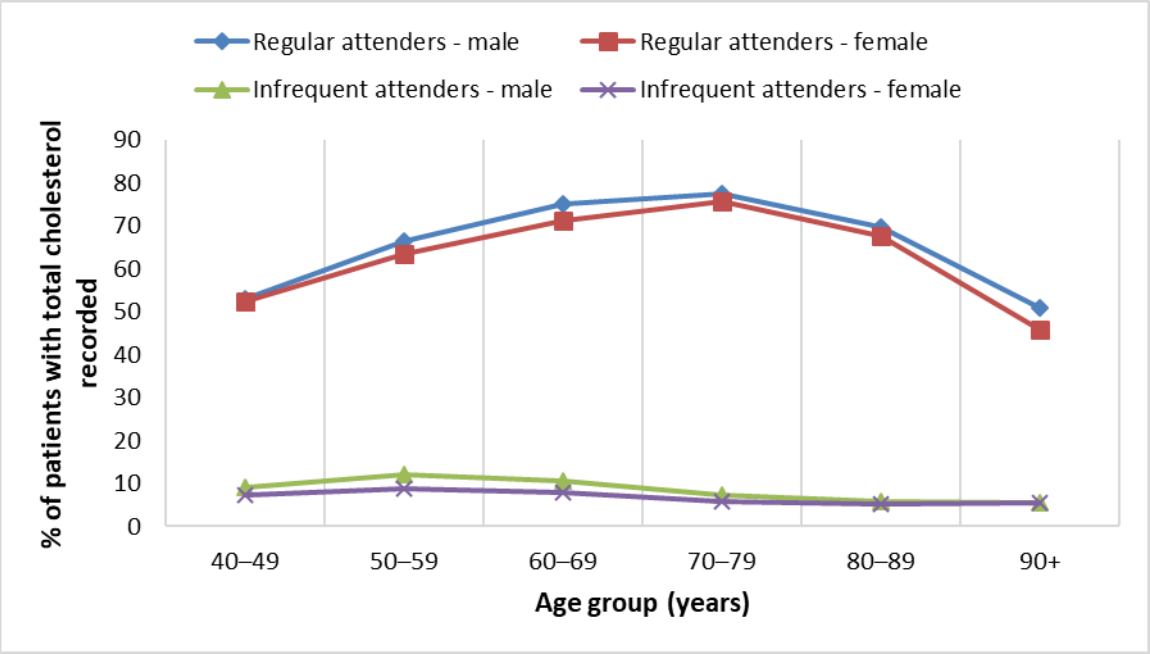

FIGURE 13: COMPLETENESS RATES OF TOTAL CHOLESTEROL RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS BY PATIENT AGE AND SEX, 2018–2020

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

Glycated haemoglobin

Glycated haemoglobin (HbA_{1c}; as defined in Table A2) was available for 13.8% of all patients in the 24-month period from 1 July 2018 to 30 June 2020, and recording rates were more than 10-fold higher in regular attenders than infrequent attenders (21.3% vs 1.9%) (Table 17). Overall, completeness rates for HbA_{1c} in regular attenders were slightly higher for males (22.3%) than females (20.4%) and mostly increased with age.

TABLE 17: PATIENTS IN FY10 WITH HBA_{1c} RECORDED DURING 1 JULY 2018 TO 30 JUNE 2020, BY PATIENT CHARACTERISTICS

	% of patients in FY10 with at least one HbA_{1c} record		
	All Patients (N = 3,161,134)	Regular attenders (N = 1,949,196	Infrequent attenders (N = 1,211,938)
HbA_{1c} available	13.8%	21.3%	1.9%
Sex			
Male	14.0%	22.3%	2.1%
Female	13.7%	20.4%	1.7%
Intersex or indeterminate	7.3%	13.6%	1.3%
Age group (years)			
0–9	0.3%	0.4%	0.0%
10–19	2.2%	3.7%	0.5%
20–29	5.2%	9.2%	1.1%
30–39	9.2%	15.0%	1.8%
40–49	15.5%	23.3%	3.0%
50–59	22.1%	31.2%	4.0%
60–69	28.7%	38.4%	3.9%
70–79	34.5%	43.2%	3.2%
80–89	34.1%	40.8%	2.9%
90+	23.4%	28.1%	3.0%
Males stratified by age group (years)			
0–9	0.3%	0.4%	0.0%
10–19	1.7%	2.8%	0.5%
20–29	4.0%	7.6%	1.0%
30–39	8.1%	14.1%	2.0%
40–49	15.5%	24.2%	3.4%
50–59	23.8%	34.0%	4.9%
60–69	31.7%	42.1%	4.9%
70–79	37.1%	46.2%	3.7%
80–89	35.6%	42.6%	3.2%
90+	25.6%	30.9%	3.1%
Females stratified by age group (years)			
0–9	0.2%	0.4%	0.0%
10–19	2.7%	4.4%	0.6%
20–29	6.1%	10.2%	1.1%
30–39	10.2%	15.5%	1.7%
40–49	15.6%	22.6%	2.6%
50–59	20.7%	29.0%	3.2%
60–69	26.1%	35.3%	3.1%
70–79	32.2%	40.5%	2.8%
80–89	33.0%	39.4%	2.7%
90+	22.2%	26.7%	3.0%

	% of patients in FY10 with at least one HbA_{1c} record		
	All Patients (N = 3,161,134)	Regular attenders (N = 1,949,196	Infrequent attenders (N = 1,211,938)
Diabetes status			
Diabetes type 2/unspecified	77.6%	86.8%	14.3%
No diabetes type 2/unspecified	10.6%	16.5%	1.7%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

Among patients with type 2 or unspecified diabetes, HbA_{1c} was available for 77.6% of all patients in the 24-month period from 1 July 2018 to 30 June 2020, and recording rates were six times greater in regular attenders with diabetes than infrequent attenders (86.8% vs 14.3%) (Table 18). Overall, completeness rates for HbA_{1c} in regular attenders were slightly higher for males (87.8%) than females (85.7%) and increased with age, with the highest recording rates in the 70–79 years age group (Figure 14). Recording rates for HbA_{1c} were high among adult regular patients with diabetes who had co-existing conditions such as CKD (90.1%), hypertension (88.7%) and CVD (88.2%).

TABLE 18: PATIENTS WITH TYPE 2 OR UNSPECIFIED DIABETES IN FY10 WITH HBA_{1C} RECORDED DURING 1 JULY 2018 TO 30 JUNE 2020, BY PATIENT CHARACTERISTICS

	% of patients with diabetes in FY10 with at least one HbA_{1c} record		
	All Patients (N = 151,980)	Regular attenders (N = 132,753)	Infrequent attenders (N = 19,227)
HbA_{1c} available for patients with diabetes type 2 or unspecified	77.6%	86.8%	14.3%
Sex			
Male	78.8%	87.8%	16.1%
Female	76.3%	85.7%	12.3%
Intersex or indeterminate	33.3%	42.9%	0.0%
Age group (years)			
0–9	8.1%	8.6%	5.6%
10–19	45.2%	51.3%	16.9%
20–29	61.5%	72.3%	14.4%
30–39	68.1%	79.6%	17.7%
40–49	72.8%	83.9%	15.9%
50–59	75.5%	86.3%	15.9%
60–69	79.5%	89.2%	15.0%
70–79	81.8%	89.7%	12.7%
80–89	79.3%	86.0%	9.8%
90+	67.1%	73.7%	10.1%
Males stratified by age group (years)			
0–9	11.3%	11.6%	10.0%
10–19	43.6%	50.3%	16.7%
20–29	61.4%	73.5%	14.2%
30–39	70.0%	81.7%	21.9%
40–49	74.0%	85.4%	17.9%
50–59	77.0%	87.3%	18.7%
60–69	80.6%	89.9%	16.7%
70–79	82.2%	90.0%	13.2%

	% of patients with diabetes in FY10 with at least one HbA_{1c} record		
	All Patients (N = 151,980)	Regular attenders (N = 132,753)	Infrequent attenders (N = 19,227)
80–89	79.7%	86.3%	10.7%
90+	69.3%	76.3%	10.7%
Females stratified by age group (years)			
0–9	5.7%	6.5%	0.0%
10–19	46.5%	52.1%	17.1%
20–29	61.6%	71.4%	14.6%
30–39	66.1%	77.4%	12.8%
40–49	71.3%	82.1%	13.1%
50–59	73.7%	85.0%	12.7%
60–69	78.1%	88.4%	13.1%
70–79	81.2%	89.3%	12.1%
80–89	78.9%	85.7%	8.9%
90+	65.5%	71.9%	9.7%
Selected chronic conditions (≥ 18 years)			
Cardiovascular disease	82.6%	88.2%	13.2%
Chronic kidney disease	85.9%	90.1%	15.9%
Dyslipidaemia[a]	83.7%	90.1%	16.3%
Hypertension[a]	81.6%	88.7%	14.6%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] The MedicineInsight algorithm for identifying patients with dyslipidaemia and hypertension is based on recorded diagnoses, not test results.

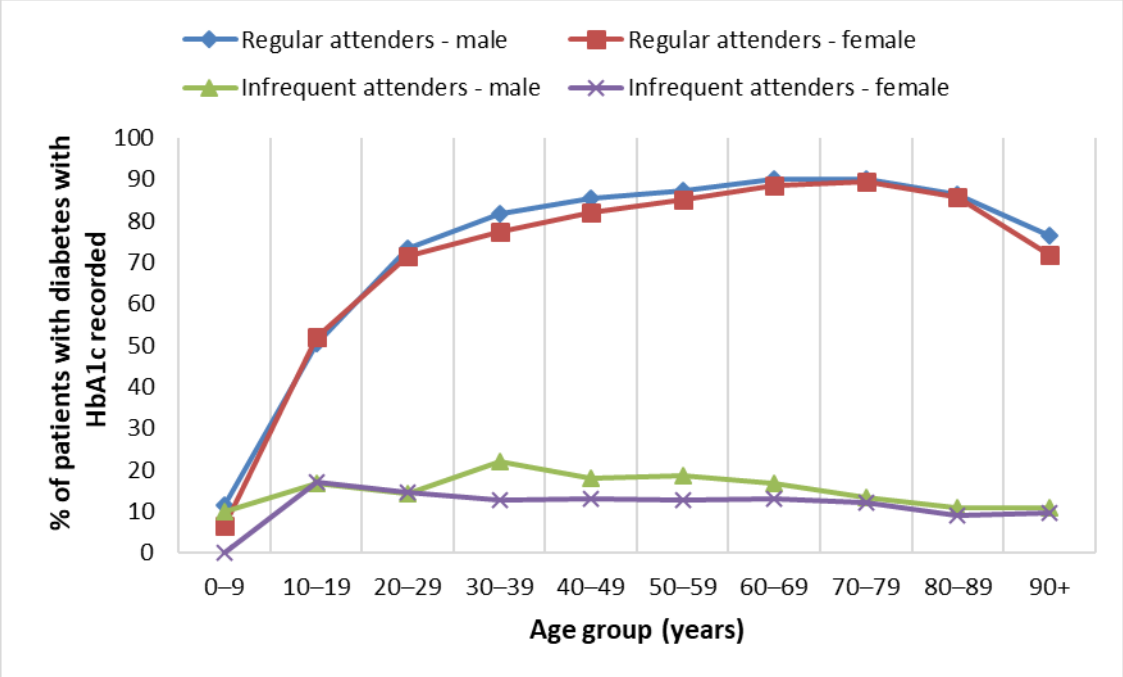

FIGURE 14: COMPLETENESS RATES OF HBA_{1C} RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS WITH TYPE 2 OR UNSPECIFIED DIABETES BY PATIENT AGE AND SEX, 2018–2020

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

Estimated glomerular filtration rate

eGFR (as defined in Table A2) was available for 53.8% of all patients aged at least 40 years in the period 1 July 2018 to 30 June 2020, and recording rates were higher in regular attenders than infrequent attenders (72.6% vs 10.3%) (Table 19). Completeness rates for eGFR were similar for males (72.9%) and females (72.4%) who were regular attenders and increased with age, with a peak in the 80–89 years age group for both genders (Figure 15). Recording rates for eGFR in regular attenders aged ≥ 40 years were high among patients with chronic conditions: 93.3% for CKD; 89.6% type 2/unspecified diabetes; 88.3% CVD; 87.0% dyslipidaemia; and 85.2% hypertension.

TABLE 19: PATIENTS AGED ≥ 40 YEARS IN FY10 WITH eGFR RECORDED DURING 1 JULY 2018 TO 30 JUNE 2020, BY PATIENT CHARACTERISTICS

	% of patients aged ≥ 40 years in FY10 with at least one eGFR record		
	All Patients (N = 1,464,585)	Regular attenders (N = 1,020,629)	Infrequent attenders (N = 443,956)
Patients aged ≥ 40 years with eGFR available	53.8%	72.6%	10.3%
Sex			
Male	53.4%	72.9%	11.3%
Female	54.0%	72.4%	9.4%
Intersex or indeterminate	25.5%	44.0%	6.3%
Age group (years)			
40–49	40.5%	59.5%	9.7%
50–59	49.8%	69.1%	11.4%
60–69	58.6%	77.4%	10.7%
70–79	67.0%	83.1%	8.7%
80–89	71.1%	84.2%	9.7%
90+	66.0%	78.2%	13.2%
Males stratified by age group (years)			
40–49	37.5%	57.2%	10.4%
50–59	49.8%	69.7%	12.9%
60–69	60.3%	78.8%	12.1%
70–79	67.6%	83.5%	9.7%
80–89	70.7%	83.8%	9.8%
90+	66.8%	79.3%	13.1%
Females stratified by age group (years)			
40–49	43.0%	61.2%	9.1%
50–59	49.7%	68.7%	10.1%
60–69	57.2%	76.2%	9.5%
70–79	66.4%	82.8%	7.9%
80–89	71.4%	84.5%	9.6%
90+	65.6%	77.7%	13.3%
Selected chronic conditions (≥ 40 years)			
Cardiovascular disease	81.6%	88.3%	15.4%
Chronic kidney disease	88.8%	93.3%	20.6%
Depression and/or anxiety	68.7%	77.3%	13.4%
Diabetes type 2/unspecified	80.5%	89.6%	15.6%
Dyslipidaemia	79.8%	87.0%	20.7%
Hypertension	76.4%	85.2%	15.1%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

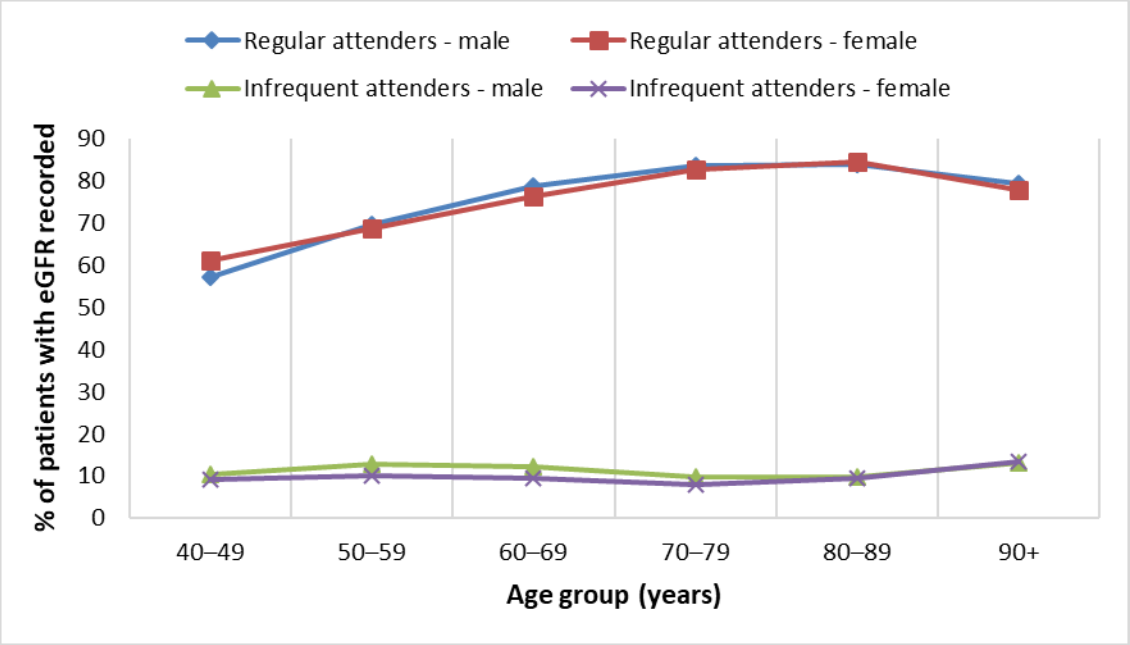

FIGURE 15: COMPLETENESS RATES OF EGFR RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS BY PATIENT AGE AND SEX, 2018–2020

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

CVD risk score

Our findings indicate low recording rates for CVD risk score (as defined in Table A2) among MedicineInsight patients who are aged at least 40 years in the 24-month period from 1 July 2018 to 30 June 2020. CVD risk score was ever recorded for 12.4% of all patients aged at least 40 years, and recording rates were higher in regular attenders than infrequent attenders overall (16.5% vs 2.9%) (Table 20) and when stratified by age group and gender (Figure 16). Completeness rates for CVD risk score were similar for males (17.0%) and females (16.0%) who were regular attenders. The highest recording rates for CVD risk score in patients aged ≥ 40 years were observed in regular attenders who had dyslipidaemia (27.3%). These results indicate relatively low recording rates for CVD risk score in patients aged at least 40 years.

TABLE 20: PATIENTS AGED ≥ 40 YEARS IN FY10 WITH CVD RISK SCORE EVER RECORDED UP TO 30 JUNE 2020, BY PATIENT CHARACTERISTICS

	% of patients aged ≥ 40 years in FY10 with at least one record of CVD risk score ever recorded		
	All Patients (N = 1,464,585)	Regular attenders (N = 1,020,629)	Infrequent attenders (N = 443,956)
Patients aged ≥ 40 years with CVD risk score available	12.4%	16.5%	2.9%
Sex			
Male	12.7%	17.0%	3.3%
Female	12.1%	16.0%	2.6%
Intersex or indeterminate	4.1%	8.0%	0.0%
Age group (years)			
40–49	7.6%	11.0%	2.1%
50–59	13.6%	18.5%	3.9%

	% of patients aged ≥ 40 years in FY10 with at least one record of CVD risk score ever recorded		
	All Patients (N = 1,464,585)	Regular attenders (N = 1,020,629)	Infrequent attenders (N = 443,956)
60–69	16.1%	21.0%	3.5%
70–79	15.0%	18.4%	2.8%
80–89	10.6%	12.5%	2.0%
90+	5.5%	6.5%	1.4%
Males stratified by age group (years)			
40–49	9.0%	13.6%	2.7%
50–59	14.2%	19.5%	4.4%
60–69	15.6%	20.2%	3.7%
70–79	14.0%	17.1%	2.6%
80–89	10.3%	12.0%	2.0%
90+	6.1%	7.1%	1.5%
Females stratified by age group (years)			
40–49	6.4%	9.0%	1.5%
50–59	13.1%	17.7%	3.3%
60–69	16.5%	21.7%	3.3%
70–79	16.0%	19.6%	3.0%
80–89	10.9%	12.8%	2.0%
90+	5.2%	6.1%	1.3%
Selected chronic conditions (≥ 40 years)			
Chronic kidney disease	18.9%	19.6%	7.2%
Depression and/or anxiety	16.5%	18.1%	5.8%
Diabetes type 2/unspecified	16.5%	18.2%	4.8%
Dyslipidaemia	25.7%	27.3%	13.3%
Hypertension	19.1%	20.9%	6.7%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

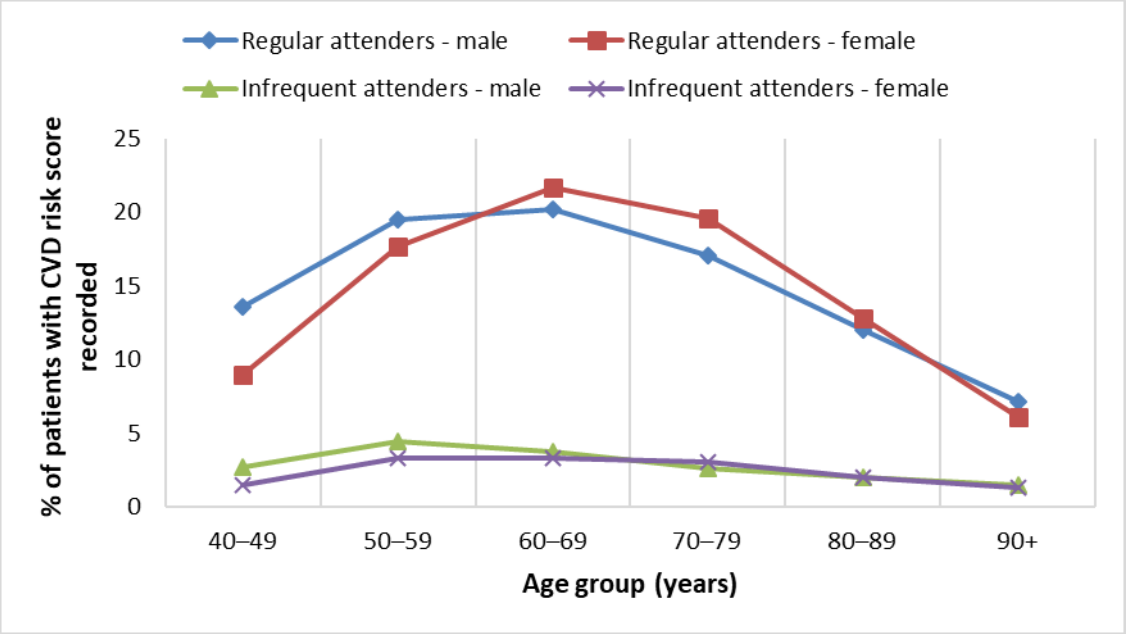

FIGURE 16: COMPLETENESS RATES OF CVD RISK SCORE EVER RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS BY PATIENT AGE AND SEX, 2020

