Electronic health record (EHR)-detectable statin intolerance phenotypes: Prevalence and validation in real-world general practice

*Shagoofa Rakhshanda [1], Joel Rhee [2], Siaw-Teng Liaw [1], Kerry-Anne Rye [3] and Jitendra Jonnagaddala [2,4 *]*

[1] School of Population Health, UNSW Sydney, Kensington, Australia

[2] Discipline of General Practice, School of Clinical Medicine, Kensington, Australia

[3] School of Biomedical Sciences, UNSW Sydney, Kensington, Australia

[4.] SREDH Consortium, Sydney, Australia

* *Corresponding author*

Shagoofa Rakhshanda, s.rakhshanda@unsw.edu.au; ORCID: 0000-0001-9172-7042

Joel Rhee, J.Rhee@unsw.edu.au; ORCID: 0000-0002-5233-2758

Siaw-Teng Liaw, siaw@unsw.edu.au; ORCID: 0000-0001-5989-3614

Kerry-Anne Rye, k.rye@unsw.edu.au; ORCID: 0000-0002-9751-917X

Jitendra Jonnagaddala, jitendra.jonnagaddala@unsw.edu.au; ORCID: 0000-0002-9912-2344

Word count in abstract: 248

Word count in full text: 4,992 (excluding references, figures and tables)

Number of figures: 4

Number of tables: 3

ABSTRACT

Aims: This study focused on patients who were prescribed statins as primary prevention of cardiovascular diseases. This study aimed to identify statin intolerant patients and determine the prevalence of statin intolerance by implementing electronic health record (EHR)-detectable statin intolerance electronic phenotyping algorithms, and to validate these algorithms.

Methods: This study used the Electronic Practice Based Research Network (ePBRN) dataset. The methodology took place in four stages: (1) literature review to identify electronic phenotypes, (2) implementation of electronic phenotypes on ePBRN, (3) development and implementation of reference standard, (4) validation of electronic phenotypes.

Results: Six EHR-detectable statin intolerance electronic phenotypes were identified, including the Minnesota Combined Rule-Based algorithm, Japan-Statin induced myopathy (SIMs), USA-SIMs, Singapore-SIMs (algorithms A, B, C, and D), Japan-Statin-associated muscle toxicity (SAMT), and NHS-UK-Statin intolerance pathway. The prevalence of statin intolerance among those prescribed statins in ePBRN was 5.09%. The Singapore SIMs-B algorithm showed the highest accuracy (57.05%), sensitivity (92.95%), negative predictive value (43.43%), and F1 (71.51%) scores, while the Japan SAMT algorithm showed the highest specificity (99.13%), positive predictive value (76.19%), and correlation coefficient (0.05%).

Conclusion: The prevalence of statin intolerance in ePBRN is at the low end of the 5–15% range reported in Australia and globally. The differences in prevalence calculations may be due to the varying definitions of intolerance. Our findings suggest that EHR-detectable phenotypes should be used as decision-support aid rather than as definitive diagnostic tools and that clinical judgement and patient engagement are necessary for the management of suspected statin intolerance.

Keywords: electronic phenotyping, statin intolerance, general practice, electronic health records, rule-based phenotyping algorithms

Key points

This study found that:

- The prevalence of statin intolerance among those prescribed statins in the ePBRN dataset was 5.09%, which is at the low end of the 5–15% range reported in Australia and globally.
- Different phenotyping algorithms show various prevalence estimations, which may be due to the varying definitions of intolerance.
- EHR-detectable phenotypes should be used as decision-support aids rather than as definitive diagnostic tools and that clinical judgement and patient engagement is necessary for the management of suspected statin intolerance.

1 INTRODUCTION

Statins or 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors are known for their ability to lower cholesterol levels and are used as pharmacological treatments for cardiovascular diseases (CVD). It is used for both primary (preventing the development of CVD) and secondary prevention (preventing the occurrence of further CV events) (1, 2). In Australia, statins are prescribed to 30% of patients aged >55 years and are taken by over 40% of patients aged >65 years (3, 4). In addition to their preventive characteristics, statins are known to cause adverse effects in patients (5). Such adverse effects include muscle damage, and in severe cases, muscle breakdown, which may release myoglobin and lead to renal damage (6). This phenomenon is known as statin intolerance and is characterized by symptoms such as muscle weakness, myopathy (muscle aches, muscle cramps, myositis, and rhabdomyolysis), and increased liver enzymes (7). While statins are generally well tolerated, statin intolerance can occur (completely or partially) in approximately 5–15% of patients (7, 8, 9). Factors associated with statin intolerance include age, gender, ethnicity, hypothyroidism, obesity, chronic liver disease, diabetes mellitus, and renal failure (5).