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

Bone mineral density

While completeness rates for bone mineral density (BMD; as defined in Table A2) for patients aged at least 50 years were very low (4.0%) in the 24-month period from 1 July 2018 to 30 June 2020, BMD was more frequently recorded in regular attenders than infrequent attenders (5.3% vs 0.8%) (Table 21). Completeness rates for BMD were slightly higher for females (6.3%) than males (4.0%) who were regular attenders and increased with age up to the 80–89 years age group for both genders (Figure 17). The highest recording rates for BMD in patients aged ≥ 50 years were observed in regular attenders who had osteoporosis (12.9%). Our findings show very low recording rates for BMD in patients aged at least 50 years. This is likely due to the way results are received by the practice (in PDF format), which is not compatible with extraction to MedicineInsight. Additionally, the MBS limitations on BMD tests – eligible for patients aged ≥ 70 years and indicated for diagnosis and monitoring of low BMD in patients with specific conditions, relevant risk factors or undergoing particular treatments – may partly explain the very few BMD records.[17]

TABLE 21: PATIENTS AGED ≥ 50 YEARS IN FY10 WITH BMD EVER RECORDED UP TO 30 JUNE 2020, BY PATIENT CHARACTERISTICS

	% of patients aged ≥ 50 years with at least one record of BMD ever recorded		
	All Patients (N = 1,063,512)	Regular attenders (N = 772,495)	Infrequent attenders (N = 291,017)
Patients aged ≥ 50 years with BMD available	4.0%	5.3%	0.8%
Sex			
Male	3.1%	4.0%	0.6%
Female	4.8%	6.3%	0.9%
Intersex or indeterminate	1.9%	3.7%	0.0%
Age group (years)			
50–59	2.2%	2.9%	0.6%
60–69	3.4%	4.5%	0.7%
70–79	5.9%	7.3%	1.0%
80–89	7.8%	9.1%	1.6%
90+	6.3%	7.2%	2.1%
Males stratified by age group (years)			
50–59	1.7%	2.3%	0.5%
60–69	2.5%	3.2%	0.6%
70–79	4.5%	5.6%	0.8%
80–89	6.3%	7.3%	1.4%
90+	5.7%	6.6%	1.8%
Females stratified by age group (years)			
50–59	2.5%	3.4%	0.7%
60–69	4.2%	5.6%	0.9%
70–79	7.2%	8.9%	1.2%
80–89	9.0%	10.6%	1.8%
90+	6.6%	7.6%	2.2%
Selected chronic conditions (≥ 50 years)			
Asthma	6.1%	6.7%	1.9%
Chronic obstructive pulmonary disease	6.5%	6.8%	2.7%
Osteoarthritis	7.3%	7.7%	3.3%
Osteoporosis	12.5%	12.9%	7.0%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

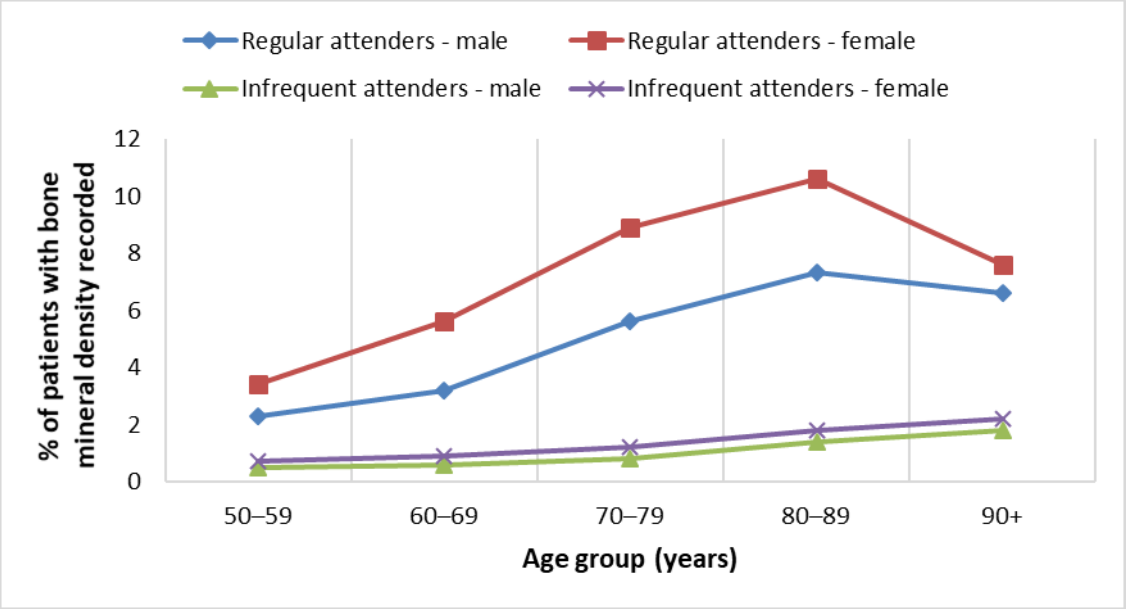

FIGURE 17: COMPLETENESS RATES OF BMD EVER RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS BY PATIENT AGE AND SEX, 2020

Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

Spirometry

Among patients with COPD and/or asthma in FY10, a spirometry record (as defined in Table A2) was available for only 2.5% of all patients with slightly higher recording rates seen in regular attenders than infrequent attenders (2.9% vs 1.0%) in the 24-month period from 1 July 2018 to 30 June 2020 (Table 22). These findings show very low completeness rates of spirometry in patients with COPD and/or asthma and suggest low recording of spirometry in the CIS fields accessible to MedicineInsight.

TABLE 22: PATIENTS WITH COPD AND/OR ASTHMA IN FY10 WITH SPIROMETRY EVER RECORDED UP TO 30 JUNE 2020

	All patients (N = 402,470)	Regular attenders (N = 320,300)	Infrequent attenders (N = 82,170)
Patients with COPD and/or asthma with spirometry available	2.5%	2.9%	1.0%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

Allergy/adverse events

An allergy or adverse event (as defined in Table A2) was ever recorded for 23.1% of all patients in the 24-month period from 1 July 2018 to 30 June 2020 FY10, and recording rates were higher in regular attenders than infrequent attenders (28.4% vs 14.6%) (Table 23). Our findings suggest low recording of allergy or adverse events in the CIS fields available to MedicineInsight even though records such as 'no known allergies' were considered as not missing and included in the completeness rates.

TABLE 23: PATIENTS IN FY10 WITH ALLERGY/ADVERSE EVENTS EVER RECORDED UP TO 30 JUNE 2020

	All patients (N = 3,161,134)	Regular attenders (N = 1,949,196)	Infrequent attenders (N = 1,211,938)
Allergy/adverse event recorded	23.1%	28.4%	14.6%

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

4.4. Recording of selected conditions

Table 24 and Figures 18 and 19 show estimates for conditions identified using two methods: the current used for MedicineInsight condition flags (code or free text record in the diagnosis, reason for encounter and reason for prescription fields) and the modified (code or free text record in the diagnosis field only). The condition estimates in the three patient cohorts – all patients, regular attenders and infrequent attenders – were lower when the modified method was used. Among regular attenders, conditions with at least a 2-point difference in estimates between the current and modified methods included anxiety disorder (17.1% vs 12.7%), cancer (10.6% vs 8.4%), depression (18.9% vs 15.9%), lower respiratory tract infection (5.6% vs 2.8%) and upper respiratory tract infection (12.0% vs 4.2%). The variation in the condition estimates identified using the two methods was less among infrequent attenders than regular attenders.

The findings also indicate that recording of conditions was significantly lower in infrequent attenders than regular attenders (Figures 18 and 19). This finding has important implications for selecting study cohorts for the purposes of estimating condition prevalence. Regular attenders are more likely to have complete records because the GP has more opportunity to collect information compared with infrequent attenders. In turn, regular attenders may be older, more unwell, and have more chronic conditions than infrequent attenders. This has the potential to lead to overestimates of condition prevalence, if younger and healthy patients who visit less frequently are excluded. However, in the Australian setting, where patients can attend multiple general practices, it is prudent to attempt to remove temporary and visitor patients from the study population when estimating condition prevalence. Medical history may not be recorded for these patients, so this approach may help to avoid underestimating prevalence.

TABLE 24: ESTIMATES FOR SELECTED CONDITIONS AMONG PATIENTS IN FY10, IDENTIFIED USING THE CURRENT AND MODIFIED METHOD

Selected conditions	All patients (N = 3,161,134)		Regular attenders (N = 1,949,196)		Infrequent attenders (N = 1,211,938)	
	Current method[a]	Modified method[b]	Current method[a]	Modified method[b]	Current method[a]	Modified method[b]
Chronic conditions (ever recorded up to 30 June 2020)						
Anxiety disorder	12.6%	9.4%	17.1%	12.7%	5.3%	4.0%
Asthma	11.2%	10.2%	14.2%	12.9%	6.4%	5.8%
Cancer	7.4%	5.9%	10.6%	8.4%	2.2%	1.8%
Cardiovascular disease	4.3%	3.9%	6.3%	5.8%	1.0%	1.0%
Chronic kidney disease	1.1%	1.0%	1.7%	1.5%	0.2%	0.2%
Chronic obstructive pulmonary disease	2.4%	2.1%	3.5%	3.1%	0.6%	0.5%
Dementia	0.6%	0.5%	0.9%	0.8%	0.2%	0.1%
Depression	14.0%	11.8%	18.9%	15.9%	6.1%	5.1%
Diabetes type 2/unspecified	4.8%	4.6%	6.8%	6.5%	1.6%	1.5%
Dyslipidaemia	11.7%	10.5%	16.9%	15.2%	3.4%	3.1%
Hypertension	14.4%	13.3%	20.3%	18.9%	4.8%	4.4%
Myocardial infarction	1.0%	0.9%	1.4%	1.4%	0.2%	0.2%
Osteoarthritis	8.4%	7.1%	12.5%	10.6%	1.7%	1.5%
Osteoporosis	3.9%	3.6%	5.9%	5.5%	0.7%	0.6%
Stroke	0.9%	0.9%	1.3%	1.2%	0.2%	0.2%
Venous thromboembolism (DVT and/or PE)	0.8%	0.6%	1.2%	0.9%	0.2%	0.2%
Acute conditions (last FY, 1 July 2019 to 30 June 2020)						
Lower respiratory tract infection	3.9%	2.0%	5.6%	2.8%	1.1%	0.6%
Otitis media	1.3%	0.7%	1.8%	0.9%	0.5%	0.3%
Upper respiratory tract infection	8.7%	3.2%	12.0%	4.2%	3.4%	1.5%
Urinary tract infection	2.4%	1.3%	3.5%	1.9%	0.6%	0.4%

[a] Current method denotes code or free text record in diagnosis/medical history, reason for encounter and reason for prescription.

[b] Modified method denotes code or free text record in diagnosis/medical history only.

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

DVT = deep vein thrombosis, PE = pulmonary embolism.

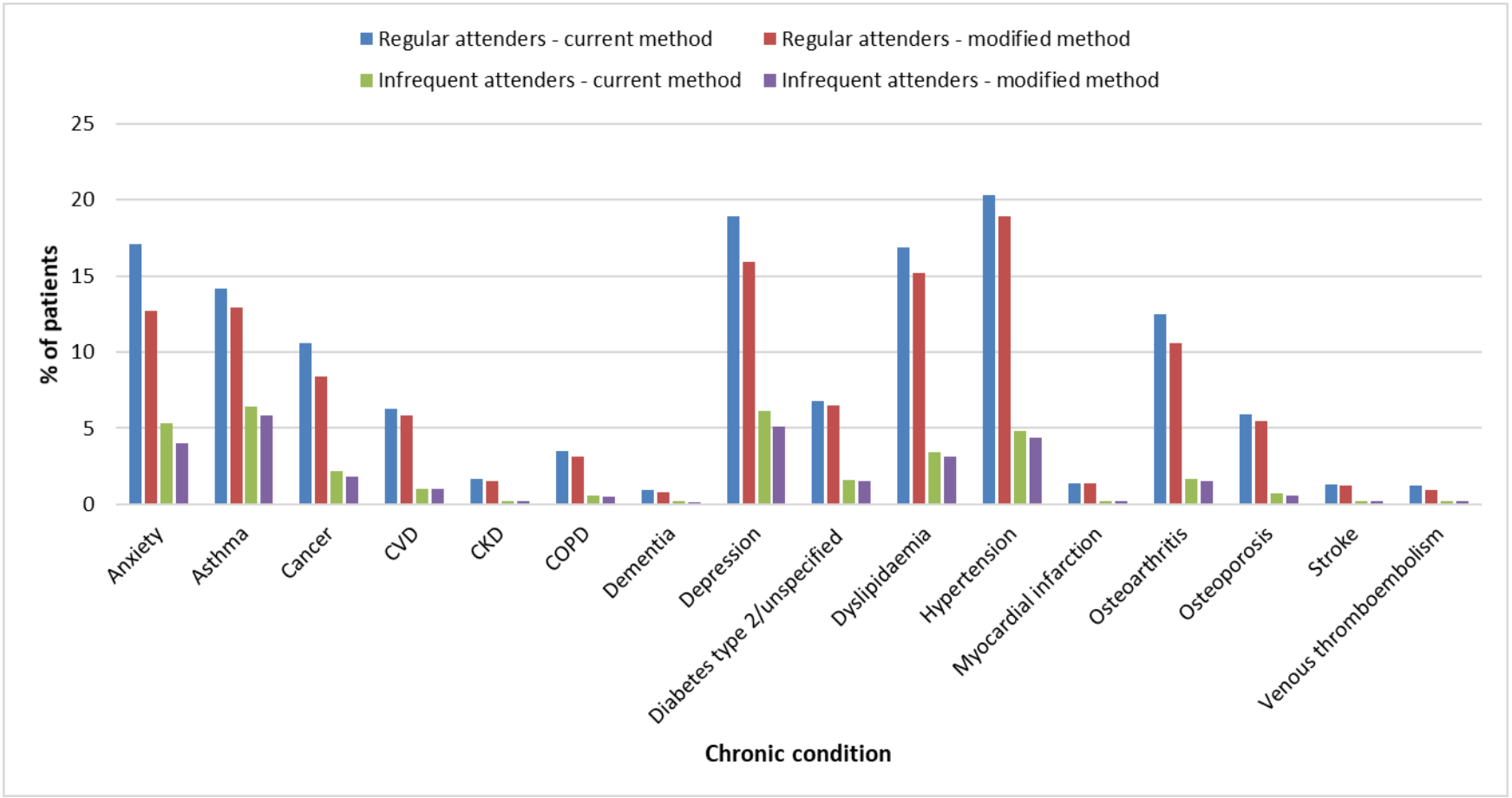

FIGURE 18: CHRONIC CONDITIONS EVER RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS BY CURRENT AND MODIFIED METHODS

Current method denotes code or free text record in diagnosis/medical history, reason for encounter and reason for prescription. Modified method denotes code or free text record in diagnosis/medical history only. Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

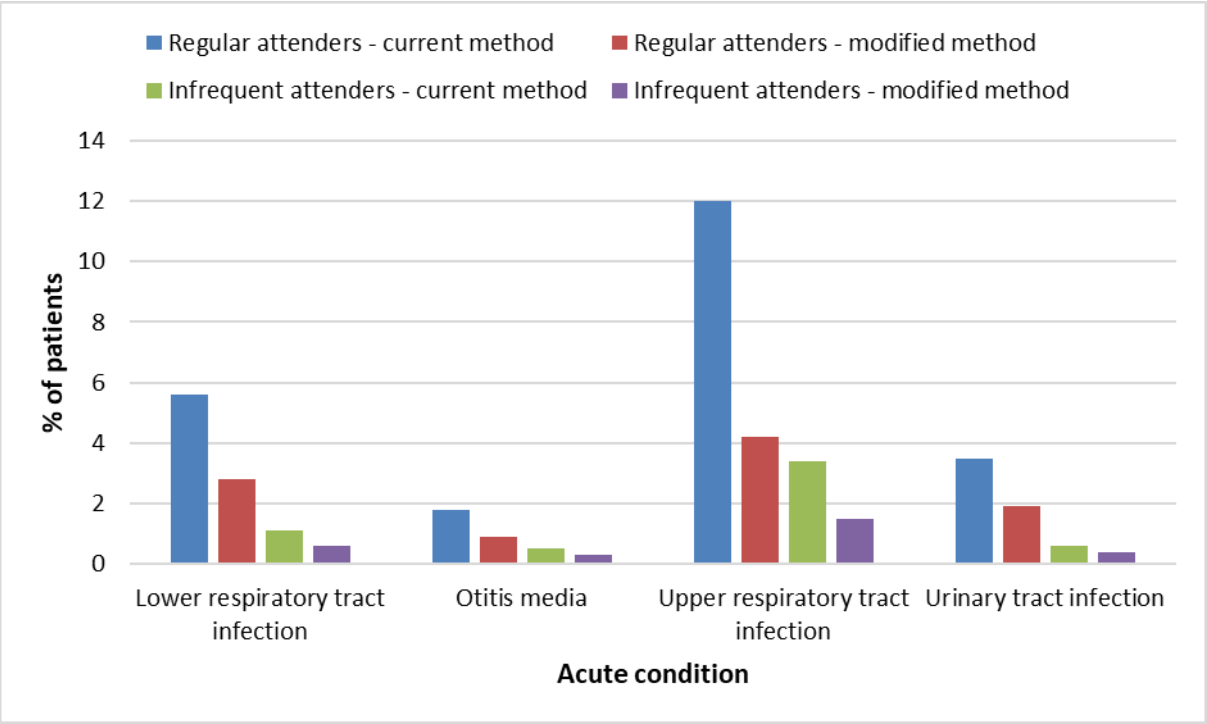

FIGURE 19: ACUTE CONDITIONS RECORDED AMONG REGULAR AND INFREQUENT ATTENDERS IN THE LAST FINANCIAL YEAR BY CURRENT AND MODIFIED METHODS

Current method denotes code or free text record in diagnosis/medical history, reason for encounter and reason for prescription. Modified method denotes code or free text record in diagnosis/medical history only.
Regular attenders are patients who had at least three clinical encounters and infrequent attenders had 1–2 clinical encounters in the previous 2 years, ending 30 June 2020.

4.5. Practice-level completeness rates for selected variables

This is an exploratory analysis to inform whether restricting to practice sites with high completeness rates for certain variables is feasible for some projects. Table 25 shows the frequency of practice sites with completeness rates of ≥ 50%, ≥ 80%, ≥ 90% and ≥ 99% for selected variables including Aboriginal and/or Torres Strait Islander status and MBS billing item for all patient age groups, and BMI and smoking status for patients aged at least 18 years in FY10.

The majority of the practice sites had at least 50% completeness rates for smoking status (98.5%), MBS billing item (97.0%) and Aboriginal and/or Torres Strait Islander status (91.6%) (Table 25). Completeness rates for smoking status of ≥ 80% were observed in 65.0% of the practice sites, just under one-third had rates of 90% and over, and fewer than 1% had rates at least 99%. Approximately 1 in 7 of the practice sites had completeness rates for MBS billing item of 80% and over and fewer than 1% had rates of at least 90%. Two-thirds of the sites (66.7%) had completeness rates of at least 80% for Aboriginal and/or Torres Strait Islander status, 42% had completeness rates of 90% and above and only 5% had completeness rates of 99% and over. Consistent with the already presented data that show low completeness rates for BMI, our findings suggest that most practice sites have poor completeness rates for BMI, with only 13.6% having rates of at least 50% and less than 1% having rates of 80% and over.

Restricting analyses to practices with at least 50% completeness rates for some key variables, such as smoking status, MBS billing item and Aboriginal and/or Torres Strait Islander status, might be considered to improve data quality without resulting in the exclusion of many practice sites (< 10%).

However, restricting analyses to practice sites with > 80% completeness for these variables would result in the exclusion of a substantial number of practice sites, which might introduce selection bias and impact on the generalisability of results. BMI should not be used as a practice-level quality marker.

Further research is required to understand whether high completeness rates for some variables at the practice level could be used as a marker of better data quality overall. Future analyses could explore practice-level completeness rates of other key variables and combinations of variables.

TABLE 25: FREQUENCY OF PRACTICE SITES WITH ≥ 50%, ≥ 80%, ≥ 90% AND ≥ 99% COMPLETENESS RATES FOR SELECTED VARIABLES

Selected variable	Number and % of practice sites with these completeness rates[a] (N = 369)			
	≥ 50%	≥ 80%	≥ 90%	≥ 99%
ATSI_NAME	336 (91.1%)	246 (66.7%)	155 (42.0%)	18 (4.9%)
BMI[b] (aged ≥ 18 years)	50 (13.6%)	< 5 (< 1.0%)	0 (0)	0 (0)
SMOKING_STATUS_NAME (aged ≥ 18 years)	363 (98.4%)	240 (65.0%)	116 (31.4%)	< 5 (< 1.0%)
MBS_BILLING_ITEM	361 (97.8%)	57 (15.4%)	< 5 (< 1.0%)	0 (0)

[a] Practice sites within each completeness rate category are not mutually exclusive. For example, practice sites within the ≥ 80%, ≥ 90% and ≥ 99% category are included in the ≥ 50% category.
[b] Included recorded and calculated (using records for height and weight) BMI.

5. GENERALISABILITY

- This chapter includes data from 441 eligible general practices, representing 5.5% of general practices nationally for the time period 1 July 2019 to 30 June 2020.
- MedicineInsight has national coverage among all states and territories, remoteness categories and PHNs (except for Western Queensland PHN). There is representative coverage of most states (NSW, VIC, QLD, WA), however practices from Tasmania (15.7% coverage) are over-represented and practices from South Australia (1.8% coverage) are under-represented.
- Practices in inner and outer regional areas are somewhat over-represented and major cities slightly under-represented.
- Coverage of PHNs fluctuates around the national estimate, 5.5%, ranging from 2.5% to 7% coverage in the majority of PHNs. MedicineInsight has less coverage in the following PHNs: Western Queensland (0.0%), Adelaide (1.9%), South Western Sydney (1.9%), Country SA (< 2.0%), Western NSW (< 2.0%), Western Sydney (2.4%), South Eastern Melbourne (2.5%) and Central and Eastern Sydney (2.6%). However, practices from the Hunter New England and Central Coast PHN (22.0%) in NSW and the Tasmania PHN (15.7%) are over-represented.
- Approximately 3.2 million eligible patients were included: 1.9 million of these were regular attenders and 1.2 million infrequent attenders.
- The demographic profile of MedicineInsight patients and MBS data on all Australian patients visiting their GP are similar in terms of age, gender and socioeconomic status.
- MedicineInsight regular attenders slightly over-represent females (56.1%) compared with MBS data (52.3%).
- Infrequent attenders over-represent younger patients aged < 40 years (63.5%) and under-represent patients aged 40+ years (36.5%) compared with MBS data (50% < 40 years).
- A similar proportion of MedicineInsight patients identified and were recorded as Aboriginal or Torres Strait Islander as in the ABS 2016 census (3.0% vs 2.8%).

5.1. Study questions

- Is the MedicineInsight patient sample comparable to the MBS patients in relation to patient characteristics such as age group, gender, socioeconomic status, geographical location and remoteness of patients?
- Are MedicineInsight general practices comparable to the Australian national estimates in relation to geographical location?

5.2. Generalisability of practices and patient characteristics

For generalisability analyses, data for patient cohorts for FY10 were used comprising approximately 3.2 million eligible patients – 1.9 million of these were regular attenders and 1.2 million infrequent attenders (see Table 6 and Figure 1 for further details on patient cohorts).

Practices

There were 441 MedicineInsight general practices that met the standard data quality criteria for inclusion in this study, representing 5.5% of general practices nationally. Table 26 presents data on MedicineInsight general practices compared with national data, by state/territory, remoteness and PHN. This table presents both the proportional geographical representation and the differences in relative coverage of MedicineInsight practices compared with national data.

MedicineInsight has national coverage among all states and territories, remoteness areas and PHNs (except for Western Queensland PHN). There is representative coverage of the largest states (NSW, VIC, QLD, WA), however practices from Tasmania (15.7% coverage) are over-represented and practices from South Australia (1.8% coverage) are under-represented. Practices in inner and outer regional areas are somewhat over-represented and major cities slightly under-represented.

Coverage of PHNs fluctuates around the national estimate, 5.5%, ranging from 2.5% to 7% coverage in the majority of PHNs. MedicineInsight has less than 3.5% coverage with less than 3.5% coverage (ie, > 2% lower coverage than the national MedicineInsight average) in the following PHNs: Adelaide (1.9%), Central and Eastern Sydney (2.6%), Country SA (< 2.0%), East Melbourne (3.0%), North Sydney (3.1%), South Eastern Melbourne (2.5%), South Western Sydney (1.9%), Western NSW (< 2.0%), Western Queensland (0.0%), and Western Sydney (2.4%). However, practices from the Hunter New England and Central Coast PHN (22.0%) in NSW and the Tasmania PHN (15.7%) are over-represented. This reflects previous active campaigns to recruit practices from these areas. To protect the confidentiality of practices, NPS MedicineWise policy is to only present results by PHN when there are three or more practices participating in the MedicineInsight program within that PHN.

TABLE 26: GEOGRAPHICAL REPRESENTATION OF MEDICINEINSIGHT GENERAL PRACTICES 2019–2020, COMPARED TO NATIONAL DATA

General practice location	MedicineInsight 2019–20		National practices 2019[a]		% Coverage of MedicineInsight practices
	n	% practices	n	% practices	%
Australian total, N	**441**	**100**	**8056[b]**	**100**	**5.5**
State/territory					
ACT	7	1.6	108	1.3	6.5
NSW	162	36.7	2763	34.3	5.9
NT	10	2.3	126	1.6	7.9
QLD	89	20.2	1633	20.3	5.5
SA	10	2.3	551	6.8	1.8
TAS	26	5.9	166	2.1	15.7
VIC	90	20.4	1986	24.7	4.5
WA	47	10.7	719	8.9	6.5
Rurality[c]					
Major city	253	57.4	5503	68.2	4.6
Inner regional	111	25.2	1396	17.3	8.0
Outer regional	65	14.7	779	9.7	8.3
Remote/very remote	12	2.7	379	4.7	3.2
Modified Monash Model					
Metropolitan (MM1)	250	56.7	-		-
Regional centre (MM2)	57	12.9	-		-
Large rural town (MM3)	46	10.4	-		-
Medium rural town (MM4)	25	5.7	-		-
Small rural town (MM5)	51	11.6	-		-
Remote/very remote (MM6 and MM7)	12	2.7	-		-
Primary Health Network					
Adelaide	7	1.6	361	4.5	1.9
Australian Capital Territory	7	1.6	108	1.3	6.5
Brisbane North	14	3.2	325	4.0	4.3
Brisbane South	20	4.5	337	4.2	5.9

General practice location	MedicineInsight 2019–20		National practices 2019[a]		% Coverage of MedicineInsight practices
	n	% practices	n	% practices	%
Central Queensland, Wide Bay, Sunshine Coast	24	5.4	276	3.4	8.7
Central and Eastern Sydney	15	3.4	579	7.2	2.6
Country SA	< 5	< 1.0	190	2.4	< 2.0
Country WA	14	3.2	198	2.5	7.1
Darling Downs and West Moreton	8	1.8	167	2.1	4.8
Eastern Melbourne	13	2.9	434	5.4	3.0
Gippsland	< 5	< 1.0	96	1.2	< 5.0
Gold Coast	9	2.0	211	2.6	4.3
Hunter New England and Central Coast	90	20.4	409	5.1	22.0
Murray	14	3.2	210	2.6	6.7
Murrumbidgee	< 5	< 1.0	87	1.1	< 5.0
Nepean Blue Mountains	5	1.1	135	1.7	3.7
North Coast	12	2.7	180	2.2	6.7
North Western Melbourne	38	8.6	568	7.1	6.7
Northern Queensland	14	3.2	253	3.1	5.5
Northern Sydney	9	2.0	292	3.6	3.1
Northern Territory	10	2.3	126	1.6	7.9
Perth North	17	3.9	266	3.3	6.4
Perth South	16	3.6	257	3.2	6.2
South Eastern Melbourne	12	2.7	482	6.0	2.5
South Eastern NSW	9	2.0	203	2.5	4.4
South Western Sydney	8	1.8	423	5.3	1.9
Tasmania	26	5.9	166	2.1	15.7
Western NSW	< 5	< 1.0	113	1.4	< 2.0
Western Queensland	-	-	64	0.8	-
Western Sydney	8	1.8	329	4.1	2.4
Western Victoria	10	2.3	210	2.6	4.8

[a] Healthdirect Australia. National Health Services Directory. Sydney: Healthdirect Australia, October 2019, https://studio.healthmap.com.au/ (accessed 21 November 2019).

[b] Including GP practices that are in the Cocos Keeling Islands and on Norfolk Island.

[c] National estimates are historical numbers from National Health Services Directory, 2017.

Patient characteristics

Approximately 3.2 million patients were eligible for inclusion in this study, representing 14.3% of all patients who visited a GP in 2019–20. Of the 3.2 million patients, 1.9 million were regular attenders (had at least three clinical encounters from 1 July 2018 to 30 June 2020) and 1.2 million were infrequent attenders (had 1–2 clinical encounters from 1 July 2018 to 30 June 2020).

MedicineInsight patients in the three cohorts – all patients, regular and infrequent attenders – are broadly similar in terms of sociodemographics when compared to national MBS information for patients who visited a GP during 2019–20, in terms of age, gender and socioeconomic status (Table 27). However, females (56.1%) are slightly over-represented in the MedicineInsight regular attenders compared with MBS data (52.3%). Furthermore, compared with MBS data, younger patients aged < 40 years (63.5%) are over-represented in infrequent attenders while patients aged ≥ 40 years (36.5%) are under-represented.

Consistent with the higher coverage of general practices from Tasmania in MedicineInsight, the proportion of MedicineInsight 'all patients' from Tasmania was higher (5.4%) when compared with the national estimate (2.1%). In line with the lower coverage of general practices from South Australia, the proportion of MedicineInsight 'all patients' from South Australia was lower (2.3%) than the national estimate (7.0%). Patients residing in inner regional areas were over-represented in MedicineInsight (22.9%, all patients) compared with national data (12.3%) (Table 27).

Aboriginal and Torres Strait Islander status was missing for 21.4% of the MedicineInsight 'all patients' population. However, a similar proportion of MedicineInsight patients identified, and were recorded as, Aboriginal or Torres Strait Islander as in the ABS 2016 national census (3.0% vs 2.8%).[18]

TABLE 27: SOCIODEMOGRAPHIC DISTRIBUTION OF MEDICINEINSIGHT PATIENTS 2019–20 COMPARED TO MBS NATIONAL DATA

Characteristic	MedicineInsight All patients		MedicineInsight Regular attenders		MedicineInsight Infrequent attenders		Australian national data (MBS)[a] 2019–2020		% Coverage of all patients
	n	%	n	%	n	%	n	%	%
Total number, N	**3,161,134**	**100**	**1,949,196**	**100**	**1,211,938**	**100**	**22,178,760**	**100**	**14.3**
Gender									
Male	1,447,775	45.8	855,299	43.9	592,476	48.9	10,583,503	47.7	13.7
Female	1,712,892	54.2	1,093,669	56.1	619,223	51.1	11,595,257	52.3	14.8
Intersex/indeterminate	467	0.0	228	0.0	239	0.0	-		
Age group (years)									
0–9	424,582	13.4	245,454	12.6	179,128	14.8	2,763,081	12.5	15.4
10–19	334,920	10.6	180,068	9.2	154,852	12.8	2,419,160	10.9	13.8
20–29	463,717	14.7	236,445	12.1	227,272	18.8	2,646,230	11.9	17.5
30–39	473,330	15.0	266,600	13.7	206,730	17.1	3,117,218	14.1	15.2
40–49	401,073	12.7	248,134	12.7	152,939	12.6	2,914,753	13.1	13.8
50–59	372,430	11.8	247,407	12.7	125,023	10.3	2,843,363	12.8	13.1
60–69	327,526	10.4	235,345	12.1	92,181	7.6	2,558,260	11.5	12.8
70–79	230,769	7.3	180,640	9.3	50,129	4.1	1,841,556	8.3	12.5
80–89	104,450	3.3	86,089	4.4	18,361	1.5	864,260	3.9	12.1
90+	28,337	0.9	23,014	1.2	5,323	0.4	210,879	1.0	13.4
Aboriginal and Torres Strait Islander status[b]									
Aboriginal and/or Torres Strait Islander	95,478	3.0	59,523	3.1	35,955	3.0	-	2.8	-
Neither Aboriginal nor Torres Strait Islander	2,389,564	75.6	1,546,100	79.3	843,464	69.6	-	91.2	-
Not recorded	676,092	21.4	343,573	17.6	332,519	27.4	-	6.0	-
State/territory									
ACT	49,425	1.6	29,234	1.5	20,191	1.7	372,196	1.7	13.3
NSW	1,152,741	36.5	727,069	37.3	425,672	35.1	7,090,958	32.0	16.3
NT	61,782	2.0	34,598	1.8	27,184	2.2	187,140	0.8	33.0
QLD	579,878	18.3	349,954	18.0	229,924	19.0	4,509,281	20.3	12.9
SA	71,150	2.3	46,263	2.4	24,887	2.1	1,551,787	7.0	4.6
TAS	171,914	5.4	115,821	5.9	56,093	4.6	467,265	2.1	36.8

Characteristic	MedicineInsight All patients		MedicineInsight Regular attenders		MedicineInsight Infrequent attenders		Australian national data (MBS)[a] 2019–2020		% Coverage of all patients
	n	%	n	%	n	%	n	%	%
VIC	705,037	22.3	420,967	21.6	284,070	23.4	5,700,158	25.7	12.4
WA	369,207	11.7	225,290	11.6	143,917	11.9	2,299,975	10.4	16.1
Rurality									
Major city	2,017,266	64.3	1,226,229	63.2	791,037	66.0	15,888,344	71.6	12.7
Inner regional	718,358	22.9	463,479	23.9	254,879	21.3	2,737,905	12.3	26.2
Outer regional	350,417	11.2	221,135	11.4	129,282	10.8	2,707,665	12.2	12.9
Remote/very remote	51,413	1.6	27,988	1.4	23,425	2.0	841,681	3.8	6.1
Missing	23,680		10,365		13,315		2546		-
Modified Monash Model									
Metropolitan (MM1)	1,978,960	62.6	1,198,776	61.5	780,184	64.4	-		-
Regional centre (MM2)	401,080	12.7	245,773	12.6	155,307	12.8	-		-
Large rural town (MM3)	304,174	9.6	203,053	10.4	101,121	8.3	-		-
Medium rural town (MM4)	129,040	4.1	81,373	4.2	47,667	3.9	-		-
Small rural town (MM5)	293,335	9.3	190,740	9.8	102,595	8.5	-		-
Remote/very remote (MM6 and MM7)	54,545	1.7	29,481	1.5	25,064	2.1	-		-
Socioeconomic status (SEIFA IRSAD quintile)									
1 (least advantaged)	497,371	15.9	318,628	16.4	178,743	14.9	3,467,086	15.6	14.3
2	603,316	19.2	378,119	19.5	225,197	18.8	3,563,822	16.1	16.9
3	681,772	21.7	431,020	22.2	250,752	20.9	4,378,392	19.7	15.6
4	637,164	20.3	381,920	19.7	255,244	21.3	4,626,996	20.9	13.8
5 (most advantaged)	717,834	22.9	429,144	22.1	288,690	24.1	6,135,506	27.7	11.7
Missing	23,677		10,365		13,312		6958		-
Concession cards									
Health Care Card	815,023	25.8	603,589	31.0	211,434	17.4	1,418,216[19]	6.5	57.5
DVA Health Card	8,392	0.3	6,653	0.3	1,739	0.1	199,123[20]	0.9	4.2
No concession card/not recorded	2,337,719	74.0	1,338,954	68.7	998,765	82.4	-		-

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] MBS data from Australian Government Department of Health.

[b] National estimates from Australian Bureau of Statistics. Census of Population and Housing – Counts of Aboriginal and Torres Strait Islander Australians, 2016.[18]

Table 28 shows the distribution of MedicineInsight patients by remoteness, stratified by state/territory compared with Australian national population estimates.[21] Although MedicineInsight patients were compared with national population estimates, the findings suggest that MedicineInsight patients are broadly comparable to national population estimates when stratified by remoteness and state/territory. However, there were variations in MedicineInsight patient and national population estimates for Tasmania, which is over-represented in MedicineInsight, and South Australia which is under-represented.

While MedicineInsight patients are generally similar to Australian population estimates[18] when stratified by Aboriginal and/or Torres Strait Islander status and state/territory, variations were observed for Tasmania which is over-represented in MedicineInsight, and South Australia which is under-represented (Table 29).

TABLE 28: MEDICINEINSIGHT PATIENTS STRATIFIED BY RURALITY AND STATE/TERRITORY

State/ territory	Major city		Inner regional		Outer regional		Remote/very remote		MedicineInsight patients with rurality missing
	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]	
All patients									
ACT	48,966 (2.4)	2.3	186 (0.0)	0.0	17 (0.0)	0.0	46 (0.1)	0.0	210 (0.9)
NSW	776,127 (38.5)	33.5	302,421 (42.1)	33.4	66,382 (18.9)	21.6	1,947 (3.8)	7.3	5,864 (24.8)
NT	59 (0.0)	0.0	48 (0.0)	0.0	47,698 (13.6)	7.2	10,440 (20.3)	20.3	3,537 (14.9)
QLD	376,007 (18.6)	17.8	116,161 (16.2)	22.0	78,721 (22.5)	33.7	4,920 (9.6)	26.0	4,069 (17.2)
SA	55,379 (2.7)	7.1	7,211 (1.0)	5.0	7,993 (2.3)	8.6	375 (0.7)	12.0	192 (0.8)
TAS	171 (0.0)	0.0	91,519 (12.7)	8.1	74,948 (21.4)	7.8	4,464 (8.7)	2.1	812 (3.4)
VIC	475,514 (23.6)	28.0	187,049 (26.0)	26.4	36,279 (10.4)	12.1	246 (0.5)	0.6	5,949 (25.1)
WA	285,043 (14.1)	11.3	13,763 (1.9)	5.1	38,379 (11.0)	9.0	28,975 (56.4)	31.7	3,047 (12.9)
Australia	2,017,266 (100)	100	718,358 (100)	100	350,417 (100)	100	51,413 (100)	100	23,680 (100)
Regular attenders									
ACT	29,005 (2.4)	2.3	106 (0.0)	0.0	8 (0.0)	0.0	18 (0.1)	0.0	97 (0.9)
NSW	482,334 (39.3)	33.5	201,134 (43.4)	33.4	40,154 (18.2)	21.6	854 (3.1)	7.3	2,593 (25.0)
NT	7 (0.0)	0.0	13 (0.0)	0.0	26,867 (12.1)	7.2	5,793 (20.7)	20.3	1,918 (18.5)
QLD	223,796 (18.3)	17.8	74,110 (16.0)	22.0	49,347 (22.3)	33.7	1,861 (6.6)	26.0	840 (8.1)
SA	36,728 (3.0)	7.1	4,120 (0.9)	5.0	5,232 (2.4)	8.6	92 (0.3)	12.0	91 (0.9)
TAS	36 (0.0)	0.0	60,011 (12.9)	8.1	52,559 (23.8)	7.8	2,922 (10.4)	2.1	293 (2.8)
VIC	279,045 (22.8)	28.0	115,905 (25.0)	26.4	23,038 (10.4)	12.1	101 (0.4)	0.6	2,878 (27.8)
WA	175,278 (14.3)	11.3	8,080 (1.7)	5.1	23,930 (10.8)	9.0	16,347 (58.4)	31.7	1,655 (16.0)
Australia	1,226,229 (100)	100	463,479 (100)	100	221,135 (100)	100	27,988 (100)	100	10,365 (100)
Infrequent attenders									
ACT	19,961 (2.5)	2.3	80 (0.0)	0.0	9 (0.0)	0.0	28 (0.1)	0.0	113 (0.8)
NSW	293,793 (37.1)	33.5	101,287 (39.7)	33.4	26,228 (20.3)	21.6	1,093 (4.7)	7.3	3,271 (24.6)
NT	52 (0.0)	0.0	35 (0.0)	0.0	20,831 (16.1)	7.2	4,647 (19.8)	20.3	1,619 (12.2)
QLD	152,211 (19.2)	17.8	42,051 (16.5)	22.0	29,374 (22.7)	33.7	3,059 (13.1)	26.0	3,229 (24.3)

State/ territory	Major city		Inner regional		Outer regional		Remote/very remote		MedicineInsight patients with rurality missing
	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]	
SA	18,651 (2.4)	7.1	3,091 (1.2)	5.0	2,761 (2.1)	8.6	283 (1.2)	12.0	101 (0.8)
TAS	135 (0.0)	0.0	31,508 (12.4)	8.1	22,389 (17.3)	7.8	1,542 (6.6)	2.1	519 (3.9)
VIC	196,469 (24.8)	28.0	71,144 (27.9)	26.4	13,241 (10.2)	12.1	145 (0.6)	0.6	3,071 (23.1)
WA	109,765 (13.9)	11.3	5,683 (2.2)	5.1	14,449 (11.2)	9.0	12,628 (53.9)	31.7	1,392 (10.5)
Australia	791,037 (100)	100	254,879 (100)	100	129,282 (100)	100	23,425 (100)	100	13,315 (100)

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] Australian population estimates as at 30 June 2018, https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3218.02017-18?OpenDocument

TABLE 29: MEDICINEINSIGHT PATIENTS STRATIFIED BY ABORIGINAL AND/OR TORRES STRAIT ISLANDER STATUS AND STATE/TERRITORY

State/territory	Aboriginal and/or Torres Strait Islander		Not Aboriginal and/or Torres Strait Islander		Not recorded	
	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]
All patients						
ACT	600 (0.6)	1.0	39,378 (1.6)	1.7	9,447 (1.4)	1.4
NSW	42,888 (44.9)	33.3	891,672 (37.3)	32.0	218,181 (32.3)	31.0
NT	4,197 (4.4)	9.0	50,568 (2.1)	0.7	7,017 (1.0)	1.6
QLD	25,438 (26.6)	28.7	462,504 (19.4)	19.7	91,936 (13.6)	21.7
SA	795 (0.8)	5.3	52,955 (2.2)	7.3	17,400 (2.6)	6.1
TAS	5,319 (5.6)	3.6	130,493 (5.5)	2.1	36,102 (5.3)	2.2
VIC	7,610 (8.0)	7.4	476,768 (20.0)	25.9	220,659 (32.6)	24.6
WA	8,631 (9.0)	11.7	285,226 (11.9)	10.5	75,350 (11.1)	11.4
Australia	95,478 (100)	100	2,389,564 (100)	100	676,092 (100)	100
Regular attenders						
ACT	371 (0.6)	1.0	24,270 (1.6)	1.7	4,593 (1.3)	1.4
NSW	27,640 (46.4)	33.3	589,850 (38.2)	32.0	109,579 (31.9)	31.0
NT	2,016 (3.4)	9.0	29,640 (1.9)	0.7	2,942 (0.9)	1.6
QLD	15,151 (25.5)	28.7	290,735 (18.8)	19.7	44,068 (12.8)	21.7
SA	496 (0.8)	5.3	35,769 (2.3)	7.3	9,998 (2.9)	6.1

State/territory	Aboriginal and/or Torres Strait Islander		Not Aboriginal and/or Torres Strait Islander		Not recorded	
	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]	MedicineInsight patients, n (%)	National population estimates (%)[a]
TAS	4,075 (6.8)	3.6	92,181 (6.0)	2.1	19,565 (5.7)	2.2
VIC	4,765 (8.0)	7.4	301,622 (19.5)	25.9	114,580 (33.3)	24.6
WA	5,009 (8.4)	11.7	182,033 (11.8)	10.5	38,248 (11.1)	11.4
Australia	59,523 (100)	100	1,546,100 (100)	100	343,573 (100)	100
Infrequent attenders						
ACT	229 (0.6)	1.0	15,108 (1.8)	1.7	4,854 (1.5)	1.4
NSW	15,248 (42.4)	33.3	301,822 (35.8)	32.0	108,602 (32.7)	31.0
NT	2,181 (6.1)	9.0	20,928 (2.5)	0.7	4,075 (1.2)	1.6
QLD	10,287 (28.6)	28.7	171,769 (20.4)	19.7	47,868 (14.4)	21.7
SA	299 (0.8)	5.3	17,186 (2.0)	7.3	7,402 (2.2)	6.1
TAS	1,244 (3.5)	3.6	38,312 (4.5)	2.1	16,537 (5.0)	2.2
VIC	2,845 (7.9)	7.4	175,146 (20.8)	25.9	106,079 (31.9)	24.6
WA	3,622 (10.1)	11.7	103,193 (12.2)	10.5	37,102 (11.2)	11.4
Australia	35,955 (100)	100	843,464 (100)	100	332,519 (100)	100

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] Australian 2016 population census estimates, http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/2075.0Main+Features12016?OpenDocument

6. PLAUSIBILITY

- The majority of MedicineInsight patients who had at least one record for each of the patient characteristics (height, weight, and BMI for adults; systolic and diastolic blood pressure and encounters per day for all patients) had plausible results, with < 1% of patients outside the plausible value range.
- The regular patient prevalence estimates for most of the conditions (atrial fibrillation, CKD, COPD, type 1 diabetes, migraine, myocardial infarction, rheumatoid arthritis and stroke) align with the 2017–18 ABS NHS, or are slightly higher (anxiety, asthma, CVD, type 2 diabetes, heart failure, low back pain, osteoarthritis and osteoporosis).
- The prevalence of all chronic and acute conditions was higher among regular attenders than infrequent attenders. As such, the infrequent patient prevalence estimates were substantially lower than the national estimates, however these findings have not been adjusted for age.
- Depending on the age group and the condition of interest, the infrequent attender cohort may not be generalisable to the Australian patient population when considering prevalence estimates.
- At ATC level 1, the proportions of total prescriptions ordered for regular attenders closely matched the proportions of prescriptions dispensed on the PBS.
- Cardiovascular medicines accounted for 31.5% of total prescriptions prescribed for MedicineInsight regular attenders and 31.3% of prescriptions dispensed on the PBS.
- Medicines for the nervous system, which include the analgesics, were the next most common prescriptions, accounting for 23.0% of total MedicineInsight prescriptions and 21.8% of PBS prescriptions, while medicines for the alimentary tract and metabolic system accounted for 14.6% and 15.9% of total prescriptions for MedicineInsight and the PBS, respectively.
- The proportions of total prescriptions for infrequent attenders do not align with the proportions of PBS prescriptions. In most cases the proportions were lower among infrequent attenders, with the exception of genitourinary system and sex hormones (which includes contraceptives) and anti-infectives for systemic use. These findings can be explained by the younger demographics of the infrequent attenders and the assumption that infrequent attenders often present for acute conditions such as those requiring antibiotics.
- Based on a preliminary analysis of a representative sample of 958,641 (30.3%) patients from the 2019/20 FY cohort who were included in the privacy preserving (bloom filter) linkage pilot project, 3.8% were identified as duplicate patients, matched to either one other patient (3.6%) or more than one other patient (0.2%). Most duplicate patients (n = 34,674) were only matched to one other patient, and of these 10% were identified as the same patient within a practice site and 90% were identified as the same patient between practice sites.