At present, there is no established taxonomy to classify statin intolerance; therefore, the condition is defined differently by different scientific groups, trialists, and societies. The International Lipid Expert Panel (ILEP) defines statin intolerance as the inability to tolerate a statin dose required to reduce a patient's CV risk, which limits the effective treatment of the patient (10). According to the National Lipid Association (NLA), statin intolerance includes any adverse effects on the quality of life that leads to a decrease or cessation of statin use (11). The European Atherosclerosis Society (EAS) defines statin intolerance as the probability of occurrence of statin-associated muscle symptoms (SAMS), considering the nature of muscle symptoms, elevation in Creatine Kinase (CK) levels, and temporal association with statin initiation and discontinuation (12). The Luso-Latin American Consortium (LLAC) and the Canadian Consensus Working Group (CCWG) define statin intolerance as the inability to tolerate two or more statins at any dose or the inability to tolerate increasing doses of statins (13, 14). However, the most recent definition of statin intolerance by NLA is "one or more adverse effects associated with statin therapy that resolves or improves with dose reduction or discontinuation and can be classified as a complete inability to tolerate any dose of a statin or partial intolerance with inability to tolerate the dose necessary to achieve the patient-specific therapeutic objective. To classify a

patient as having statin intolerance, a minimum of two statins should have been attempted, including at least one at the lowest approved daily dosage" (15).

Statin intolerance can be measured or quantified as the number and severity of adverse events/reactions to statin therapy, severity of reports of muscle pain and weakness, discontinuation of statins due to adverse effects, creatinine kinase values, and symptom measures, including patient-reported myalgia (16, 17). Furthermore, the clinical measures of statin intolerance include the global symptom scale score (Likert scale of symptoms), visual analogue scale (VAS) score (to assess pain), brief pain inventory (BPI), and SAMS-CI scale (18, 19, 20, 21). The management of statin intolerance includes several interventions, including changing the statin dose, switching statins, combining statins with non-statin drug therapy, changing to non-statin drugs, and providing dietary interventions (8, 22, 23).

Identification of statin intolerance can be difficult, not only because of the variability in the definition of the condition but also due to the negative representation of statins by the media, close associates, and other sources (24). Fear of intolerance can lead to nonadherence or discontinuation (5). However, recent studies have shown that this issue can be addressed in electronic health records (EHRs) in primary care using electronic phenotyping. Electronic phenotyping involves the systematic identification and categorization of patient health data to create specific disease phenotypic profiles. The process involves using EHR data to classify patients based on the presence or absence of specific health-related traits or diseases and to leverage structured and unstructured data from EHR systems (25, 26). There are two types of electronic phenotyping: rule-based and machine learning-based. Rule-based phenotyping involves using predefined algorithms to classify patients and relies on specified criteria (such as codes and laboratory values) to define a phenotype, which may involve significant manual effort to validate and implement (27, 28). Machine learning-based approaches involve the application of machine learning algorithms to identify phenotypes, where machines automatically learn from EHR data and improve the identification accuracy of different patient groups over time (25, 29). Rule-based algorithms can be useful in the use of EHR datasets as the information in them can be implicit or discrepant (30). Thus, to ensure the use of correct information to identify disease cohorts within the EHR system, researchers use rule-based algorithms with their own inclusion and exclusion criteria, which are created in collaboration with clinicians (25, 31). The primary objective of this study was to identify statin intolerant patients and determine the prevalence of statin intolerance using electronic

health records from general practice (GP) by implementing rule-based algorithms (or EHR-detectable statin intolerance electronic phenotypes) that apply various definitions, criteria, or guidelines of statin intolerance and are used in different countries. The secondary objective was to validate these algorithms. This study focused on patients who were prescribed statins as primary prevention of CVD.