6.1. Study questions

- Are the MedicineInsight prevalence estimates of selected conditions comparable to Australian national estimates such as estimates from the ABS NHS?
- Are the MedicineInsight medicine utilisation estimates comparable to the Australian national estimates such as the PBS dispensing data?
- What are the plausibility rates for selected patient characteristics such as height, weight, BMI, blood pressure and clinical encounters per day?

6.2. Plausibility of MedicineInsight data

For plausibility analyses, data for patient cohorts for FY10 were used comprising approximately 3.2 million eligible patients: 1.9 million of these were regular attenders and 1.2 million were infrequent attenders (see Table 6 and Figure 1 for further details on patient cohorts).

Plausible values for selected patient characteristics

The plausibility of selected patient characteristics (height, weight, and BMI for adults; systolic and diastolic blood pressure and encounters per day for all patients) recorded in MedicineInsight in the period 1 July 2018 to 30 June 2020 is presented in Table 30. The majority of the 'all patients' cohort with at least one record for each of the patient characteristics had plausible results, with only 0.02% (systolic blood pressure) to 0.4% (height and clinical encounters per day) of patients outside the plausible value range. Regular attenders were more likely than infrequent attenders to have implausible values for clinical encounters per day (0.6% vs 0.1%). This is probably due to regular attenders being more likely to have more records available than infrequent attenders, for example, 61.2% of regular attenders had a BP recorded compared with 23.0% of irregular attenders.

TABLE 30: PLAUSIBILITY OF PATIENT CHARACTERISTICS IN MEDICINEINSIGHT, 2018–2020

Characteristic[a]	MedicineInsight		Plausible values		MedicineInsight patients outside plausible range[c]		Total number of patients with records (% of all patients)[d]
	Average [b]	Range (min–max)	Minimum	Maximum	% patients	95% CI	
All patients							
Height (cm; adult)	169.5	0–19481	112	251	0.4	(0.3, 0.4)	730,399 (29.5%)
Weight (kg; adult)	83.7	0–94173	25	610	0.2	(0.1, 0.2)	849,943 (34.3%)
BMI (kg/m^2; adult)	30.8	0–320818	12	70	0.2	(0.2, 0.2)	702,765 (28.4%)
Systolic BP (mm Hg)[e]	127.6	0–13080	50	250	0.02	(0.02, 0.03)	1,471,860 (46.6%)
Diastolic BP (mm Hg)[e]	77.9	0–991	30	140	0.1	(0.1, 0.1)	1,471,181 (46.5%)
Clinical encounters per patient per day	1.1	1–18	1	3	0.4	(0.3, 0.5)	3,161,134 (100%)
Regular attenders							
Height (cm; adult)	168.9	0–19481	112	251	0.3	(0.3, 0.4)	615,367 (39.3)
Weight (kg; adult)	83.9	0–94173	25	610	0.2	(0.1, 0.2)	720,968 (46.1)
BMI (kg/m^2; adult)	31.2	0–320818	12	70	0.2	(0.2, 0.2)	593,173 (37.9)
Systolic BP (mm Hg)[d]	127.8	0–13080	50	250	0.02	(0.02, 0.03)	1,192,878 (61.2)
Diastolic BP (mm Hg)[d]	77.8	0–991	30	140	0.1	(0.1, 0.1)	1,192,447 (61.2)
Clinical encounters per patient per day	1.1	1–18	1	3	0.6	(0.5, 0.8)	1,949,196 (100%)
Infrequent attenders							
Height (cm; adult)	172.5	0–18195	112	251	0.4	(0.3, 0.5)	115,032 (12.6)
Weight (kg; adult)	82.5	0–8403	25	610	0.1	(0.1, 0.2)	128,975 (14.1)
BMI (kg/m^2; adult)	28.6	0–24005	12	70	0.2	(0.1, 0.2)	109,592 (12.0)
Systolic BP (mm Hg)[e]	126.7	1–905	50	250	0.03	(0.02, 0.04)	278,982 (23.0)
Diastolic BP (mm Hg)[e]	78.6	0–987	30	140	0.1	(0.1, 0.1)	278,734 (23.0)
Clinical encounters per patient per day[f]	1.1	1–8[e]	1	3	0.1	(0, 0.2)	1,211,938 (100%)

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

[a] Latest record of height, weight, BMI, systolic and diastolic BP between 1 July 2018 and 30 June 2020 was used for each patient.

[b] Excluding patients with missing records.

[c] The averages were calculated based on all values, regardless of plausibility, and significant outliers have impacted results.

[d] Patient populations: a) All patients (adult, ≥ 18 years) = 3,161,134 (2,476,035); b) Regular (adult) = 1,949,196 (1,564,247); c) Infrequent (adult) = 1,211,938 (911,788).

[e] If more than one value for systolic or diastolic BP were recorded on the same day, the highest value was used. Included patients of all ages.

[f] Some infrequent attenders (n=6,648) had more than two clinical encounters on the same day.

External validity of condition prevalence estimates

For plausibility, MedicineInsight prevalence estimates for all patients, regular attenders and infrequent attenders were compared to the ABS NHS estimates or other available Australian estimates. It is intended that these findings will assist Australian researchers in both designing and interpreting studies of condition prevalence using EHRs.

The patient prevalence of most of the conditions was strikingly similar for both the MedicineInsight (MI) all patients cohort and the 2017–18 ABS NHS[8] (Table 31), including anxiety disorder (12.6% MI; 13.1% ABS) asthma (11.2% MI; 11.2% ABS), atrial fibrillation (2.0% MI; 1.9% ABS), CKD (1.1% MI; 1.0% ABS), COPD (2.4% MI: 2.5% ABS), CVD (4.3% MI; 4.8% ABS), type 1 diabetes (0.5% MI; 0.6% ABS), type 2 diabetes (4.8% MI; 4.1% ABS), osteoarthritis (8.4% MI; 9.3% ABS), osteoporosis (3.9% MI; 3.8% ABS) and stroke (0.9% MI; 1.2% ABS). As expected, the prevalence of every chronic and acute condition was higher among regular attenders than infrequent attenders (Tables 31 and 32).

Among regular attenders, the patient prevalence of most of the conditions was similar to the 2017–18 ABS NHS,[8] including for atrial fibrillation, CKD, COPD, type 1 diabetes, migraine, myocardial infarction, rheumatoid arthritis and stroke. Myocardial infarction was recorded for 1.4% of MedicineInsight regular patients and 1.6% of ABS NHS participants. Similarly, stroke was recorded for 1.3% of MedicineInsight regular patients and 1.2% of ABS NHS participants (Table 31).

Patient prevalence estimates were slightly higher for a number of conditions in the MedicineInsight regular attenders cohort than in the 2017–18 ABS NHS, including anxiety, asthma, CVD, type 2 diabetes, heart failure, low back pain, osteoarthritis and osteoporosis. The difference between the two data sets was greatest for depression, dermatitis/eczema, dyslipidaemia and hypertension. In the MedicineInsight cohort, the proportion of regular attenders with hypertension was 20.3%, compared with 10.6% among ABS NHS participants. The proportion of regular patients with dyslipidaemia was 16.9% compared with 6.1% of ABS NHS participants reporting high cholesterol. While the all patient (regular and infrequent attenders) estimates for hypertension (14.4%) and dyslipidaemia (11.7%) were closer, they were still higher than ABS estimates. The differences in prevalence may be a reflection of the different populations from which the data are drawn (patient vs general population), the different collection methods (self-reported data compared with secondary use of electronic health records) and the definitions of the conditions used in the two studies.

MedicineInsight data include people visiting their GPs, while the ABS NHS data were collected via self-report from people randomly selected from the general population. Therefore, as a group, patients from the MedicineInsight population will differ from those in the ABS NHS. Additionally, the ABS NHS asked respondents mostly about ‘current conditions’ (defined as medical conditions that have lasted, or are expected to last, for 6 months or more), while MedicineInsight data were based on whether a GP had ‘ever’ recorded a condition in a patient’s medical history. Prevalence estimates using MedicineInsight data do not account for conditions resolving over time, as depression can, so they have the potential to overestimate prevalence. This is less of an issue for chronic conditions which are unlikely to resolve, such as type 2 diabetes.

The prevalence estimates among infrequent patients were substantially lower than the national estimates, however these findings have not been adjusted for age (Table 31). Infrequent attenders are younger than the national MBS patient comparator, with 63.5% of the infrequent cohort aged < 40 years compared with 50% of the MBS patients (Table 27). Therefore the prevalence of most chronic conditions is expected to be lower in the infrequent attender cohort. However, after adjusting for age, the prevalence of most conditions is expected to be underestimated due to the expected under-recording of conditions in this cohort, which might include temporary and visitor patients. Depending on the age group and the condition of interest, the infrequent attender cohort may not be fit for purpose when considering prevalence estimates.

TABLE 31: PROPORTION OF MEDICINEINSIGHT PATIENTS WITH SELECTED CHRONIC AND OTHER CONDITIONS EVER RECORDED UP TO 30 JUNE 2020 COMPARED WITH AUSTRALIAN NATIONAL ESTIMATES

Condition	MedicineInsight All patients (N = 3,161,134)		MedicineInsight Regular attenders (N = 1,949,196)		MedicineInsight Infrequent attenders (N = 1,211,938)		ABS NHS[a]	Other Australian estimates
	%	95% CI	%	95% CI	%	95% CI	% patients	% patients
Anxiety disorder	12.6	(12.0, 13.2)	17.1	(16.4, 17.7)	5.3	(5.0, 5.6)	13.1	
Asthma	11.2	(10.7, 11.7)	14.2	(13.7, 14.7)	6.4	(6.0, 6.7)	11.2	
Atrial fibrillation	2.0	(1.8, 2.2)	2.9	(2.7, 3.2)	0.5	(0.4, 0.5)	1.9[b]	
Breast cancer	1.0	(0.9, 1.1)	1.4	(1.3, 1.5)	0.3	(0.3, 0.4)	0.3[c]	
Chronic kidney disease	1.1	(1.0, 1.2)	1.7	(1.5, 1.9)	0.2	(0.2, 0.2)	1.0	
Chronic obstructive pulmonary disease	2.4	(2.2, 2.5)	3.5	(3.2, 3.7)	0.6	(0.5, 0.6)	2.5	
Cardiovascular disease[d]	4.3	(3.9, 4.7)	6.3	(5.8, 6.8)	1.0	(0.9, 1.1)	4.8	
Dementia	0.6	(0.5, 0.7)	0.9	(0.8, 1.0)	0.2	(0.1, 0.2)		0.9[22]
Depression	14.0	(13.3, 14.7)	18.9	(18.1, 19.7)	6.1	(5.6, 6.5)	10.4	
Dermatitis/eczema	6.5	(6.1, 6.9)	8.5	(8.0, 9.0)	3.3	(3.1, 3.6)	1.0	
Diabetes (gestational)	0.6	(0.6, 0.7)	0.9	(0.8, 1.0)	0.3	(0.2, 0.3)		0.2[23]
Diabetes (type 1)	0.5	(0.5, 0.5)	0.6	(0.6, 0.7)	0.2	(0.2, 0.3)	0.6	
Diabetes (type 2/unspecified)	4.8	(4.5, 5.1)	6.8	(6.5, 7.1)	1.6	(1.5, 1.7)	4.1	
Dyslipidaemia	11.7	(11.0, 12.4)	16.9	(16.1, 17.6)	3.4	(3.2, 3.7)	6.1[e]	
Gastro-oesophageal reflux disease	11.1	(10.5, 11.7)	15.8	(15.1, 16.5)	3.5	(3.3, 3.8)		11.6[24]
Heart failure	1.0	(0.9, 1.1)	1.5	(1.4, 1.7)	0.2	(0.2, 0.2)	0.5	
Hypertension	14.4	(13.6, 15.2)	20.3	(19.4, 21.3)	4.8	(4.5, 5.1)	10.6[f]	
Low back pain	13.1	(12.5, 13.7)	18.3	(17.6, 19.1)	4.6	(4.3, 5.0)	16.4[g]	
Migraine	4.3	(4.1, 4.5)	5.8	(5.6, 6.0)	1.8	(1.7, 1.9)	6.2	
Myocardial infarction	1.0	(0.9, 1.1)	1.4	(1.3, 1.5)	0.2	(0.2, 0.3)	1.6	
Osteoarthritis	8.4	(7.7, 9.0)	12.5	(11.7, 13.3)	1.7	(1.6, 1.9)	9.3	
Osteoporosis	3.9	(3.5, 4.3)	5.9	(5.4, 6.4)	0.7	(0.6, 0.8)	3.8	
Acute pancreatitis	0.1	(0.1, 0.1)	0.1	(0.1, 0.2)	0.02	(0.02, 0.03)		1.0[25]
Prostate cancer	0.7	(0.7, 0.8)	1.1	(1.0, 1.2)	0.2	(0.1, 0.2)	0.3[c]	
Rheumatoid arthritis	0.7	(0.7, 0.7)	1.0	(0.9, 1.0)	0.2	(0.2, 0.3)	1.9	
Stroke	0.9	(0.8, 1.0)	1.3	(1.2, 1.4)	0.2	(0.2, 0.3)	1.2	

Condition	MedicineInsight All patients (N = 3,161,134)		MedicineInsight Regular attenders (N = 1,949,196)		MedicineInsight Infrequent attenders (N = 1,211,938)		ABS NHS[a]	Other Australian estimates
	%	95% CI	%	95% CI	%	95% CI	% patients	% patients
Suicide[h]	0.4	(0.3, 0.4)	0.5	(0.5, 0.6)	0.1	(0.1, 0.1)		0.013[26]
Venous thromboembolism	0.8	(0.7, 0.9)	1.2	(1.1, 1.3)	0.2	(0.2, 0.2)		

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.
[a] Defined as a current medical condition that has lasted, or is expected to last, for 6 months or more, unless otherwise stated. Non-age-standardised rate provided.
[b] Rapid or irregular heartbeat, tachycardia or palpitations.
[c] Estimated from results presented in the ABS NHS report, https://www.abs.gov.au/statistics/health/health-conditions-and-risks/cancer/latest-release
[d] Includes coronary artery disease, peripheral vascular disease stroke and transient ischaemic attack.
[e] Self-reported high cholesterol only.
[f] Self-reported hypertension only. Excludes measured high blood pressure.
[g] Includes sciatica, disc disorders, back pain/problems not elsewhere classified and curvature of the spine.
[h] Note that MedicineInsight estimates include both suicide and self-harm.

TABLE 32: PROPORTION OF MEDICINEINSIGHT PATIENTS WITH SELECTED ACUTE CONDITIONS RECORDED IN 2019–2020 COMPARED WITH AUSTRALIAN NATIONAL ESTIMATES

Condition	MedicineInsight All patients		MedicineInsight Regular attenders		MedicineInsight Infrequent attenders		ABS NHS	Other estimates
	%	95% CI	%	95% CI	%	95% CI	% patients	% patients
Lower respiratory tract infection	3.9	(3.7, 4.1)	5.6	(5.3, 5.9)	1.1	(1.0, 1.2)		
Otitis media	1.3	(1.2, 1.4)	1.8	(1.7, 1.9)	0.5	(0.5, 0.5)	0.4	
Upper respiratory tract infection	8.7	(8.1, 9.3)	12.0	(11.2, 12.7)	3.4	(3.1, 3.7)		
Urinary tract infection	2.4	(2.3, 2.5)	3.5	(3.4, 3.6)	0.6	(0.6, 0.7)		

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients.

External validity of prescribed medicines

Approximately 29.5 million total prescriptions (issued plus repeat prescriptions) with an assigned ATC code were prescribed to the MedicineInsight all patients cohort in FY 2019–2020 (Table 33). During the same period, there were approximately 208 million prescriptions dispensed on the PBS (ATC level 1 excluding under co-payment prescriptions).[9] Of note, MedicineInsight captures prescriptions that have been issued – whether they are private, PBS subsidised or under co-payment – while the PBS data capture prescriptions when the medicine has been dispensed on the PBS or is under co-payment.

At ATC level 1, the proportions of total prescriptions ordered for regular attenders closely match the proportions of prescriptions dispensed on the PBS, demonstrating good external validity (Table 33). However, the proportions of total prescriptions ordered for infrequent attenders do not align with the proportions of PBS prescriptions. In most cases the proportions were lower among infrequent attenders, with some notable exceptions: genitourinary system and sex hormones (which includes contraceptives) at 7.6% of infrequent attender prescriptions and 2.0% of PBS dispensing, and anti-infectives for systemic use at 19.9% of infrequent attender prescriptions and 5.7% of PBS dispensing. These findings can be explained by the younger demographics of the infrequent attenders and the assumption that infrequent attenders often present for acute conditions such as those requiring antibiotics.

Cardiovascular medicines accounted for 31.5% of total prescriptions prescribed to MedicineInsight regular attenders and 31.3% of prescriptions dispensed on the PBS. Medicines for the nervous system, which include the analgesics, were the next most common prescriptions, accounting for 23.0% of total MedicineInsight prescriptions and 21.8% of PBS prescriptions, while medicines for the alimentary tract and metabolic system accounted for 14.6% and 15.9% of total prescriptions for MedicineInsight and the PBS, respectively.

Differences between the regular attender MedicineInsight cohort and PBS prescribing data were observed. This is likely to reflect the nature of prescribing for patients seen in primary care compared with the medicines dispensed on prescriptions from all types of prescribers (including specialists, other health professionals and medicines dispensed under the PBS from a hospital). For example, medicines from the ATC G (genitourinary system and sex hormones) group account for 4.1% of total prescriptions prescribed for MedicineInsight regular patients but only 2.0% of dispensed PBS medicines. This is most likely because this group includes contraceptives and hormone replacement therapies which are much more likely to be prescribed by GPs than other prescribers. In contrast, medicines to treat cancer (ATC L group), which are most likely to be prescribed in a specialist setting, are less commonly prescribed for MedicineInsight patients (0.5%) than dispensed on the PBS (2.2%).

Other possible explanations for differences between MedicineInsight and PBS estimates are:

- MedicineInsight includes private prescriptions which are not captured by the PBS.
- MedicineInsight captures information on all prescriptions that are written, but these may not necessarily all be dispensed.