2 METHODS

This study delved into the various EHR-detectable statin intolerance electronic phenotyping algorithms that implements various definitions of statin intolerance to identify statin intolerant patients. The identification of statin intolerant patients depends a lot on the elevation of the creatine kinase (CK) levels (32).

This study methodology took place in four stages. Firstly, a literature review was conducted to identify EHR-detectable statin intolerance electronic phenotypes. Secondly, the implementation of these electronic phenotypes on dataset was performed. Thirdly, the reference standard data was developed and implemented. And finally, the EHR-detectable statin intolerance electronic phenotypes were validated using performance measures. The overview of the methodology is shown in **Figure 1**.

Figure 1: Overview of the study methodology

ePBRN: Electronic Practice Based Research Network; CK: Creatine kinase; EHR: Electronic health record

2.1 Dataset

The dataset that was used for this study was the Electronic Practice Based Research Network (ePBRN) Linked Dataset, which is a comprehensive electronic health record (EHR) derived from GP facilities in the Fairfield and Wollondilly regions of the South-Western Sydney local

health district (SWS LHD). This dataset contains a vast amount of information about a significant patient population, consisting of about 249,345 patients spread across 11 general practitioner sites, who had at least one visit between 2012 to 2019 (33, 34).

The ePBRN dataset consists of a total of 76 tables which are categorized based on the data extraction tools employed (Best Practice or Medical Director). Thirty-nine of the 76 tables are sourced from Best Practice and 37 tables are sourced from Medical Director. The research objective of this study requires the extraction of data on statin intolerance, which is associated with patient diagnosis, procedures, conditions, visit details, and demographics. For this purpose, 6 tables were strategically chosen from Best Practice—namely, Careplan Goal, Current Rx, Visit Reason, Visits, Past History, and Patients—along with 5 tables from Medical Directors—Diagnosis, Prescription, History, Consultations, and Patients. This meticulous selection enhances the precision and relevance of the experimental data. The implementation involved the use of Python to generate and execute SQL queries, with the creation of a Master table to facilitate the application of electronic phenotyping algorithms. The Master table consists of patients who were prescribed statins. The selection criteria of patients in the ePBRN dataset were patients who initiated with statins, were aged 18 years or above at index date (date when an individual is prescribed a statin for the first time), with no history of statin prescription before index date. Patients who were continuously observed for at least two years after index date with three or more visits for any reason or any medication prescription (active patients) were selected. Patients with no gender or birth year recorded, aged more than 80 years, or who did not survive during the observation period were excluded. Moreover, patients who suffered any CVD events before the index date or during the observation period of 2 years after index date were also excluded. These restrictions were applied to reduce clinical heterogeneity and to focus the study on primary prevention of CVD, where statin prescription and intolerance are more consistent and comparable across individuals.

2.2 Identification and implementation of EHR-detectable statin intolerance electronic phenotypes

A rapid literature review was conducted to identify past studies on EHR-detectable statin intolerance electronic phenotypes. Inclusion and exclusion criteria of the rule-based algorithms were tabulated. The codes for the identified phenotyping algorithms were run on

the prepared dataset using R. The numbers derived from each algorithm was obtained and put in a table.

2.3 Development and implementation of reference standard

The reference standard was developed by taking a subset of all patients from the ePBRN dataset who were prescribed statins. It consisted of patients who were initially identified as potential cases of statin intolerance, and it included all the statin intolerant patients who were detected using the identified phenotyping algorithms (n = 1,369). Two individual researchers (annotators) manually reviewed the records to find (annotate) patients with statin intolerance based on the EAS definition of statin intolerance (12). The EAS definition was treated as a clinical reference definition rather than an electronic phenotype. The manual review was undertaken to identify reference cases of statin intolerance in accordance with the definition. The EAS definition was used as it is the most stringent and clinically reasoned definition, which is suitable since in the EHR data symptoms are under-reported and nocebo effects are indistinguishable from true intolerance. Moreover, the EAS prioritizes specificity over sensitivity and thereby reduces the misclassification of statin intolerance. Any discrepancies in opinion were discussed and resolved. Following this, the Inter Annotator Agreement (IAA) was calculated, which is a measure of how well the same annotation was made by the two researchers. The IAA score was calculated for the two annotators as annotator 1 vs annotator 2. The Cohen's Kappa method was used for this calculation since both the annotators used binomial (yes/no) annotation for the data. A higher IAA scores means a better agreement and compliance with the annotation guidelines (35, 36). For the identification of statin intolerant patients, the EAS definition was adopted (12). Furthermore, the Deviation Score (DS) was also calculated to find the agreement between the final reference standard and each annotator. The DS for Annotator 1 was the agreement between the reference standard and Annotator 1, and DS for Annotator 2 was the agreement between the reference standard and Annotator 2 (35, 37). The development and implementation of the reference standard is shown in **Figure 2**. The threshold for CK was different for male and female patients, as well as those with and without muscle symptoms.

The prevalence of statin intolerance was calculated based on the final annotation by the researchers who assessed the reference standard. The demographic distribution of the reference standard was presented in a table. Area-based social advantage and disadvantage

was classified using the Socio-Economic Indexes for Australia (SEIFA) 2021 Index of Relative Socio-Economic Advantage and Disadvantage (IRSAD) which is based on suburbs and localities (38).

Figure 2: Development and implementation of reference standard

ePBRN: Electronic Practice Based Research Network; CK: Creatine kinase; IAA: Inter Annotator Agreement

* The threshold for creatine kinase (CK) was different for patients with and without muscle symptoms

2.4 Validation of the EHR-detectable statin intolerance phenotypes

Separate from the manual identification of statin intolerant patients using EAS definition, established rule-based electronic phenotyping algorithms were implemented to enable automated identification of statin intolerance using the ePBRN dataset. The phenotypes were then validated against the manually identified reference standard. For the validation of each algorithm, the performance measures used were accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F1 score and correlation coefficient scores. Accuracy is the number of correctly identified statin tolerant and intolerant patients between the measured values from each algorithm and the true values from the reference standard (the number of correct predictions over the total number of predictions) (39). Sensitivity is the number of correctly identified patients with statin intolerance (the number of true positives over the sum of the true positives and false negatives) (40). Specificity is the number of correctly identified patients without statin intolerance (the number of true negatives over the sum of true negatives and false positives) (40). Positive predictive value is the number of identified patients with statin intolerance who are actually statin intolerant (the number of true positives over the sum of the true positives and false positives) (40). Negative predictive value is the number of identified patients without statin intolerance who are actually statin tolerant (the number of true negatives over the sum of the true negatives and false negatives)

(40). The reference standard, consisting of a subset of 1,369 patients, was considered as the correct or true values.

2.5 Ethical considerations

This study and the use of data from ePBRN was approved by the UNSW Human Research Ethics Advisory Panel (HC230066; 23 June 2023).

3 RESULTS

3.1 Identification of EHR-detectable statin intolerance electronic phenotypes

Literature review helped to identify EHR-detectable statin intolerance electronic phenotypes. Six rule-based phenotyping algorithms were employed in this study, namely, Minnesota Combined Rule-Based (CRB) algorithm (41), Japan - Statin induced myopathy (SIMs) (42), USA - SIMs (16), Singapore - SIMs (algorithms A, B, C, and D) (43), Japan - Statin-associated muscle toxicity (SAMT) (44), and NHS - UK - Statin intolerance pathway (45). The selection of these algorithms was based on their utilization of structured data elements such as diagnosis codes, medications, and procedure codes for establishing inclusion and exclusion criteria. The overview of the identified phenotyping algorithms is shown in **Table 1**.

Table 1: Overview of the identified EHR-detectable statin intolerance electronic phenotyping algorithms

Algorithm	Definition	Organization of the SI definition	Setting	Algorithm rule
Minnesota Combined Rule-Based **(Minnesota CRB)**	SAMS and CK > 3 times ULN (145 IU/L for females and 170 IU/L for males)	NLA and EAS	Primary	**Inclusion:** • Patients with Age > 18 **Exclusion:** • Patients with no BP, weight and laboratory values 3 months preceding statin treatment (baseline) • Patients with no BP, weight and laboratory values 1 year after statin treatment (follow-up period)
Japan - Statin induced myopathy – Suspected **(Japan SIMs-S)**	The types of myopathy targeted were myositis and rhabdomyolysis, which are detectable by an increase in CK values (750 IU/L for males and 500 IU/L for females)	EAS	Secondary / Tertiary	**Inclusion:** • Patients with CK level 3 times higher than (750 IU/L for male and 500 IU/L for female) after discontinuation of statins • Patients having CK level normal or lower than 250 IU/L (for males) and 170 IU/L (for females) before starting statins • Patients having CK levels 3 times the normal (above 750 IU/L for male and 500 IU/L for female) within 15 days after the last statin dose **Exclusion:** • Patients with non-statins related conditions with elevated CK within 7 days before/after/throughout event period
USA - Statins induced myopathy **(USA SIMs)**	≥1 incident of all-cause elevated CK after statin initiation with no elevation within the 5 years prior to statin initiation.	(Custom defined)	Primary/ Secondary / Tertiary	**Inclusion:** • Patients with CK level 4 times higher than ULN (336 IU/L in males and 176 IU/L in females) after initiation of statins • Patients with no elevated CK level before statin treatment **Exclusion:** • Patients continue statins after elevated CK • Patients with non-statins related conditions but with elevated CK