TABLE 33: NUMBER AND PROPORTION (%) OF MEDICINEINSIGHT TOTAL (ISSUED PLUS REPEATS) PRESCRIPTIONS FOR ATC LEVEL 1 COMPARED TO NUMBER AND PROPORTION (%) OF ALL PBS MEDICINES DISPENSED, 2019–20

ATC level 1	ATC level 1	MedicineInsight total prescriptions (issued plus repeats) 2019–2020						PBS 2019–2020[a]	
		All patients		Regular attenders		Infrequent attenders			
		n	%	n	%	n	%	N	%
A	Alimentary tract and metabolism	4,259,661	14.4	4,181,350	14.6	78,311	9.1	33,040,857	15.9
B	Blood and blood-forming organs	1,060,279	3.6	1,048,006	3.7	12,273	1.4	10,388,487	5.0
C	Cardiovascular system	9,157,746	31.0	9,014,841	31.5	142,905	16.6	65,309,786	31.3
D	Dermatologicals	709,743	2.4	658,991	2.3	50,752	5.9	2,982,967	1.4
G	Genitourinary system and sex hormones	1,253,348	4.2	1,187,451	4.1	65,897	7.6	4,148,639	2.0
H	Systemic hormonal preparations, excl. sex hormones and insulins	563,223	1.9	540,196	1.9	23,027	2.7	3,615,132	1.7
J	Anti-infectives for systemic use	1,990,260	6.7	1,819,099	6.3	171,161	19.9	11,979,502	5.7
L	Antineoplastic and immunomodulating agents	133,197	0.5	130,023	0.5	3174	0.4	4,510,311	2.2
M	Musculoskeletal system	1,081,337	3.7	1,047,996	3.7	33,341	3.9	6,851,278	3.3
N	Nervous system	6,751,160	22.9	6,591,252	23.0	159,908	18.5	45,444,822	21.8
P	Antiparasitic products, insecticides and repellents	45,421	0.2	42,297	0.1	3124	0.4	72,937	0.0
R	Respiratory system	2,199,244	7.5	2,100,354	7.3	98,890	11.5	12,396,771	6.0
S	Sensory organs (eye/ear)	311,716	1.1	292,371	1.0	19,345	2.2	7,408,131	3.6
V	Various	2645	0.01	2616	0.01	29	0.003	194,230	0.1
	Total	**29,518,980[b]**	**100**	**28,656,843[b]**	**100**	**862,137[b]**	**100**	**208,343,850[c]**	**100**

All patients had at least one clinical encounter in the previous 2 years, ending 30 June 2020; regular attenders (had at least three clinical encounters) and infrequent attenders (had 1–2 clinical encounters) are sub-sets of all patients. Note that 48.4% of all patients, 29.8% of all regular attenders and 78.2% of all infrequent attenders did not have any prescription recorded during 1 July 2019 to 30 June 2020.

[a] Excludes under co-payment prescriptions. These accounted for another 96,374,178 prescriptions but these are not reported according to ATC class. Approximately 304 million prescriptions were dispensed if under co-payment prescriptions are also counted.

[b] Excludes prescriptions that do not have an ATC code: All patients – 1,822,602 prescriptions, Regular attenders – 1,753,807 prescriptions, Infrequent attenders – 68,795 prescriptions.

[c] Excludes 125,822 prescriptions that do not have an ATC code and are designated as 'unless otherwise classified'.

Preliminary analysis of duplicate patients (uniqueness plausibility)

NPS MedicineWise commissioned a proof-of-concept project with the Centre for Data Linkage at Curtin University to provide privacy preserving (bloom filter) record linkage services among 150 general practice sites that use the INCA extraction tool to contribute data to MedicineInsight. The output of this project was the creation of a single linkage map, identifying the same individuals within and between practices. Of the 441 general practices included in this report, 112 practices were also included in the proof-of-concept bloom filter project.

Among the 3.2 million patients in the FY 2019/20 cohort, who had at least one clinical encounter from 1 July 2018 to 30 June 2020, 958,641 (30.3%) patients were included in the proof-of-concept bloom filter project and could therefore be included in the duplicate patient analysis (Figure 20). The sociodemographic characteristics of the 30.3% subset of patients included in this preliminary analysis (Appendix B Table B4) were similar to that of all patients in the FY 2019/20 cohort (Appendix B Table B1).

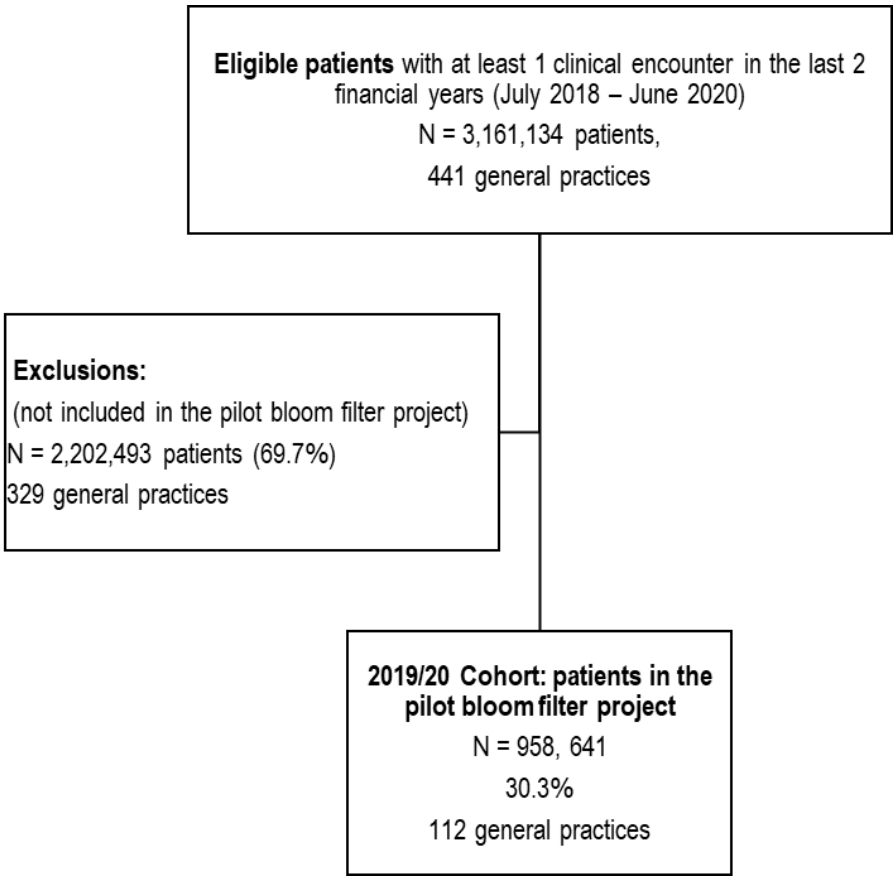

FIGURE 20: STUDY SELECTION FLOWCHART FOR THE 2019/20 COHORT DUPLICATE PATIENT ANALYSIS

Among the 958,641 patients in the duplicate patient analysis 922,370 (96.2%) were considered unique patients and 36,271 (3.8%) were identified as duplicate patients, matched to either one other patient (3.6%) or more than one other patient (0.2%) (Figure 21). Most duplicate patients (n = 34,674) were only matched to one other patient, and of these 10% were identified as the same patient within practice sites and 90% were identified as the same patient between practice sites (Figure 21).

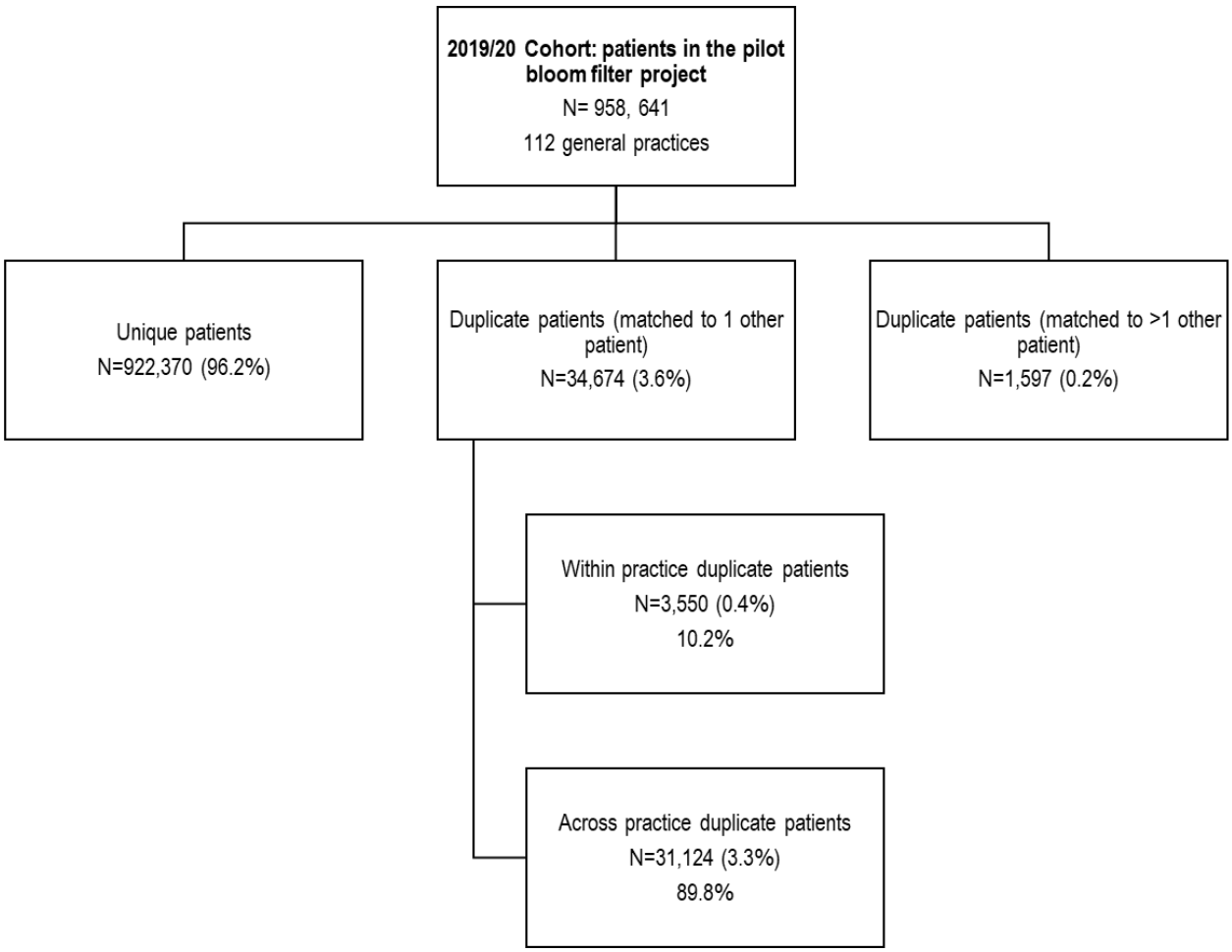

FIGURE 21: RESULTS FLOWCHART FOR THE 2019/20 COHORT DUPLICATE PATIENT ANALYSIS

As a preliminary assessment of the quality of the privacy preserving linkage, the concordance (or agreement) of a selection of patient characteristics between patients who were matched to at least one other patient was measured (Table 34). There were 17,337 pairs of patients who were matched to one other patient and of these 96.8% were concordant for year of birth, 98.7% concordant for gender and 90.8% concordant for the presence or absence of a diagnosis of hypertension. Lower concordance was observed for postcode, at 71.1%. This is not surprising, as patients who visit more than one practice may have changed their address throughout the year.

TABLE 34: PRELIMINARY QUALITY ASSESSMENT OF THE BLOOM FILTER LINKAGE MATCHES; CONCORDANCE OF KEY PATIENT CHARACTERISTICS AMONG 17,337 DUPLICATE PATIENT PAIRS

Patient characteristic	Duplicate patient pairs with matching records (n = 17,337)
	Concordance, %
Year of birth	96.8%
Gender	98.7%
Postcode	71.1%
Hypertension	90.8%

These findings align with the previous analyses conducted by NPS MedicineWise in the series of General Practice Insights Reports,[27,28] where patient loyalty data, in combination with estimates of the proportion of practices in MedicineInsight, were used to estimate the likely number of duplicate patients in MedicineInsight. Patient loyalty data provided by the Department of Health indicate that 63% of all patients attend only one practice. Another 26% attend two practices and 11% attend three or more practices (data on file, Australian Government Department of Health). Assuming no change in patient behaviour, we estimated that 2.2% of patients who visited a MedicineInsight practice at least once in FY 2018/19[28] and 3.0% of patients who visited a MedicineInsight practice at least once in FY 2017/18,[27] had two or more unique patient ID numbers, due to visiting more than one MedicineInsight practice site.

RECOMMENDATIONS

- Consider measures to improve data entry at the:
  - general practice level (including data quality feedback reports, training modules for practice staff, incentives)
  - CIS vendor level (including making fields mandatory, enabling edit checks during data entry), with a focus on the following key variables: reason for prescription, reason for encounter, visit type, Aboriginal and Torres Strait Islander status, smoking status and smoking ceased date (for ex-smokers), year of death (for deceased patients), alcohol use, BMI, CVD risk score and allergy/adverse events.
- Further research is required to understand whether high completeness rates for some variables at the practice level could be used as a marker of better data quality overall.
- When identifying conditions, it is advisable to use information from the diagnosis/medical history, reason for encounter and reason for prescription fields – both coded and free text.
- Consider targeted recruitment of new practices to the MedicineInsight program to improve geographical under-representation among PHNs with less than 3.5% coverage (ie, > 2% lower coverage than the national MedicineInsight average), including Western Queensland (0.0%), Adelaide (1.9%), South Western Sydney (1.9%), Western NSW (< 2.0%), Country SA (< 2.0%), Western Sydney (2.4%), South Eastern Melbourne (2.5%), Central and Eastern Sydney (2.6%), East Melbourne (3.0%) and North Sydney (3.1%).
- Depending on the research question, consider using the regular attender cohort, while acknowledging the potential impact of selection bias and generalisability on study estimates.
- In the Australian setting, where patients can attend multiple general practices, when estimating condition prevalence it is prudent to attempt to exclude temporary and visitor patients, whose medical history might not be recorded, to avoid underestimates. Acknowledge the potential to overestimate condition prevalence when using a regular attender cohort, as regular attenders are more likely to have complete records, be older, and have more chronic conditions than infrequent attenders. Infrequent attenders are generally younger and healthier patients who visit less frequently and are less likely to have a regular GP. Depending on the research question, consider age and sex standardisation of the regular attender cohort to the Australian population or MBS patient population.
- Conduct future validation studies focusing on other domains of data quality such as the accuracy of the information recorded in MedicineInsight.
- Consider linkage of MedicineInsight to other datasets to identify duplicate patients and to improve the capture of data with low completeness rates or from care that occurs outside the MedicineInsight practice, such as: the Australian Institute of Health and Welfare's national death index to improve death recording, PBS data to capture prescriptions from outside the MedicineInsight practice (other GPs and specialists), MBS data to capture MBS billing and test

orders outside the MedicineInsight practice, and hospital data/registries, to more accurately identify serious health outcomes and hospital episodes.

GUIDE TO INTERPRETING THE DATA

When interpreting the information presented in this report, readers should note the following caveats and/or assumptions related to the MedicineInsight data:

- MedicineInsight data are dependent on the accuracy and completeness of data recorded in, and available for extraction from, the general practice clinical systems.
- Completeness of the data is dependent on the various stages at which the data may have been cleaned or cleansed eg, in the CIS, data warehouse and during data analysis. For example, variables such as gender and date of birth have 100% completeness rates because they were used in the selection of patient cohorts. Also, variables derived by NPS MedicineWise will typically have 100% completeness rates.
- The information in this report represents completeness of data recorded in fields accessible to MedicineInsight and may not indicate non-recording of data. It is possible that some GPs may record information in different places within the CIS, for example in the progress notes (which are not available to MedicineInsight), and this can affect completeness rates in MedicineInsight data.
- Identification of conditions is dependent on GPs recording these items in their CIS. Conditions may be under-reported in MedicineInsight data depending on GPs' recording practices.
- Calculation of the relative proportion of different indications assumes that non-recording of conditions occurs at random.
- Identification of risk factor information is dependent on whether this information has been recorded in fields from which data can be extracted and analysed.
- Due to confidentiality issues, NPS MedicineWise does not have access to progress notes, which may contain further information on symptoms, family history, reasons for encounters, diagnoses and test results.
- Patients are free to visit multiple other practices. We do not have data on patients from non-MedicineInsight clinics. Currently we cannot identify patients who have attended multiple MedicineInsight practices.

REFERENCES

1. University of Melbourne. GRHANITE Health Informatics Unit. Melbourne: University of Melbourne, 2020. http://www.grhanite.com/ (accessed 11 April 2020).
2. Precedence Health Care. INCA. Melbourne: Precedence Health Care, 2020. https://precedencehealthcare.com/inca/ (accessed 15 June 2020).
3. Havard A, Manski-Nankervis JA, Thistlethwaite J, et al. Validity of algorithms for identifying five chronic conditions in MedicineInsight, an Australian national general practice database. BMC Health Serv Res 2021;21:551. https://www.ncbi.nlm.nih.gov/pubmed/34090424.
4. Australian Bureau of Statistics. ABS Data Quality Framework. Canberra, ACT: ABS, 2009. https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1520.0Main%20Features1May%202009?opendocument&tabname=Summary&prodno=1520.0&issue=May%202009&num=&view= (accessed 30 June 2020).
5. Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS (Wash DC) 2016;4:1244. https://www.ncbi.nlm.nih.gov/pubmed/27713905.
6. Busingye D, Gianacas C, Pollack A, et al. Data Resource Profile: MedicineInsight, an Australian national primary health care database. Int J Epidemiol 2019;48:1741-h. https://www.ncbi.nlm.nih.gov/pubmed/31292616.
7. The Royal Australian College of General Practitioners. Improving health record quality in general practice: How to create and maintain health records that are fit for purpose. East Melbourne: RACGP, 2018. https://www.racgp.org.au/FSDEDEV/media/documents/Running%20a%20practice/Practice%20resources/Improving-health-record-quality-in-general-practice.pdf (accessed 30/06/2020).
8. Australian Bureau of Statistics. National Health Survey: First Results, 2017-18 (Catalogue No. 4364.0.55.001). Canberra: ABS, 2018. http://www.abs.gov.au/ausstats/abs@.nsf/mf/4364.0.55.001 (accessed 22 February 2019).
9. Department of Health. PBS Expenditure and Prescriptions Report 1 July 2019 to 30 June 2020. Canberra, 2020. https://www.pbs.gov.au/statistics/expenditure-prescriptions/2019-2020/PBS_Expenditure_and_Prescriptions_Report_1-July-2019_to_30-June-2020.pdf (accessed 17 December 2020).
10. Cheng FW, Gao X, Mitchell DC, et al. Body mass index and all-cause mortality among older adults. Obesity (Silver Spring) 2016;24:2232-9. https://www.ncbi.nlm.nih.gov/pubmed/27570944.
11. Li W, Kelsey JL, Zhang Z, et al. Small-area estimation and prioritizing communities for obesity control in Massachusetts. Am J Public Health 2009;99:511-9. https://www.ncbi.nlm.nih.gov/pubmed/19150906.
12. Littman AJ, Boyko EJ, McDonell MB, et al. Evaluation of a weight management program for veterans. Prev Chronic Dis 2012;9:E99. https://www.ncbi.nlm.nih.gov/pubmed/22595323.
13. Australian Bureau of Statistics. Australian Statistical Geography Standard (ASGS): Volume 5 - Remoteness Structure, July 2016. Canberra: ABS, 2018. http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1270.0.55.005Main%20Features1Ju

ly%202016?opendocument&tabname=Summary&prodno=1270.0.55.005&issue=July%202016&num=&view= (accessed 7 November).

14. Australian Bureau of Statistics. 2033.0.55.001 – Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016. Canberra: ABS, 2018. https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001 (accessed 12 August 2020).
15. Nicholson BD, Aveyard P, Bankhead CR, et al. Determinants and extent of weight recording in UK primary care: an analysis of 5 million adults' electronic health records from 2000 to 2017. BMC Med 2019;17:222. https://www.ncbi.nlm.nih.gov/pubmed/31783757.
16. Royal Australian College of General Practitioners. Guidelines for preventive activities in general practice 9th edition (Red Book). East Melbourne: RACGP, 2018. https://www.racgp.org.au/clinical-resources/clinical-guidelines/key-racgp-guidelines/view-all-racgp-guidelines/red-book (accessed 24 February 2020).
17. The Australian Government Department of Health. Bone Densitometry. Canberra: Department of Health, 2017. https://www1.health.gov.au/internet/main/publishing.nsf/Content/diagnosticimaging-bd.htm (accessed 23 December 2020).
18. Australian Bureau of Statistics. Census of Population and Housing – Counts of Aboriginal and Torres Strait Islander Australians, 2016 Canberra, 2017. http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/2075.0Main+Features12016?OpenDocument (accessed 15 December 2020).
19. Department of Social Services. DSS Payment Demographic Data June 2019. Canberra, 2019. https://data.gov.au/data/dataset/dss-payment-demographic-data (accessed 5 March 2020).
20. Department of Veterans' Affairs. DVA Pensioner Summary – June 2019. Canberra, 2019. https://www.dva.gov.au/sites/default/files/files/publications/datastatistical/pensionsummary/pensumm_jun2019.pdf (accessed 5 March 2020).
21. Australian Bureau of Statistics. 3218.0 – Regional Population Growth, Australia, 2017–18. Canberra: ABS, 2019. https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3218.02017-18?OpenDocument (accessed 15 December 2020).
22. Australian Bureau of Statistics. Disability, Ageing and Carers, Australia: Summary of Findings, 2018 https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4430.02018?OpenDocument (accessed 17 December 2020).
23. National Diabetes Service Scheme. The Australian Diabetes Map. Canberra: Diabetes Australia, 2020. https://map.ndss.com.au/#!/ (accessed 21 December 2020).
24. Harrison C, Britt H, Miller G, et al. Prevalence of chronic conditions in Australia. PLoS One 2013;8:e67494. https://www.ncbi.nlm.nih.gov/pubmed/23935834.
25. Sundararajan K, Schoeman T, Hughes L, et al. Predictors and outcomes of acute pancreatitis in critically ill patients presenting to the emergency department of a tertiary referral centre in Australia. Emerg Med Australas 2017;29:184-91. https://www.ncbi.nlm.nih.gov/pubmed/28125855.

26. Australian Institute of Health and Welfare. Deaths by suicide over time. Canberra: AIHW, 2020. https://www.aihw.gov.au/suicide-self-harm-monitoring/data/deaths-by-suicide-in-australia/suicide-deaths-over-time (accessed 21 December 2020).
27. NPS MedicineWise. General Practice Insights Report July 2017–June 2018. Sydney: NPS MedicineWise, 2019. https://www.nps.org.au/assets/NPS/pdf/General-Practice-Insights-Report_2017-18.pdf (accessed 25 February 2020).
28. NPS MedicineWise. General Practice Insights Report July 2018–June 2019. Sydney: NPS MedicineWise, 2020. https://www.nps.org.au/assets/Report-2018-19-GPIR-final-signed-off-by-DOH.pdf (accessed 23 December 2020).

APPENDIX A: DEFINITIONS

Completeness definitions

Table A1: Completeness definitions for selected variables

Data table	Variable name	Completeness definition
Patient	ASGS_RA_NAME_2011	Value is not blank
	ASGS_RA_NAME_2016	Value is not blank
	ATSI_NAME	Value is not blank, and **does not** equal any of the following; • 'Not recorded' • 'Not stated/inadequately described'
	CIS_PATIENT_STATUS_NAME	Value is not blank and **does not** equal 'Not recorded'
	CLIN_REGULAR_ATTENDER_FLAG	Value is not blank
	DECEASED_INDICATOR	Value is not blank
	GENDER_NAME	Value is not blank, and **does not** equal any of the following: • 'Not recorded' • 'Not stated/inadequately described'
	PATIENT_AGE	Value is not blank
	PATIENT_CITY	Value is not blank
	PATIENT_POSTCODE	Value is not blank, and contains only numbers, and is one of the following; • 4 digits, with the first digit 1 to 9 (inclusive) • 4 digits, with the first two digits 02, 08 or 09 • 3 digits, with the first digit 2, 8 or 9
	PENSION_CODE	Value is not blank and does not equal 'Not recorded'
	PHN_CODE	Value is not blank, and **does not** equal any of the following; • 'Unknown' • 'Undefined'
	SMOKING_CEASED_DATE *(among Ex-Smokers only)*	Value is not blank, and **does not** equal any of the following system default dates: • 01JAN1800 • 01JAN1900

Data table	Variable name	Completeness definition
	SMOKING_STATUS_NAME	Value is not blank, and does not equal any of the following: • 'Not recorded' • 'Not stated/inadequately described'
	VALID_PATIENT_FLAG	Value is not blank and is one of the following: • 0 (No) • 1 (Yes)
	YEAR_OF_BIRTH	Value is not blank, and is between 1898 and 2020 (inclusive)
	YEAR_OF_DEATH	Value is not blank, and is between 1898 and 2020 (inclusive)
Encounter	PROVIDER_ID	Value is not blank
	VISIT_DATE	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	VISIT_TYPE	Value is not blank
	VISIT_TYPE_MD *(Derived variable for this analysis)*	Where SOURCE_SYSTEM contains 'MD', and Value is not blank
	VISIT_TYPE_BP *(Derived variable for this analysis)*	Where SOURCE_SYSTEM contains 'BP', and Value is not blank
	ENCOUNTER_REASON	When linked to the ENCOUNTER_REASON table by PATIENT_NUMBER and ENCOUNTER_ID, at least one record has an ENCOUNTER_REASON value that is not blank.
	IS_CLINICAL	Value is one of the following: • 0 (No) • 1 (Yes)
	PROVIDER_IS_DR_OR_NURSE	Value is one of the following: • 0 (No) • 1 (Yes)
Encounter reason	VISIT_DATETIME	Date value is not blank, and **does not equal any of the following**: • 01JAN1800 • 01JAN1900
	ENCOUNTER_REASON_CODE	Value is not blank, and **does not** equal '0'
	ENCOUNTER_REASON	Value is not blank

Data table	Variable name	Completeness definition
Diagnosis	ADMIN_FLAG	Value is one of the following: • 0 (No) • 1 (Yes)
	DIAGNOSIS_REASON	Value is not blank
	DIAGNOSIS_STATUS_ACTIVE_FLAG	Value is not blank
	DIAGNOSIS_TYPE	Value is not blank
	CREATED_DATETIME	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	DIAGNOSIS_DATE	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	CIS_CODED_STATUS	Value is not blank
	DIAGNOSIS_REASON_CODE	Value is not blank, and **does not** equal '0'
	DIFFERENTIAL_FLAG	Value is not blank
Pathology results header	RESULT_NAME	Value is not blank
	COLLECTION_DATE	Date value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	REPORT_DATE	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	IMPORT_DATE	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	COMPLETION_FLAG	Value is not blank
	NORMAL_FLAG	Value is not blank
Requested investigations	REQUESTED_TESTS	Value is not blank
	TEST_REASON	Value is not blank
	REQUEST_DATE	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	BILLING	Value is not blank
	COPIES	Value is not blank

Data table	Variable name	Completeness definition
Pathology results detail	PATHOLOGY_RESULT_ID	Value is not blank
	RESULT_DATE	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	DATA_TYPE	Value is not blank
	LOINC_CODE	Value is not blank
	RESULT_NAME	Value is not blank
	RESULT_VALUE	Value is not blank
	UNITS	Value is not blank
	NORMAL_RANGE	Value is not blank
	ABNORMAL_FLAG	Value is not blank
	RECORD_STATUS	Value is not blank
	CREATED_DATETIME	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	UPDATED_DATETIME	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
Medicine history	CREATED_BY	Value is not blank
	FIRST_DATE	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	LAST_DATE	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	MEDICINE_ACTIVE_INGREDIENT	Value is not blank
	MEDICINE_NAME	Value is not blank
	PRODUCT_NAME	Value is not blank
	ATC_CODE	Value is not blank
	DOSE	Value is not blank
	STRENGTH	Value is not blank
	FORM	Value is not blank
	ROUTE	Value is not blank
	QUANTITY	Value is not blank
	FREQUENCY	Value is not blank

Data table	Variable name	Completeness definition
	INSTRUCTIONS	Value is not blank
	REPEAT_INTERVAL	Value is not blank
	REPEATS	Value is not blank
	PBS_STATUS	Value is not blank, and **does not** equal 'TBD'
	RESTRICTION_CODE	Value is not blank
	AUTHORITY_INDICATION	Value is not blank
	PREVIOUS_AUTHORITY	Value is not blank
	REASON	Value is not blank
	REASON_CODE	Value is not blank
	IS_CURRENT	Value is one of the following: • 0 (No) • 1 (Yes)
	RX_STATUS_LIMITED_MEDICATION	Value is not blank
	RECORD_STATUS	Value is not blank
Prescription issued	SCRIPT_DATE	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	MEDICINE_ACTIVE_INGREDIENT	Value is not blank
	MEDICINE_NAME	Value is not blank
	ATC_CODE	Value is not blank
	DOSE	Value is not blank
	ROUTE	Value is not blank
	STRENGTH	Value is not blank
	FREQUENCY	Value is not blank
	QUANTITY	Value is not blank
	INSTRUCTIONS	Value is not blank
	PRN	Value is one of the following: • 0 (No) • 1 (Yes)
	REGULATION_24	Value is not blank
	REPEATS	Value is not blank
	REPEAT_INTERVAL	Value is not blank
	RESTRICTION_CODE	Value is not blank
	AUTHORITY_INDICATION	Value is not blank
	AUTHORITY_INDICATION_CODE_A_OR_X *(Derived variable for this analysis)*	AUTHORITY_INDICATION value is not blank, and RESTRICTION_CODE equals 'A' or 'X'

Data table	Variable name	Completeness definition
MBS billing	ITEM_NUMBER	Value is not blank
	SERVICE_PATIENT_COUNT	Value is not blank
	SERVICE_RECORD_STATUS	Value is not blank
	CREATED_DATETIME	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	SERVICE_DATETIME	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	UPDATED_DATETIME	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
	VISIT_DATETIME	Value is not blank, and **does not** equal any of the following: • 01JAN1800 • 01JAN1900
Site	STATE	Value is not blank, and Equals one of the following: • WA • SA • NT • QLD • NSW • VIC • ACT • TAS
	PHN_CODE	Value is not blank, and **does not** equal any of the following: • 'UNKNOWN' • 'UNDEFINED'
	MULTI_PRACTICE_FLAG	Value is not blank
	ASGS_RA_CODE_2016	Value is not blank
Practice recruitment	PRACTICE_CATEGORY	Value is not blank
	PRACTICE_NUMBER_OF_GP	Value is not blank
	PRACTICE_POSTCODE	Value is not blank, and contains only numbers, and is one of the following: • 4 digits, with the first digit 1 to 9 (inclusive) • 4 digits, with the first two digits 02, 08 or 09 • 3 digits, with the first digit 2, 8 or 9

Data table	Variable name	Completeness definition
	PRACTICE_STATE	Value is not blank, and equals one of the following: • WA • SA • NT • QLD • NSW • VIC • ACT • TAS
	PRACTICE_SUBURB	Value is not blank
Clinical user	CLINICAL_USER_TYPE_NAME	Value is not blank, and **does not** equal 'Unknown type'
	DOCTOR_INDICATOR	Value is not blank, and **does not** equal 'Unknown'
	NURSE_INDICATOR	Value is not blank, and **does not** equal 'Unknown'

Table A2. Completeness definitions for patient characteristics

Characteristic	Definition
Patient has a clinical encounter in the last FY	Patient has at least one record in EMI_ENCOUNTER table where **VISIT_DATE** is between 01JUL2019 and 30JUN2020 and IS_CLINICAL=1
Patient has an encounter reason in the last FY	Patient has at least one record in EMI_ENCOUNTER_REASON table where **VISIT_DATETIME** is between 01JUL2019 and 30JUN2020 and ENCOUNTER_REASON is not blank
Patient has a diagnosis in the last FY	Patient has at least one record in EMI_DIAGNOSIS table where **DIAGNOSIS_DATE** is between 01JUL2019 and 30JUN2020 and DIAGNOSIS_REASON is not blank
Patient has a prescription issued in the last FY	Patient has at least one record in EMI_SCRIPT_ITEM table where **SCRIPT_DATE** is between 01JUL2019 and 30JUN2020
Patient has a pathology results header record in the last FY	Patient has at least one record in EMI_PATHOLOGY table where **REPORT_DATE** is between 01JUL2019 and 30JUN2020
Patient has an MBS billing record in the last FY	Patient has at least one record in EMI_BILLING_SERVICE table where **SERVICE_DATETIME** is between 01JUL2019 and 30JUN2020
BMI recorded in the last 2 years (1 July 2018 to 30 June 2020)	Patient has at least one record in EMI_OBSERVATION table where Observation_Type is 'BMI' and the observation_value is not missing and observation_datetime is between 01JUL2018 and 30JUN2020 OR Patient has at least one record in EMI_OBSERVATION table where Observation_Type is 'HEIGHT' and the observation_value is not missing and observation_datetime is between 01JUL2018 and 30JUN2020 AND at least one record in EMI_OBSERVATION where Observation_Type is 'WEIGHT'

	and the observation_value is not missing and observation_datetime is between 01JUL2018 and 30JUN2020
Smoking status recorded as at date of download (ever)	The patient record in EMI_PATIENT table has a value in their SMOKING_STATUS_NAME field of 'Smoker' 'Non Smoker' or 'Ex Smoker'
Alcohol status recorded as at date of download	Patient has at least one record in EMI_ALCOHOL_STATUS table where DRINKS_PER_DAY or PAST_DRINKS_PER_DAY is not missing
Blood pressure among patients aged 40+ years recorded in the last 2 years (1 July 2018 to 30 June 2020)	Patient has at least one record in EMI_OBSERVATION table where Observation_Type is BLOOD PRESSU' or 'BP' and the observation_value is not missing and observation_datetime is between 01JUL2018 and 30JUN2020 OR Patient has at least one record in EMI_OBSERVATION table where Observation_Type is 'SYSTOLIC' or 'SYSSTAND' or 'SYSSLIE' and the observation_value is not missing and observation_datetime is between 01JUL2018 and 30JUN2020 AND at least one record in EMI_OBSERVATION table where Observation_Type is 'DIASTOLIC' or 'DIASSTAND' or 'DIASLIE' and the observation_value is not missing and observation_datetime is between 01JUL2018 and 30JUN2020 AND a diastolic and systolic result are on the same day
BMD among patients aged 50+ years as at 30 June 2020 (ever recorded)	Patient has at least one record in any of the following locations [table (variable)]: • EMI_OBSERVATION (Observation_Type) • EMI_PATHOLOGY_RESULT_ATOM (Result_Name) • EMI_ENCOUNTER_REASON (Encounter_Reason) • EMI_REQUESTED_TEST (Requested_Tests) • EMI_DIAGNOSIS (Diagnosis_Reason) That contain any of the following terms: • *BONE MINERAL DENSITY* • *BMD* • *BONE DENSITOMETRY REPORT* • *DUAL ENERGY X-RAY ABSORPTIOMETRY* • *DXA* • *T-SCORE* • *DEXA* • *Z-SCORE* • *DUAL ENERGY X-RAY ABSORPTIOMETRY* • *BONE DENSITY* • *BONE MASS* And the relevant date variable is before or on 30JUN2020
CVD risk score among patients aged 40+ years as at 30 June 2020 (ever recorded)	Patient has at least one record in EMI_OBSERVATION table where the OBSERVATION_TYPE is 'ACVRISK' or 'CVRISK' and the cleaned OBSERVATION_VALUE is a number, and OBSERVATION_DATETIME is before or on 30JUN2020
eGFR among patients aged 40+ years recorded in the last 2 years (1 July 2018 to 30 June 2020)	Patient has at least one record in EMI_PATHOLOGY_RESULT_EGFR table where PATHOLOGY_TEST_DATETIME is between 01JUL2018 and 30JUN2020 and the RESULT_TRANSMISSION_TYPE = 'HL7'
Total cholesterol among patients aged 40+ years recorded in the last 2 years (1 July 2018 to 30 June 2020)	Patient has at least one record in EMI_PATHOLOGY_RESULT_LIPIDS table where PATHOLOGY_TEST_DATETIME is between 01JUL2018 and

	30JUN2020 and PATHOLOGY_TEST_TYPE = 'Total cholesterol measurement'
HbA_{1c} recorded in the last 2 years (1 July 2018 to 30 June 2020)	Patient has at least one record in EMI_PATHOLOGY_RESULT_HBA1C table where PATHOLOGY_TEST_DATETIME is between 01JUL2018 and 30JUN2020
Spirometry recorded as at 30 June 2020 (ever recorded)	Patient has both a POSTFEV1 and POSTFVC record with the same date in EMI_OBSERVATION table, or a single POSTFEV1FVC record, where OBSERVATION_DATETIME is before or on 30JUN2020
Allergy/adverse events as at 30 June 2020 (ever recorded)	Patient has any record in the EMI_ALLERGY_REACTION table where the ALLERGY_ITEM_NAME is not blank, and CREATED_DATETIME is before or on 30JUN2020. Records such as 'no known allergies' were considered not missing

Conditions

In MedicineInsight, patients are defined as having a condition if they had a relevant coded (Docle, Pyefinch) or free text entry recorded in one of the three diagnosis fields ('Diagnosis/medical history field ', 'Reason for visit' or 'Reason for prescription' fields).

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 actually refers to the condition indicated or is present in another context (eg, a search for 'cancer' may identify 'partner died from cancer'). Each record is flagged accordingly.

Relevant terms used to identify the included condition are shown in Table A3.

Table A3. Definitions for conditions

Conditions	Definition
Acute pancreatitis	Includes: acute pancreatitis, pancreatitis acute
Asthma	Includes: allergic asthma, allergy induced asthma, asthma, asthma action plan, asthma care plan, asthma cycle of care, asthma exacerbation, asthma review, exercise induced asthma, exertional asthma, occupational asthma, Samter's triad or thunderstorm asthma. Excludes (when recorded in isolation): wheezy bronchitis
Anxiety	Includes: adjustment disorder with anxiety, adjustment disorder with mixed anxiety and depressed mood, anxiety, anxiety (generalised or neurosis or phobia or PTSD or social), anxiety disorder, anxiety with panic attacks, anxiety/depression, depressive anxiety disorder, GAD, generalised anxiety disorder, mixed anxiety depression, nervous anxiety, neurotic anxiety, phobic anxiety disorder, social anxiety disorder, social phobia or substance induced anxiety disorder. Excludes (when recorded in isolation): anxiety feeling, adjustment disorder, (parental or performance or separation) anxiety, neurosis, OCD, PTSD, phobias or panic disorders
Atrial fibrillation	Includes: AF, A FIB, atrial f, atrial fibrillation, atrial fibrillation (isolated episode or paroxysmal or ablation or non-valvular or valvular), fibrillation or rapid atrial fibrillation
Breast cancer	Includes: breast (adenocarcinoma or cancer or carcinoma), breast ca, (colloid or intraductal or lobular) carcinoma, DCIS, disseminated peritoneal adenocarcinoma, ductal carcinoma(in situ or infiltrating), infiltrating lobular carcinoma of breast, lobular ca, lobular carcinoma in-situ, mammary carcinoma, mucinous cystadenocarcinoma, Paget's disease

	of breast, peritoneal mucinous carcinomatosis, pseudomyxoma peritonei or signet ring cell carcinoma of breast
Cancer	Includes: terms for all types of cancer or carcinoma or tumours, except skin cancer or melanoma
Cardiovascular disease	Includes: atherosclerosis, coronary heart disease, peripheral vascular disease, stroke and transient ischaemic attack.
Chronic kidney disease	Includes: anaemia - chronic renal failure, capd, catheterisation of peritoneum, chronic kidney disease or CKD (all stages), chronic renal disease (all stages), chronic renal failure, chronic renal failure – hyperparathyroidism, chronic renal insufficiency, continuous ambulatory peritoneal dialysis, CRF, dialysis, haemodialysis, hemodialysis, peritoneal catherisation for dialysis, peritoneal dialysis renal dialysis or surgery - abdomen - dialysis - catheterisation
Chronic obstructive pulmonary disease	Includes: acute exacerbation of copd, cal, chronic airways limitation, chronic bronchitis, chronic obstructive airways disease, chronic obstructive pulmonary disease, coad, copd, emphysema
Dementia	Includes: alzh, alzheimer disease, behavioural and psychological symptoms of dementia, binswanger (disease or encephalopathy), demen, dementia, (early onset or frontotemporal or jakob creutzfeldt or korsakoff or lewy-body or multi infarct or pick or semantic or subcortical or substance-induced or vascular or young onset) dementia, major neurocognitive disorder due to alzheimer disease, parkinson disease with lewy body dementia, psychosis (korsakoff or dementia related), senile dementia with psychosis, subcortical arteriosclerotic encephalopathy
Depression	Includes: adjustment disorder with depressed +/- anxious mood, anxiety/depression, depres, depression, (endogenous or major or melancholic or minor or non melancholic or organic or postnatal or psychotic or reactive or recurrent or subsyndromal) depression or depressive disorder or depressive episode, melancholia
Dermatitis/eczema	Includes (allergic or asteatotic or atopic or chronic or contact or discoid or dyshidrotic or exfoliative or infantile or infected or nummular or varicose or venous) eczema, atopic dermatitis, autoeczematisation, dyshidrosis, eczema, eczema craquele, flexural eczema, gravitational eczema, pompholyx, pompholyx eczema, psoriatic eczema
Diabetes (gestational)	Includes: gestational (diabetes or diabetes mellitus)
Diabetes (type 1)	Includes: diabetes mellitus (iddm or type I or type 1), iddm, insulin dependent diabetes mellitus, juvenile onset diabetes
Diabetes (type 2/unspecified)	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
Dyslipidaemia	Includes: dyslipidaemia, dyslip, familial (hypercholesterolaemia or hypercholesterolemia), hdl, high cholesterol, high cholest, high lipids, hypercholesterolaemia, hyperlipidaemia, hyperlipoproteinaemia (type 2 or type iv or type iia), hypertriglyceridaemia, hypercho, hyperlip, hypertr
Gastroesophageal reflux disease	Includes: acid reflux, acid regurgitation, gastro-oesophageal reflux, gor, gord, heartburn, laryngopharyngeal reflux, non-erosive reflux disease, oesophageal reflux, reflux laringitis, reflux oesophagitis
Heart failure	Includes: acute cardiac failure, biventricular heart failure, cardiac failure, CCF, chronic heart failure, congestive cardiac failure, congestive heart failure, cor pulmonale, diastolic cardiac dysfunction, diastolic heart failure, heart failure, HFmrEF, HFpEF, HFrEF, Hhgh output cardiac failure, high output heart failure, hypertensive heart failure, left heart failure,

	left ventricular failure, LHF (left heart failure), LVF (left ventricular failure), pulmonary oedema, RHF (right heart failure), right heart failure, right ventricular failure, RVF (right ventricular failure), systolic cardiac dysfunction, systolic heart failure, ventricular diastolic dysfunction
Hypertension	Includes: antihypertensive agent prescription, (blood pressure or bp) and (labile or review or unstable), hbp, high blood pressure, ht, hypertension, hypertension (controlled or diastolic or essential or isolated systolic or labile or lifestyle management or malignant or pregnancy or primary or renal or renovascular or review or unstable), pih, pregnancy induced hypertension or severe refractory hypertension
Low back pain	Includes: back (ache or injury or muscle strain or pain or spasm or strain), back and buttock pain, back and leg pain, back pain, back pain (acute or acute on chronic or buttock or degenerative spine or leg or lumbar or lumbo-sacral or sacral or radiating to buttock or radiating to leg), back pain syndrome, back pain with (radiculopathy or referred leg pain) back pain without leg pain, degenerative lumbar disc disease, foraminal stenosis, lumbar, intervertebral disc prolapse , disc prolapse, nerve root compression, loin pain, low back injury, low back pain, low back strain, lumbago, lumbar back (injury or muscle strain or pain or prolapse), lumbar (radiculopathy or spondylosis or lumbar sprain), lumbosacral back pain, lumbosacral spondylosis, lumbosacral stenosis, mechanical back pain, mechanical low back pain, mononeuropathy - sciatic nerve, sacral spinal pain, sacro-iliac joint pain, sciatic (mononeuropathy or pain), sciatica, spinal disc protrusion, spinal pain, strained back
Lower respiratory tract infection	Includes: bronchiolitis, cavitating tuberculosis in lungs, cavitating tuberculosis of lung, chest airway infection, chest infection, chest inf, croup, croup (cough or spasmodic or viral), pleural tuberculosis, TB, tuberculosis, tuberculosis of the lung, viral lower respiratory tract infection, viral LRTI, all bronchitis and pneumonia terms
Migraine	Patients were defined as having migraine, if they had a relevant coded (Docle, Pyefinch) or free text entry in one of the three diagnosis fields recorded at any time from the patient's earliest record up to the download date. Relevant terms include: antimigraine prescription, botox treatment for migraine, cluster headache, migraine, migraine aura, migraineur or vascular headache
Myocardial infarction	Includes: myocardial infarction, myocardial infarct, MI, acute myocardial infarction, AMI, heart attack, myocardial damage, non-st-elevation myocardial infarction, NSTEMI, st elevation myocardial infarction, STEMI, subendocardial infarct
Osteoarthritis	Includes: aneurysm-osteoarthritis syndrome, ankylosing spondylitis, generalised osteoarthritis, oa, osteoarthritis, osteoarthritis (ankle or cervical spine or elbow or fingers or foot or glenohumeral joint or hands or hip or knee or lumbar spine or midfoot or neck or patellofemoral joint or sacroiliac joints or shoulder or spine or sternoclavicular joint or thoracic spine or tmj or wrist or 1st carpometacarpal joint or osteoarthritis of 1st metatarsophalangeal joint), osteoarthrosis (hip or knee), spondylosis, wear and tear arthritis
Osteoporosis	Includes: osteoporosis, osteoporosis (corticosteroid induced or no fracture or with fracture or disuse or steroid induced), pathological fracture due to osteoporosis, post menopausal osteoporosis, steroid osteopathy
Otitis media	Includes: otitis media, middle ear infection, bullous myringitis, viral myringitis, ear effusion
Prostate cancer	Includes: prostate or prostatic (adenocarcinoma or ca or cancer or carcinoma or carcinosarcoma), (family history or FH) of prostate cancer, signet ring cell carcinoma of prostate
Rheumatoid arthritis	Includes: arthritis (juvenile rheumatoid or rheumatoid or seronegative), caplan syndrome, jra, lipoid dermatoarthritis, lipoid rheumatism, multicentric reticulohistiocytosis, RA, rheumatoid arthritis – pneumoconiosis, seronegative rheumatoid arthritis, stills disease

Stroke	Includes: cerebral (haemorrhage or infarction), cerebrovascular accident, cva, haemorrhage intracerebral, haemorrhagic (cva or stroke), intracerebral (bleed or haemorrhage or haemorrhage), ischaemic stroke, lacunar infarct, lacunar stroke, migrainous stroke, migranous stroke, stroke, thrombotic stroke, visual cortex stroke
Suicide/self-harm	Includes: suicide, suicidal ideation, suicidal thoughts, thinking of suicide, attempted suicide, suicide ideas, suicidal tendencies, self-harm, self-mutilation
Upper respiratory tract infection	Includes: acute pharyngitis, bacterial pharyngitis, (infection or inflammation) of pharynx or larynx, laryngitis, pain in throat, pharyngitis, rhinitis, sore throat, throat pain, upper respiratory congestion, upper respiratory tract congestion, upper respiratory tract infection or urti
Urinary tract infection	Includes: bladder infection, cystitis, pyuria, recurrent urinary tract infection, recurrent UTI, urinary tract infection, UTI, UTI recurrent, U.T.I, urinary tract infect, urine - pus
Venous thromboembolism (ie DVT and/or PE)	Includes: deep venous thrombosis, DVT, thrombosis deep vein, pulmonary embolism, saddle pulmonary embolism, pulmonary emb, PE, VTE

DVT = deep vein thrombosis, PE = pulmonary embolism

Prescriptions

Prescription data are restricted to medicines prescribed by GPs using their CIS to print the prescription. 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. Prescription data do not necessarily indicate whether a medicine was dispensed or used by the patient.