Singapore - Statin induced myopathy -Algorithm A **(Singapore SIMs-A)**	Custom defined (CK elevation >4x ULN)	CCWG	Tertiary	**Inclusion:** • Patients with CK level 4 times higher than ULN (250 IU/L in males and 150 IU/L in females) after initiation of statins **Exclusion:** • Patients with troponin level > 40 and CK MB level > 3
Singapore - Statin induced myopathy -Algorithm B **(Singapore SIMs-B)**				**Inclusion:** • Patients with statin related muscle symptoms
Singapore - Statin induced myopathy -Algorithm C **(Singapore SIMs-C)**				**Inclusion:** • Patients with CK level 4 times higher than ULN (250 IU/L in males and 150 IU/L in females) after initiation of statins • Patients with statin related muscle symptoms
Singapore - Statin induced myopathy -Algorithm D **(Singapore SIMs-D)**				**Inclusion:** • Patients with CK level 4 times higher than ULN (250 IU/L in males and 150 IU/L in females) after initiation of statins • Patients with statin related muscle symptoms **Exclusion:** • Patients with troponin level > 40 and CK MB level > 3
Japan - Statin-associated muscle toxicity -Criteria A **(Japan SAMT-A)**	Custom defined	(Custom defined)	Secondary / Tertiary	**Inclusion:** • Patients with CK level 4 times higher than ULN (250 IU/L in males and 150 IU/L in females) after initiation of statins • Patients prescribed with drugs which share some pharmacokinetic interactions with statins • Patients with muscle related symptoms

				Exclusion: • Patients with non-statins related conditions which leads to elevated CK level 3 months prior to statin prescription • Patients prescribed nitrates and levothyroxine prescriptions within 3 days after concurrant elevation of CK
NHS - UK - Statin intolerance pathway **(NHS-UK)**	SAMS and CK elevation >4x ULN and <10x ULN	EAS and NICE guidelines	Primary/ Secondary	**Inclusion:** • Patients with Statins related muscle symptoms • Patient with creatine kinase level between 4x ULN (250 IU/L in males and 150 IU/L in females) and 10x ULN (250 IU/L in males and 150 IU/L in females)

SI: Statin intolerance; CRB: Combined rule-based; SAMS: Statin-associated muscle symptoms; CK: Creatine kinase; ULN: Upper limit of normal; IU/L: International units per litre; NLA: National Lipid Association; EAS: European Atherosclerosis Society; BP: Blood pressure; SIMs: Statin induced myopathy; USA: United States of America; CCWG: Canadian Consensus Working Group; CK MB: Creatine kinase myocardial band; SAMT: Statin-associated muscle toxicity; NHS-UK: National Health Service-United Kingdom; NICE: National Institute for Health and Care Excellence

3.2 Prevalence of statin intolerance

The prevalence of statin intolerance among those who were prescribed statins in the ePBRN dataset was found to be 5.09% (794/15,583). The ePBRN data consists of a total of 249,345 patients with 3,103,072 visits, of whom 18,377 patients with 110,147 visits were prescribed with statins. After excluding patients who were less than 18 years of age at index date (date when an individual was prescribed a statin for the first time), there were a total of 18,371 patients with 110,083 visits. Based on the further exclusion criteria 2,788 patients with 24,358 visits were excluded, leaving 15,583 GP patients with 85,725 visits for analysis. The details of patient selection process are shown in **Figure 3 and** demographics of patients with statin intolerance in **Table 2**.