There are two potential sources of information about prescriptions in MedicineInsight – Medicine history table and Prescription issued table – which are all linked to the patient. The Medicine history table contains details of medications prescribed to a patient and may also include medicines prescribed by specialists, hospital etc. The Prescription issued table contains details of individual prescriptions issued or printed for patients by their GP.

Prescription data are available for issued prescriptions and a stated number of repeats recorded in the CIS. To provide context, when a prescription for a medicine with five repeats is entered in the CIS it will be counted once when the analysis focuses on issued prescriptions, and will be counted six times when the analysis is for the issued-plus-repeat prescriptions, which we refer to here as the total prescriptions. Completeness of selected variables from the two prescriptions tables was assessed and total prescriptions were compared with PBS data. For the plausibility analysis, we assessed total prescriptions from the Prescription issued table.

Pathology

Most Australian practices receive pathology test results electronically, transferred directly into the CIS from pathology providers. There are three potential sources of pathology information within the CIS – tests requested, result summaries and the associated result details – which are all linked to the patient. The Requested investigations table contains details of any medical tests requested for a patient and does not contain any test results, but only the details provided on the test request form.

The result summaries (Pathology results header table) contains general details (a header record) of pathology results received by a practice. The result details (Pathology results detail table) contains result values from specific pathology tests. The result summaries and result details also include data from tests ordered by specialists or doctors outside the practice, when they have requested that a GP receive a copy of a result.

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. Completeness of selected variables from the three pathology tables was assessed.

Details of other data tables are provided in Table A4.

Table A4: Other MedicineInsight data tables

Data table [TABLE NAME]	Description	Data fields examples
CONDITIONS [EMI_CONDITIONS_DETAIL] [EMI_CONDITIONS_SUMMARY]	Derived tables. Identifies specific conditions (eg, asthma, diabetes, etc) documented in any of the Diagnosis, Encounter Reason or Prescription tables.	Condition 1 Condition 2 Condition 3
ALLERGIES/REACTIONS [EMI_ALLERGY_REACTION]	Allergies and adverse reactions.	Date recorded Allergy substance Reaction type
IMMUNISATIONS [EMI_IMMUNISATION]	Vaccine and immunisation details.	Vaccine name Date given Batch number Sequence number
OBSERVATIONS [EMI_OBSERVATION]	Observations recorded about the patient. eg, blood pressure, height, weight, BMI, temperature, blood sugar etc.	Observation date Observation type Observation value

Table A5: Plausible values used in MedicineInsight data portal

Observation	Source system	Observation type	Minimum value	Maximum value
Body height	BP	Height	24.0	251.0
Body height	MD	HEIGHT	24.0	251.0
Body height	BP_PHC	Height	24	251
Body height	MD_PHC	HEIGHT	24	251
Body weight	BP	Weight	0.2	610.0
Body weight	MD	WEIGHT	0.2	610.0
Body weight	BP_PHC	Weight	0.2	610.0
Body weight	MD_PHC	WEIGHT	0.2	610.0
Body mass index	BP	BMI	7.50	204.00
Body mass index	MD	BMI	7.50	204.00
Body mass index	BP_PHC	BMI	7.50	204.00
Body mass index	MD_PHC	BMI	7.50	204.00
Sitting systolic blood pressure	BP	Systolic	50	250
Sitting systolic blood pressure	MD	SYSTOLIC	50	250
Standing systolic blood pressure	BP	Systolic	50	250
Standing systolic blood pressure	MD	SYSSTAND	50	250
Lying systolic blood pressure	BP	Systolic	50	250
Lying systolic blood pressure	MD	SYSLIE	50	250
Sitting diastolic blood pressure	BP	Diastolic	30	140
Sitting diastolic blood pressure	MD	DIASTOLIC	30	140
Standing diastolic blood pressure	BP	Diastolic	30	140
Standing diastolic blood pressure	MD	DIASSTAND	30	140
Lying diastolic blood pressure	BP	Diastolic	30	140
Lying diastolic blood pressure	MD	DIASLIE	30	140
Sitting systolic blood pressure	BP_PHC	Systolic	50	250
Standing systolic blood pressure	BP_PHC	Systolic	50	250
Lying systolic blood pressure	BP_PHC	Systolic	50	250
Sitting diastolic blood pressure	BP_PHC	Diastolic	30	140
Standing diastolic blood pressure	BP_PHC	Diastolic	30	140
Lying diastolic blood pressure	BP_PHC	Diastolic	30	140
Sitting systolic blood pressure	MD_PHC	SYSTOLIC	50	250
Standing systolic blood pressure	MD_PHC	SYSSTAND	50	250
Lying systolic blood pressure	MD_PHC	SYSLIE	50	250
Sitting diastolic blood pressure	MD_PHC	DIASTOLIC	30	140
Standing diastolic blood pressure	MD_PHC	DIASSTAND	30	140
Lying diastolic blood pressure	MD_PHC	DIASLIE	30	140

Note that these plausible values cover all age groups.

APPENDIX B: COHORT CHARACTERISTICS

The tables in this appendix show the sociodemographic characteristics of all patients eligible for inclusion in this study as well as regular and infrequent attenders, and the subset of all patients in the preliminary duplicate patient analysis.

Table B1: Sociodemographic characteristics of 'all patient' cohorts for each financial year

Characteristic	Number (%)									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
Total patient count	**1,854,173**	**2,317,003**	**2,470,769**	**2,598,641**	**2,715,010**	**2,853,751**	**2,988,493**	**3,096,365**	**3,173,076**	**3,161,134**
Gender										
Female	1,018,767 (54.9%)	1,264,500 (54.6%)	1,346,185 (54.5%)	1,416,061 (54.5%)	1,479,103 (54.5%)	1,552,661 (54.4%)	1,624,915 (54.4%)	1,682,255 (54.3%)	1,722,546 (54.3%)	1,712,892 (54.2%)
Male	835,361 45.1%)	1,052,447 (45.4%)	1,124,517 (45.5%)	1,182,510 (45.5%)	1,235,818 (45.5%)	1,300,996 (45.6%)	1,363,450 (45.6%)	1,413,924 (45.7%)	1,450,268 (45.7%)	1,447,775 (45.8%)
Intersex/indeterminate	45 (0.0%)	56 (0.0%)	67 (0.0%)	70 (0.0%)	89 (0.0%)	94 (0.0%)	128 (0.0%)	186 (0.0%)	262 (0.0%)	467 (0.0%)
Age group (years)										
0–9	254,629 (13.7%)	319,693 (13.8%)	341,923 (13.8%)	361,211 (13.9%)	379,632 (14.0%)	398,054 (13.9%)	413,689 (13.8%)	425,258 (13.7%)	435,764 (13.7%)	424,582 (13.4%)
10–19	210,297 (11.3%)	254,845 (11.0%)	268,156 (10.9%)	276,742 (10.6%)	286,695 (10.6%)	303,559 (10.6%)	318,877 (10.7%)	330,715 (10.7%)	339,002 (10.7%)	334,920 (10.6%)
20–29	256,190 (13.8%)	344,358 (14.9%)	372,159 (15.1%)	392,658 (15.1%)	410,170 (15.1%)	432,535 (15.2%)	452,410 (15.1%)	466,237 (15.1%)	475,033 (15.0%)	463,717 (14.7%)
30–39	257,855 (13.9%)	328,872 (14.2%)	349,261 (14.1%)	368,960 (14.2%)	387,397 (14.3%)	409,680 (14.4%)	433,201 (14.5%)	454,574 (14.7%)	472,266 (14.9%)	473,330 (15.0%)
40–49	257,810 (13.9%)	316,980 (13.7%)	335,887 (13.6%)	349,032 (13.4%)	359,899 (13.3%)	373,769 (13.1%)	387,636 (13.0%)	399,487 (12.9%)	404,824 (12.8%)	401,073 (12.7%)
50–59	234,214 (12.6%)	284,907 (12.3%)	303,160 (12.3%)	318,310 (12.2%)	331,025 (12.2%)	345,369 (12.1%)	358,478 (12.0%)	366,944 (11.9%)	372,255 (11.7%)	372,430 (11.8%)
60–69	190,920 (10.3%)	231,945 (10.0%)	249,280 (10.1%)	265,227 (10.2%)	278,970 (10.3%)	293,786 (10.3%)	306,064 (10.2%)	316,372 (10.2%)	323,716 (10.2%)	327,526 (10.4%)
70–79	115,634 (6.24%)	138,288 (6.0%)	148,238 (6.0%)	158,480 (6.1%)	169,075 (6.2%)	180,193 (6.3%)	195,771 (6.6%)	210,214 (6.8%)	220,685 (7.0%)	230,769 (7.3%)
80–89	64,454 (3.5%)	79,460 (3.4%)	83,380 (3.4%)	86,836 (3.3%)	89,517 (3.3%)	92,750 (3.3%)	96,765 (3.2%)	99,758 (3.2%)	101,806 (3.2%)	104,450 (3.3%)
90+	12,170 (0.7%)	17,655 (0.8%)	19,325 (0.8%)	21,185 (0.8%)	22,630 (0.8%)	24,056 (0.8%)	25,602 (0.9%)	26,806 (0.9%)	27,725 (0.9%)	28,337 (0.9%)
Remoteness										
Major city	1,101,162 (59.4%)	1,380,912 (59.6%)	1,473,826 (59.7%)	1,566,030 (60.3%)	1,649,086 (60.7%)	1,746,787 (61.2%)	1,846,400 (61.8%)	1,934,455 (62.5%)	2,008,044 (63.3%)	2,017,266 (63.8%)
Inner regional	445,782 (24.0%)	557,954 (24.1%)	611,666 (24.8%)	644,718 (24.8%)	672,412 (24.8%)	698,565 (24.5%)	718,064 (24.0%)	726,900 (23.5%)	728,215 (22.9%)	718,358 (22.7%)
Outer regional	263,240 (14.2%)	319,872 (13.8%)	325,239 (13.2%)	327,426 (12.6%)	330,232 (12.2%)	336,970 (11.8%)	346,480 (11.6%)	356,089 (11.5%)	358,174 (11.3%)	350,417 (11.1%)
Remote or very remote	26,763 (1.4%)	34,632 (1.5%)	36,052 (1.5%)	36,431 (1.4%)	39,301 (1.5%)	47,117 (1.7%)	52,721 (1.8%)	54,103 (1.8%)	53,709 (1.7%)	51,413 (1.6%)
Indigenous status										
Aboriginal and/or Torres Strait Islander	35,683 (1.9%)	45,403 (2.0%)	52,885 (2.1%)	60,381 (2.3%)	67,473 (2.5%)	75,662 (2.7%)	83,611 (2.8%)	90,107 (2.9%)	94,978 (3.0%)	95,478 (3.0%)
Other Australian	1,184,683 (63.9%)	1,453,466 (62.7%)	1,604,568 (64.9%)	1,750,198 (67.4%)	1,905,605 (70.2%)	2,070,204 (72.5%)	2,211,837 (74.0%)	2,321,995 (75.0%)	2,396,996 (75.5%)	2,389,564 (75.6%)
Not known	633,807 (34.2%)	818,134 (35.3%)	813,316 (32.9%)	788,062 (30.3%)	741,932 (27.3%)	707,885 (24.8%)	693,045 (23.2%)	684,263 (22.1%)	681,102 (21.5%)	676,092 (21.4%)
State/territory										
Australian Capital Territory	20,113 (1.1%)	26,980 (1.2%)	33,114 (1.3%)	37,902 (1.5%)	43,381 (1.6%)	49,028 (1.7%)	50,643 (1.7%)	51,153 (1.7%)	50,993 (1.6%)	49,425 (1.6%)

Characteristic	Number (%)									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
New South Wales	666,535 (35.9%)	823,553 (35.5%)	885,169 (35.8%)	942,371 (36.3%)	980,609 (36.1%)	1,022,021 (35.8%)	1,074,381 (36.0%)	1,116,889 (36.1%)	1,150,468 (36.3%)	1,152,714 (36.5%)
Northern Territory	53,584 (2.9%)	67,236 (2.9%)	61,828 (2.5%)	56,354 (2.2%)	56,942 (2.1%)	60,789 (2.1%)	64,403 (2.2%)	66,760 (2.2%)	65,369 (2.1%)	61,782 (2.0%)
Queensland	302,865 (16.3%)	385,253 (16.6%)	401,449 (16.2%)	417,845 (16.1%)	441,473 (16.3%)	475,352 (16.7%)	509,544 (17.1%)	549,104 (17.7%)	572,507 (18.0%)	579,878 (18.3%)
South Australia	42,747 (2.31%)	49,645 (2.1%)	51,453 (2.1%)	52,390 (2.0%)	51,670 (1.9%)	54,659 (1.9%)	59,705 (2.0%)	66,109 (2.1%)	70,160 (2.2%)	71,150 (2.3%)
Tasmania	108,658 (5.9%)	124,083 (5.4%)	131,065 (5.3%)	135,593 (5.2%)	144,453 (5.3%)	151,548 (5.3%)	157,232 (5.3%)	163,825 (5.3%)	167,094 (5.3%)	171,914 (5.4%)
Victoria	433,574 (23.4%)	556,272 (24.0%)	595,435 (24.1%)	631,318 (24.3%)	659,858 (24.3%)	684,446 (24.0%)	707,812 (23.7%)	719,829 (23.2%)	730,099 (23.0%)	705,037 (22.3%)
Western Australia	226,079 (12.2%)	283,953 (12.3%)	311,222 (12.6%)	324,820 (12.5%)	336,575 (12.4%)	355,865 (12.5%)	364,741 (12.2%)	362,666 (11.7%)	366,349 (11.5%)	369,189 (11.7%)
Other territories	18 (0.0%)	28 (0.0%)	34 (0.0%)	48 (0.0%)	49 (0.0%)	43 (0.0%)	32 (0.0%)	30 (0.0%)	37 (0.0%)	45 (0.0%)
Socioeconomic status (SEIFA quintile)										
1 (least advantaged)	325,755 (17.6%)	400,063 (17.3%)	425,935 (17.2%)	443,778 (17.1%)	454,620 (16.7%)	471,922 (16.5%)	487,669 (16.3%)	503,347 (16.3%)	508,896 (16.0%)	497,371 (15.7%)
2	369,889 (19.9%)	464,994 (20.1%)	507,182 (20.5%)	535,693 (20.6%)	553,122 (20.4%)	571,600 (20.0%)	589,137 (19.7%)	599,312 (19.4%)	606,738 (19.1%)	603,316 (19.1%)
3	378,162 (20.4%)	473,972 (20.5%)	503,979 (20.4%)	533,043 (20.5%)	559,139 (20.6%)	596,142 (20.9%)	634,651 (21.2%)	663,083 (21.4%)	682,213 (21.5%)	681,772 (21.6%)
4	320,845 (17.3%)	408,485 (17.6%)	437,419 (17.7%)	468,574 (18.0%)	504,674 (18.6%)	548,428 (19.2%)	585,333 (19.6%)	614,203 (19.8%)	636,057 (20.0%)	637,164 (20.2%)
5 (most advantaged)	442,333 (23.9%)	545,910 (23.6%)	572,333 (23.2%)	593,562 (22.8%)	619,499 (22.8%)	641,365 (22.5%)	666,888 (22.3%)	691,616 (22.3%)	714,248 (22.5%)	717,834 (22.7%)

FY = financial year, SEIFA = Socioeconomic Index for Areas. Missing are included in the denominator for calculating the proportions.

Table B2: Sociodemographic characteristics of 'regular attenders' cohort for each financial year

Characteristic	Number (%)									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
Total patient count	**1,220,576**	**1,375,787**	**1,483,697**	**1,572,564**	**1,649,365**	**1,734,080**	**1,810,354**	**1,875,033**	**1,930,420**	**1,949,196**
Gender										
Female	695,958 (57.0%)	782,661 (56.9%)	841,530 (56.7%)	891,386 (56.7%)	933,036 (56.6%)	978,688 (56.4%)	1,020,874 (56.4%)	1,056,053 (56.3%)	1,086,049 (56.3%)	1,093,669 (56.1%)
Male	524,592 (43.0%)	593,089 (43.1%)	642,122 (43.3%)	681,129 (43.3%)	716,275 (43.4%)	755,328 (43.6%)	789,407 (43.6%)	818,873 (43.7%)	844,217 (43.7%)	855,299 (43.9%)
Intersex or indeterminate	26 (0.0%)	37 (0.0%)	45 (0.0%)	49 (0.0%)	54 (0.0%)	64 (0.0%)	73 (0.0%)	107 (0.0%)	154 (0.0%)	228 (0.0%)
Age group (years)										
0–9	153,743 (12.6%)	172,539 (12.5%)	187,244 (12.6%)	198,430 (12.6%)	210,214 (12.7%)	221,672 (12.8%)	228,884 (12.6%)	236,767 (12.6%)	246,673 (12.8%)	245,454 (12.6%)
10–19	114,749 (9.4%)	126,567 (9.2%)	135,623 (9.1%)	142,448 (9.1%)	149,133 (9.0%)	160,039 (9.2%)	167,984 (9.3%)	173,625 (9.3%)	179,111 (9.3%)	180,068 (9.2%)
20–29	143,316 (11.7%)	167,524 (12.2%)	185,205 (12.5%)	198,424 (12.6%)	208,960 (12.7%)	221,140 (12.8%)	229,281 (12.7%)	234,874 (12.5%)	238,710 (12.4%)	236,445 (12.1%)
30–39	157,152 (12.9%)	177,532 (12.9%)	191,166 (12.9%)	204,757 (13.0%)	216,602 (13.1%)	230,070 (13.3%)	242,038 (13.4%)	253,037 (13.5%)	263,164 (13.6%)	266,600 (13.7%)
40–49	169,875 (13.9%)	189,396 (13.8%)	202,607 (13.7%)	212,863 (13.5%)	220,555 (13.4%)	228,869 (13.2%)	236,537 (13.1%)	243,867 (13.0%)	248,399 (12.9%)	248,134 (12.7%)
50–59	168,810 (13.8%)	187,764 (13.6%)	200,489 (13.5%)	210,980 (13.4%)	219,520 (13.3%)	228,151 (13.2%)	236,234 (13.0%)	241,793 (12.9%)	245,970 (12.7%)	247,407 (12.7%)
60–69	148,582 (12.2%)	166,617 (12.1%)	180,035 (12.1%)	191,343 (12.2%)	200,212 (12.1%)	209,317 (12.1%)	217,725 (12.0%)	224,696 (12.0%)	230,528 (11.9%)	235,345 (12.1%)
70–79	97,069 (8.0%)	108,154 (7.9%)	116,699 (7.9%)	124,390 (7.9%)	131,955 (8.0%)	139,550 (8.1%)	151,657 (8.4%)	162,787 (8.7%)	171,675 (8.9%)	180,640 (9.3%)
80–89	56,525 (4.6%)	65,299 (4.8%)	68,811 (4.6%)	71,595 (4.6%)	73,676 (4.5%)	75,650 (4.4%)	79,160 (4.4%)	81,799 (4.4%)	83,652 (4.3%)	86,089 (4.2%)
90+	10,755 (0.9%)	14,395 (1.1%)	15,818 (1.1%)	17,334 (1.1%)	18,538 (1.1%)	19,622 (1.1%)	20,854 (1.2%)	21,788 (1.2%)	22,538 (1.2%)	23,014 (1.2%)
Remoteness										
Major City	713,612 (58.5%)	804,399 (58.5%)	870,195 (58.7%)	928,270 (59.0%)	979,344 (59.4%)	1,038,073 (59.9%)	1,095,485 (60.5%)	1,150,962 (61.4%)	1,200,429 (62.2%)	1,226,229 (62.9%)
Inner Regional	309,091 (25.3%)	351,812 (25.6%)	386,680 (26.1%)	411,045 (26.1%)	430,164 (26.1%)	446,566 (25.8%)	457,397 (25.3%)	461,853 (24.6%)	465,226 (24.1%)	463,479 (23.8%)
Outer Regional	172,976 (14.2%)	190,575 (13.9%)	196,566 (13.2%)	201,802 (12.8%)	207,145 (12.6%)	212,639 (12.3%)	217,168 (12.0%)	221,575 (11.8%)	224,719 (11.6%)	221,135 (11.3%)
Remote or Very Remote	15,876 (1.3%)	18,837 (1.4%)	19,675 (1.3%)	20,543 (1.3%)	21,839 (1.3%)	25,888 (1.5%)	29,495 (1.6%)	29,941 (1.6%)	29,215 (1.5%)	27,988 (1.4%)
Indigenous status										
Aboriginal and/or Torres Strait Islander	23,515 (1.9%)	28,327 (2.1%)	33,361 (2.3%)	37,973 (2.4%)	42,450 (2.6%)	47,195 (2.72%)	51,609 (2.9%)	55,307 (3.0%)	58,542 (3.0%)	59,523 (3.1%)
Other Australian	843,027 (69.1%)	964,432 (70.1%)	1,064,905 (71.8%)	1,153,635 (73.4%)	1,241,997 (75.3%)	1,336,782 (77.1%)	1,417,205 (78.3%)	1,481,536 (79.0%)	1,530,483 (79.3%)	1,546,100 (79.3%)
Not known	354,034 (29.0%)	383,028 (27.8%)	385,431 (26.0%)	380,956 (24.2%)	364,918 (22.1%)	350,103 (20.2%)	341,540 (18.9%)	338,190 (18.0%)	341,395 (17.7%)	343,573 (17.6%)
State/territory										
Australian Capital Territory	12,843 (1.1%)	14,521 (1.1%)	17,565 (1.2%)	20,151 (1.3%)	23,084 (1.4%)	26,764 (1.5%)	28,487 (1.57%)	29,394 (1.6%)	29,305 (1.5%)	29,234 (1.5%)
New South Wales	448,437 (36.7%)	502,310 (36.5%)	546,067 (36.8%)	582,797 (37.1%)	606,643 (36.8%)	630,372 (36.4%)	660,186 (36.5%)	688,058 (36.7%)	713,382 (37.0%)	727,062 (37.3%)
Northern Territory	29,417 (2.4%)	31,976 (2.3%)	29,826 (2.0%)	28,623 (1.8%)	30,568 (1.9%)	33,448 (1.9%)	36,024 (2.0%)	37,252 (2.0%)	36,833 (1.9%)	34,598 (1.8%)
Queensland	194,544 (15.9%)	222,675 (16.2%)	234,343 (15.8%)	245,050 (15.6%)	260,208 (15.8%)	280,485 (16.2%)	299,126 (16.5%)	322,143 (17.2%)	338,029 (17.5%)	349,954 (18.0%)
South Australia	29,565 (2.4%)	31,259 (2.3%)	32,294 (2.2%)	32,994 (2.1%)	32,853 (2.0%)	34,238 (2.0%)	36,762 (2.0%)	40,447 (2.2%)	43,658 (2.3%)	46,263 (2.4%)
Tasmania	80,867 (6.6%)	86,916 (6.3%)	90,591 (6.1%)	94,157 (6.0%)	97,444 (5.9%)	100,526 (5.8%)	103,848 (5.7%)	108,164 (5.8%)	113,238 (5.9%)	115,821 (5.9%)