Figure 3: Identification of eligible patients from the ePBRN dataset

ePBRN: Electronic Practice Based Research Network; CVD: Cardiovascular disease

Table 2: Demographic distribution of patients with statin intolerance

Variable	Category	Statin intolerant (N = 794)	Total population (N = 15,583)
Age [mean (SD)]		58.40 (10.80)	55.80 (12.40)
Age [n (%)]	< 35	8 (1.01)	815 (5.23)
	35 – 44	75 (9.45)	2,208 (14.17)
	45 – 54	231 (29.09)	4,171 (26.77)
	55 – 64	249 (31.36)	4,556 (29.24)
	65 – 74	185 (23.30)	2,920 (18.74)
	75 – 84	43 (5.42)	901 (5.78)
	85 – 94	< 5 (< 0.63)	12 (0.08)
Gender [n (%)]	Male	349 (43.95)	7,866 (50.48)
	Female	444 (55.92)	7,704 (49.44)

	Others	< 5 (< 0.63)	< 5 (< 0.03)
	Invalid	< 5 (< 0.63)	10 (0.06)
Ethnicity [n (%)]	ATSI	< 5 (< 0.63)	169 (1.08)
	Non-ATSI	719 (90.55)	11,274 (72.35)
	Not provided/recorded	71 (8.94)	4,140 (26.57)
Employment status [n (%)]	Employed	156 (19.65)	3,471 (22.27)
	Unemployed	105 (13.22)	2,577 (16.54)
	Retired/pensioner	160 (20.15)	1,332 (8.55)
	Not recorded	373 (46.98)	8,203 (52.64)
SEIFA IRSAD category (decile) [n (%)]	1	156 (19.65)	1,703 (10.93)
	2	10 (1.26)	1,046 (6.71)
	3	226 (28.46)	3,772 (24.21)
	4	95 (11.96)	2,688 (17.25)
	8	275 (34.63)	5,417 (34.76)
	9	32 (4.03)	957 (6.14)
Smoking status [n (%)]	Smoker	222 (27.96)	4,112 (26.39)
	Ex-smoker	81 (10.20)	1,076 (6.90)
	Non-smoker	401 (50.50)	6,538 (41.96)
	Not recorded	90 (11.33)	3,857 (24.75)
Statin intensity at index date [n (%)] *	High	164 (20.65)	3,547 (22.76)
	Moderate	571 (71.91)	10,507 (67.46)
	Low	51 (6.42)	574 (3.68)
Statin type at index date [n (%)] *	Atorvastatin	308 (38.79)	6,233 (40.00)
	Rosuvastatin	305 (38.41)	5,941 (38.12)
	Simvastatin	132 (16.62)	1,924 (12.35)
	Fluvastatin	4 (0.50)	33 (0.21)
	Pravastatin	37 (4.66)	497 (3.19)

SD: Standard deviation; ATSI: Aboriginals and Torres Strait Islanders; SIEFA: Socio-Economic Indexes for Australia; IRSAD: Index of Relative Socio-economic Advantage and Disadvantage

* Missing data: 8 patients in statin intolerant group; 955 patients in total population group

3.3 Eligible patient selection

There were 1,369 patients in the reference standard developed. The Cohen's Kappa (IAA) between the two annotators was 0.894 **(Supplement 1)**. The DS for Annotator 1 was 0.968 and for Annotator 2 was 0.989, with an average DS of 0.978. The precision and recall for Annotator 1 were 0.944 and 0.994, and those for Annotator 2 were 0.985and 0.992, respectively **(Supplement 2)**.

In the reference standard, 90.36% patients were active, 7.38% patients were inactive, and 2.26% patients had deceased. The demographic distribution of the patients in the reference standard is shown in **Table 2**. Of the patients in the reference standard, statin intolerance was found in 794 patients. The mean age was 58.40 (±10.80) years for statin intolerant patients and 55.80 (±12.40) years for the total patients. The majority of statin intolerant patients were female (55.92%), aged 55 – 64 years (31.36%), non-smokers (50.50%) and residing in SEIFA IRSAD category 8 (34.63%). Similarly, among all patients, most were male (50.48%), within the same age group of 55 – 64 years (29.24%), non-smokers (41.96%) and predominantly from SEIFA IRSAD category 8 (34.76%).

3.4 Implementation and validation of the EHR-detectable statin intolerance electronic phenotyping algorithms

The six rule-based phenotyping algorithms were Minnesota Combined Rule-Based (CRB) algorithm, Japan - Statin induced myopathy (SIMs), USA - SIMs, Singapore - SIMs (algorithms A, B, C, and D), Japan - Statin-associated muscle toxicity, NHS - UK - Statin intolerance pathway. The prevalence of statin intolerance, as identified by each of the algorithm is shown in **Figure 4**. The frequency of matches by each algorithm with the reference standard is shown in **Supplement 3**. **Supplement 4** shows all the overlaps among the algorithms and the reference standard.

Figure 4: Prevalence of statin intolerance, as identified by each algorithm (N = 15,583)

NHS-UK: National Health Service-United Kingdom; SAMT: Statin-associated muscle toxicity; SIMs: Statin induced myopathy; USA: United States of America; CRB: Combined rule-based

The results of the validation of each of the electronic phenotyping algorithms are shown in **Table 3.** The algorithms correctly classified statin intolerance in 44.12%, 42.07%, 42.88%, 42.73%, 57.05%, 42.22%, 42.51%, 42.80%, and 42.29% of the patients respectively (accuracy scores). Among all the patients with statin intolerance, the algorithms identified 17.63%, 1.76%, 2.90%, 2.77%, 92.95%, 3.02%, 2.90%, 2.02%, and 3.40% patients respectively (sensitivity scores). Among all the patients with statin tolerance, the algorithms identified 80.70%, 97.74%, 98.09%, 97.91%, 7.48%, 96.35%, 97.22%, 99.13%, and 96.00% patients respectively (specificity scores). Among all the patients the algorithms identified as having statin intolerance, 55.78%, 51.85%, 67.65%, 64.71%, 58.11%, 53.33%, 58.97%, 76.19%, and 54.00% truly had statin intolerance, respectively (PPV scores). Among all the patients the algorithms identified as statin tolerant, 41.50%, 41.88%, 42.25%, 42.17%, 43.43%, 41.84%, 42.03%, 42.28%, and 41.85% truly were statin tolerant, respectively (NPV scores). Singapore SIMs-B showed the highest accuracy, sensitivity, and NPV, while Japan SAMT showed the highest specificity and PPV. The true positives, true negatives, false positives and the false negatives are shown in **Supplement 5**.

Additional metrics when compared to reference standard showed that Singapore SIMs-B algorithm achieved the highest F1 score (71.51%), indicating the strongest predictive values for statin intolerance. However, the Japan SAMT algorithm had the strongest correlation with the reference standard as displayed by the Matthew's correlation coefficient (MCC) and Pearsons's correlation coefficient (PCC) value of 0.046.

The area under the curve (AUC) for Receiver operating characteristic (ROC) curve of sensitivity against specificity shows very similar performance for all algorithms. All algorithms had relatively poor discriminating ability to identify statin intolerance (ROC AUC 0.50 – 0.52). However, the algorithm Japan SAMT were better than the rest of the algorithms at identifying true statin intolerant patients, based on its Precision–recall (PR) AUC value of 0.593. The ROC AUC and PR AUC values for each of algorithm is shown in **Supplement 6**. Furthermore, the graphs of the area under the curves are shown in **Supplement 7**. The heatmap table showing the performance metrics of each algorithm for identifying patients with statin intolerance compared to the reference standard is shown in **Supplement 8**.

Table 3: Summary performance metrics of each algorithm to identify patients with statin intolerance compared to reference standard (N = 1,369)