Characteristic	Number (%)									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
Victoria	276,342 (22.6%)	318,952 (23.2%)	348,731 (23.5%)	375,104 (23.9%)	397,957 (24.1%)	416,277 (24.0%)	427,573 (23.6%)	430,121 (22.9%)	433,917 (22.5%)	420,967 (21.6%)
Western Australia	148,555 (12.2%)	167,166 (12.2%)	184,263 (12.4%)	193,669 (12.3%)	200,593 (12.2%)	211,956 (12.2%)	218,336 (12.1%)	219,443 (11.7%)	222,049 (11.5%)	225,288 (11.6%)
Other territories	6 (0.0%)	12 (0.0%)	17 (0.0%)	19 (0.0%)	15 (0.0%)	14 (0.0%)	12 (0.0%)	11 (0.0%)	9 (0.0%)	9 (0.0%)
Socioeconomic status (SEIFA quintile)										
1 (least advantaged)	223,556 (18.3%)	250,678 (18.2%)	268,823 (18.1%)	281,101 (17.9%)	290,722 (17.6%)	300,438 (17.3%)	308,745 (17.1%)	317,040 (16.9%)	321,602 (16.7%)	318,628 (16.3%)
2	247,100 (20.2%)	282,783 (20.6%)	311,868 (21.0%)	332,646 (21.2%)	343,762 (20.8%)	354,124 (20.4%)	362,546 (20.0%)	367,807 (19.6%)	374,307 (19.4%)	378,119 (19.4%)
3	250,695 (20.5%)	283,713 (20.6%)	304,632 (20.5%)	325,204 (20.7%)	343,713 (20.8%)	367,634 (21.2%)	391,060 (21.6%)	410,334 (21.9%)	425,044 (22.0%)	431,020 (22.1%)
4	206,391 (16.9%)	235,462 (17.1%)	255,727 (17.2%)	275,151 (17.5%)	296,961 (18.0%)	323,260 (18.6%)	344,274 (19.0%)	360,756 (19.2%)	375,795 (19.5%)	381,920 (19.6%)
5 (most advantaged)	283,833 (23.3%)	313,008 (22.8%)	332,079 (22.4%)	347,561 (22.1%)	363,328 (22.0%)	377,706 (21.8%)	392,913 (21.7%)	408,394 (21.8%)	422,839 (21.9%)	429,144 (22.0%)

FY = financial year, SEIFA = socioeconomic index for areas. Missing are included in the denominator for calculating the proportions.

Table B3: Sociodemographic characteristics of 'infrequent attenders' cohort for each financial year

Characteristic	Number (%)									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
Total patient count	**633,597**	**941,216**	**987,072**	**1,026,077**	**1,065,645**	**1,119,671**	**1,178,139**	**1,221,332**	**1,242,656**	**1,211,938**
Gender										
Female	322,809 (50.9%)	481,839 (51.2%)	504,655 (51.1%)	524,675 (51.1%)	546,067 (51.2%)	573,973 (51.3%)	604,041 (51.3%)	626,202 (51.3%)	636,497 (51.2%)	619,223 (51.1%)
Male	310,769 (49.0%)	459,358 (48.8%)	482,395 (48.9%)	501,381 (48.9%)	519,543 (48.8%)	545,668 (48.7%)	574,043 (48.7%)	595,051 (48.7%)	606,051 (48.8%)	592,476 (48.9%)
Intersex/indeterminate	19 (0.0%)	19 (0.0%)	22 (0.0%)	21 (0.0%)	35 (0.0%)	30 (0.0%)	55 (0.0%)	79 (0.0%)	108 (0.0%)	239 (0.0%)
Age group (years)										
0–9	100,886 (15.9%)	147,154 (15.6%)	154,679 (15.7%)	162,781 (15.9%)	169,418 (15.9%)	176,382 (15.8%)	184,805 (15.7%)	188,491 (15.4%)	189,091 (15.2%)	179,128 (14.8%)
10–19	95,548 (15.1%)	128,278 (13.6%)	132,533 (13.4%)	134,294 (13.1%)	137,562 (12.9%)	143,520 (12.8%)	150,893 (12.8%)	157,090 (12.9%)	159,891 (12.9%)	154,852 (12.8%)
20–29	112,874 (17.8%)	176,834 (18.8%)	186,954 (18.9%)	194,234 (18.9%)	201,210 (18.9%)	211,395 (18.9%)	223,129 (18.9%)	231,363 (18.9%)	236,323 (19.0%)	227,272 (18.8%)
30–39	100,703 (15.9%)	151,340 (16.1%)	158,095 (16.0%)	164,203 (16.0%)	170,795 (16.0%)	179,610 (16.0%)	191,163 (16.2%)	201,537 (16.5%)	209,102 (16.8%)	206,730 (17.1%)
40–49	87,935 (13.9%)	127,584 (13.6%)	133,280 (13.5%)	136,169 (13.3%)	139,344 (13.1%)	144,900 (12.9%)	151,099 (12.8%)	155,620 (12.7%)	156,425 (12.6%)	152,939 (12.6%)
50–59	65,404 (10.3%)	97,143 (10.3%)	102,671 (10.4%)	107,330 (10.5%)	111,505 (10.5%)	117,218 (10.5%)	122,244 (10.4%)	125,151 (10.2%)	126,285 (10.2%)	125,023 (10.3%)
60–69	42,338 (6.7%)	65,328 (6.9%)	69,245 (7.0%)	73,884 (7.2%)	78,758 (7.4%)	84,469 (7.5%)	88,339 (7.5%)	91,676 (7.51%)	93,188 (7.5%)	92,181 (7.6%)
70–79	18,565 (2.9%)	30,134 (3.2%)	31,539 (3.2%)	34,090 (3.3%)	37,120 (3.5%)	40,643 (3.6%)	44,114 (3.7%)	47,427 (3.9%)	49,010 (3.9%)	50,129 (4.1%)
80–89	7,929 (1.3%)	14,161 (1.5%)	14,569 (1.5%)	15,241 (1.5%)	15,841 (1.5%)	17,100 (1.5%)	17,605 (1.5%)	17,959 (1.5%)	18,154 (1.5%)	18,361 (1.5%)
90+	1,415 (0.2%)	3,260 (0.3%)	3,507 (0.4%)	3,851 (0.4%)	4,092 (0.4%)	4,434 (0.4%)	4,748 (0.4%)	5,018 (0.41%)	5,187 (0.4%)	5,323 (0.4%)
Remoteness										
Major city	387,550 (61.2%)	576,513 (61.3%)	603,631 (61.2%)	637,760 (62.2%)	669,742 (62.8%)	708,714 (63.3%)	750,915 (63.7%)	783,493 (64.2%)	807,615 (65.0%)	791,037 (65.3%)
Inner regional	136,691 (21.6%)	206,142 (21.9%)	224,986 (22.8%)	233,673 (22.8%)	242,248 (22.7%)	251,999 (22.5%)	260,667 (22.1%)	265,047 (21.7%)	262,989 (21.2%)	254,879 (21.0%)
Outer regional	90,264 (14.2%)	129,297 (13.7%)	128,673 (13.0%)	125,624 (12.2%)	123,087 (11.6%)	124,331 (11.1%)	129,312 (11.0%)	134,514 (11.0%)	133,455 (10.7%)	129,282 (10.7%)
Remote or very remote	10,887 (1.7%)	15,795 (1.7%)	16,377 (1.7%)	15,888 (1.6%)	17,462 (1.6%)	21,229 (1.9%)	23,226 (2.0%)	24,162 (2.0%)	24,494 (2.0%)	23,425 (1.9%)
Indigenous status										
Aboriginal and/or Torres Strait Islander	12,168 (1.9%)	17,076 (1.8%)	19,524 (2.0%)	22,408 (2.2%)	25,023 (2.4%)	28,467 (2.549%)	32,002 (2.729%)	34,800 (2.9%)	36,436 (3.0%)	35,955 (3.0%)
Other Australian	341,656 (53.9%)	489,034 (52.0%)	539,663 (54.7%)	596,563 (58.1%)	663,608 (62.3%)	733,422 (65.5%)	794,632 (67.4%)	840,459 (68.8%)	866,513 (69.7%)	843,464 (69.6%)
Not known	279,773 (44.2%)	435,106 (46.2%)	427,885 (43.3%)	407,106 (39.7%)	377,014 (35.4%)	357,782 (32.0%)	351,505 (29.8%)	346,073 (28.3%)	339,707 (27.3%)	332,519 (27.4%)
State/territory										
Australian Capital Territory	7,270 (1.2%)	12,459 (1.3%)	15,549 (1.58%)	17,751 (1.7%)	20,297 (1.9%)	22,264 (2.0%)	22,156 (1.9%)	21,759 (1.8%)	21,688 (1.8%)	20,191 (1.7%)
New South Wales	218,098 (34.4%)	321,243 (34.1%)	339,102 (34.4%)	359,574 (35.0%)	373,966 (35.1%)	391,649 (35.0%)	414,195 (35.2%)	428,831 (35.1%)	437,086 (35.2%)	425,652 (35.1%)
Northern Territory	24,167 (3.8%)	35,260 (3.8%)	32,002 (3.2%)	27,731 (2.7%)	26,374 (2.5%)	27,341 (2.4%)	28,379 (2.4%)	29,508 (2.4%)	28,536 (2.3%)	27,184 (2.2%)
Queensland	108,321 (17.1%)	162,578 (17.3%)	167,106 (16.9%)	172,795 (16.8%)	181,265 (17.0%)	194,867 (17.4%)	210,418 (17.9%)	226,961 (18.6%)	234,478 (18.9%)	229,924 (19.0%)
South Australia	13,182 (2.1%)	18,386 (2.0%)	19,159 (1.9%)	19,396 (1.9%)	18,817 (1.8%)	20,421 (1.8%)	22,943 (2.0%)	25,662 (2.1%)	26,502 (2.1%)	24,887 (2.1%)
Tasmania	27,791 (4.4%)	37,167 (4.0%)	40,474 (4.1%)	41,436 (4.0%)	47,009 (4.4%)	51,022 (4.6%)	53,384 (4.5%)	55,661 (4.6%)	53,856 (4.3%)	56,093 (4.6%)

Characteristic	Number (%)									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
Victoria	157,232 (24.8%)	237,320 (25.2%)	246,704 (25.0%)	256,214 (25.0%)	261,901 (24.6%)	268,169 (24.0%)	280,239 (23.8%)	289,708 (23.7%)	296,182 (23.8%)	284,070 (23.4%)
Western Australia	77,524 (12.2%)	116,787 (12.4%)	126,959 (12.9%)	131,151 (12.8%)	135,982 (12.8%)	143,909 (12.9%)	146,405 (12.4%)	143,223 (11.7%)	144,300 (11.6%)	143,901 (11.9%)
Other territories	12 (0.0%)	16 (0.0%)	17 (0.0%)	29 (0.0%)	34 (0.0%)	29 (0.0%)	20 (0.0%)	19 (0.0%)	28 (0.0%)	36 (0.0%)
Socioeconomic status (SEIFA quintile)										
1 (least advantaged)	102,199 (16.1%)	149,385 (15.9%)	157,112 (15.9%)	162,677 (15.9%)	163,898 (15.4%)	171,484 (15.3%)	178,924 (15.2%)	186,307 (15.3%)	187,294 (15.1%)	178,743 (14.7%)
2	122,789 (19.4%)	182,211 (19.4%)	195,314 (19.8%)	203,047 (19.8%)	209,360 (19.6%)	217,476 (19.4%)	226,591 (19.2%)	231,505 (19.0%)	232,431 (18.7%)	225,197 (18.6%)
3	127,467 (20.1%)	190,259 (20.2%)	199,347 (20.2%)	207,839 (20.3%)	215,426 (20.2%)	228,508 (20.4%)	243,591 (20.7%)	252,749 (20.7%)	257,169 (20.7%)	250,752 (20.7%)
4	114,454 (18.1%)	173,023 (18.4%)	181,692 (18.4%)	193,423 (18.9%)	207,713 (19.5%)	225,168 (20.1%)	241,059 (20.5%)	253,447 (20.8%)	260,262 (20.9%)	255,244 (21.1%)
5 (most advantaged)	158,500 (25.0%)	232,902 (24.7%)	240,254 (24.3%)	246,001 (24.0%)	256,171 (24.0%)	263,659 (23.5%)	273,975 (23.3%)	283,222 (23.2%)	291,409 (23.5%)	288,690 (23.8%)

FY = financial year, SEIFA = socioeconomic index for areas. Missing are included in the denominator for calculating the proportions.

Table B4: Sociodemographic characteristics of the subset of patients from the 2019/20 cohort included in the preliminary duplicate patient analysis

Characteristic	Number	%
Total patient count	**957,044**	**100**
Gender		
Male	438,748	45.8
Female	518,151	54.1
Intersex/indeterminate	145	0.02
Age group (years)		
0–9	134,408	14.0
10–19	101,887	10.6
20–29	138,533	14.5
30–39	146,423	15.3
40–49	123,123	12.9
50–59	110,021	11.5
60–69	94,865	9.9
70–79	67,665	7.1
80–89	31,006	3.2
90+	9,113	1.0
Remoteness		
Major city	656,785	68.6
Inner regional	166,141	17.4
Outer regional	102,591	10.7
Remote or very remote	25,045	2.6
Indigenous status		
Aboriginal and/or Torres Strait Islander	31,893	3.3
Other Australian	719,864	75.2
Not known	205,287	21.5
State/territory		
Australian Capital Territory	11,633	1.2
New South Wales	312,325	32.6
Northern Territory	35,382	3.7
Queensland	195,234	20.4
South Australia	19,472	2.0
Tasmania	66,888	7.0
Victoria	200,980	21.0
Western Australia	115,130	12.0
Socioeconomic status (SEIFA quintile)		
1 (least advantaged)	115,601	12.1
2	162,382	17.0
3	240,986	25.2
4	206,037	21.5
5 (most advantaged)	225,550	23.6

SEIFA = socioeconomic index for areas. Missing are included in the denominator for calculating the proportions.

APPENDIX C: DENOMINATORS USED FOR ASSESSING COMPLETENESS

Table C1: Denominators used for assessment of completeness for each data table and financial year

Selected data tables	Denominator for each data table and financial year									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
All patients										
Encounter	11,002,046	11,938,593	12,743,244	13,640,330	14,524,211	15,478,926	16,287,842	17,287,264	17,481,570	18,013,469
Encounter reason	7,955,730	9,148,705	10,688,561	12,416,881	13,962,730	15,445,420	16,754,009	17,925,485	18,634,598	19,309,924
Diagnosis	2,671,497	2,901,560	3,078,918	3,463,399	3,754,243	3,947,644	4,420,875	4,631,937	4,510,448	4,369,517
Pathology results header	7,157,544	7,985,566	8,889,988	9,682,845	10,316,712	10,948,107	11,672,135	12,538,023	13,207,873	13,114,692
Pathology results detail	21,736,571	26,817,950	31,685,419	37,579,016	42,465,260	47,234,692	51,761,041	56,501,883	59,843,794	60,247,408
Requested investigations	1,539,750	1,922,970	2,226,917	2,492,326	2,749,378	2,986,755	3,186,042	3,472,256	3,692,551	3,762,024
Medicine history	3,644,165	3,927,103	4,165,173	4,398,747	4,663,213	4,808,448	4,900,439	5,058,466	5,119,768	4,840,823
Prescription issued	7,729,740	8,268,735	8,732,206	9,275,043	9,694,092	9,912,620	9,980,438	10,122,839	10,204,130	9,972,277
MBS billing	6,605,549	8,285,282	10,246,618	11,883,177	13,774,627	15,339,663	16,500,132	18,077,018	19,231,986	20,551,710
Regular attenders										
Encounter	9,957,046	10,794,053	11,602,628	12,467,432	13,276,509	14,169,111	14,974,453	15,938,308	16,222,218	16,833,037
Encounter reason	7,177,311	8,250,967	9,704,025	11,345,805	12,769,634	14,150,995	15,399,400	16,510,454	17,265,664	18,034,194
Diagnosis	2,276,396	2,454,484	2,631,208	2,996,775	3,263,752	3,439,799	3,873,952	4,097,093	4,030,863	3,968,588
Pathology results header	6,602,013	7,353,178	8,246,400	8,996,925	9,601,507	10,214,361	10,964,042	11,819,138	12,538,992	12,541,114
Pathology results detail	20,288,909	25,002,317	29,585,811	35,173,108	39,839,182	44,466,239	48,967,853	53,638,966	57,130,958	57,959,765
Requested investigations	1,390,395	1,734,671	2,025,692	2,274,787	2,513,052	2,742,014	2,951,040	3,228,712	3,456,847	3,548,914
Medicine history	3,175,888	3,408,959	3,639,518	3,857,022	4,089,997	4,230,890	4,334,293	4,502,712	4,596,272	4,411,751
Prescription issued	7,134,961	7,617,235	8,092,653	8,634,867	9,027,624	9,255,885	9,366,938	9,539,095	9,676,218	9,542,835
MBS billing	6,091,618	7,611,954	9,429,747	10,943,723	12,687,667	14,166,190	15,290,135	16,796,853	17,973,185	19,383,232
Infrequent attenders										
Encounter	1,045,000	1,144,540	1,140,616	1,172,898	1,247,702	1,309,815	1,313,389	1,348,956	1,259,352	1,180,432

Selected data tables	Denominator for each data table and financial year									
	FY1	FY2	FY3	FY4	FY5	FY6	FY7	FY8	FY9	FY10
Encounter reason	778,419	897,738	984,536	1,071,076	1,193,096	1,294,425	1,354,609	1,415,031	1,368,934	1,275,730
Diagnosis	395,101	447,076	447,710	466,624	490,491	507,845	546,923	534,844	479,585	400,929
Pathology results header	555,531	632,388	643,588	685,920	715,205	733,746	708,093	718,885	668,881	573,578
Pathology results detail	1,447,662	1,815,633	2,099,608	2,405,908	2,626,078	2,768,453	2,793,188	2,862,917	2,712,836	2,287,643
Requested investigations	149,355	188,299	201,225	217,539	236,326	244,741	235,002	243,544	235,704	213,110
Medicine history	468,277	518,144	525,655	541,725	573,216	577,558	566,146	555,754	523,496	429,072
Prescription issued	594,779	651,500	639,553	640,176	666,468	656,735	613,500	583,744	527,912	429,442
MBS billing	513,931	673,328	816,871	939,454	1,086,960	1,173,473	1,209,997	1,280,165	1,258,801	1,168,478