Algorithm	Counts (n)	Accuracy [% (95% CI)]	Sensitivity [% (95% CI)]	Specificity [% (95% CI)]	PPV [% (95% CI)]	NPV [% (95% CI)]	Balanced Accuracy [% (95% CI)]	F1 score [% (95% CI)]	MCC score (95% CI)	PCC score (95% CI)
Singapore SIMs-B	1270	57.05 (54.38 – 59.69)	92.95 (90.94 –94.63)	7.48 (5.46 – 9.94%)	58.11 (55.34 – 60.84)	43.43 (33.50 – 53.77)	50.21 (14.24 – 86.19)	71.51 (69.51 – 73.45)	0.01 (-0.04 – 0.06)	0.01 (-0.05 – 0.06)
Minnesota CRB	251	44.12 (41.47 – 46.80)	17.63 (15.04 – 20.46)	80.70 (77.23 – 83.84)	55.78 (49.4 – 62.02)	41.50 (38.6 – 44.45)	49.16 (-4.60 – 102.93)	26.79 (24.13 – 29.59)	-0.02 (-0.08 – 0.03)	-0.02 (-0.07 – 0.03)
Japan SIMs	34	42.88 (40.24 –45.55)	2.90 (1.84 – 4.31)	98.09 (96.6 – 99.04)	67.65 (49.47 – 82.61)	42.25 (39.58 – 44.95)	50.49 (29.27 – 71.71)	5.56 (4.10 – 7.34)	0.03 (-0.02 – 0.08)	0.03 (-0.02 – 0.08)
Japan SAMT	21	42.80 (40.17 – 45.48)	2.02 (1.16 – 3.25)	99.13 (97.98 – 99.72)	76.19 (52.83 – 91.78)	42.28 (39.63 – 44.97)	50.57 (34.07 – 67.08)	3.93 (2.7 – 5.5)	0.05 (-0.01 – 0.09)	0.05 (-0.01 – 0.10)
Singapore SIMs-A	34	42.73 (40.09 – 45.40)	2.77 (1.74 – 4.17)	97.91 (96.38 – 98.92)	64.71 (46.49 – 80.25)	42.17 (39.51 – 44.87)	50.34 (29.01 – 71.67)	5.31 (3.89 – 7.07)	0.02 (-0.03 – 0.07)	0.02 (-0.03 – 0.08)
Singapore SIMs-D	39	42.51 (39.88 – 45.18)	2.90 (1.84 – 4.31)	97.22 (95.52 – 98.4)	58.97 (42.10 – 74.43)	42.03 (39.36 – 44.74)	50.06 (27.04 – 73.08)	5.52 (4.07 – 7.30)	0.00 (-0.05 – 0.06)	0.00 (-0.05 – 0.06)
NHS-UK	50	42.29 (39.66 – 44.96)	3.40 (2.25 – 4.91)	96.00 (94.06 – 97.45)	54.00 (39.32 – 68.19)	41.85 (39.17 – 44.57)	49.7 (23.54 – 75.86)	6.40 (4.84 – 8.27)	-0.02 (-0.07 – 0.04)	-0.02 (-0.07 – 0.04)
Singapore SIMs-C	45	42.22 (39.59 – 44.89)	3.02 (1.95 – 4.46)	96.35 (94.47 – 97.73)	53.33 (37.87 – 68.34)	41.84 (39.17 – 44.55)	49.69 (24.80 – 74.57)	5.72 (4.25 – 7.51)	-0.02 (-0.07 – 0.04)	-0.02 (-0.07 – 0.04)
USA SIMs	27	42.07 (39.44 – 44.74)	1.76 (0.97 – 2.94)	97.74 (96.16 – 98.79)	51.85 (31.95 – 71.33)	41.88 (39.22 – 44.57)	49.75 (30.29 – 69.21)	3.41 (2.28 – 4.89)	-0.02 (-0.07 – 0.04)	-0.02 (-0.07 – 0.04)

CRB: Combined rule-based; USA: United States of America; SIMs: Statin induced myopathy; SAMT: Statin-associated muscle toxicity; NHS-UK: National Health Service-United Kingdom; PPV: Positive predictive value; NPV: Negative predictive value; MCC: Matthew's correlation coefficient; PCC: Pearson's correlation coefficient

4 DISCUSSION

This study found that the prevalence of statin intolerance in the ePBRN dataset was 5.09%, which aligns marginally with the prevalence of statin intolerance in Australia which is in between 5–15% of patients (7, 8). The observed low prevalence could be due to the varying definitions used for statin intolerance by the different algorithms (such as SIMs and SAMT) (46). Statin intolerance is often defined as muscle-related symptoms or other adverse effects such as elevated liver enzyme or gastrointestinal disturbances that lead to dose reduction, switch, or discontinuation and several studies suggest that the estimate of statin intolerance is highly dependent on the definition, setting, and ascertainment of the case (5, 47, 48). Meta-analysis shows that the prevalence can be as low as 3-5%, when strict definitions are used, and can be as high as 10-30% when broader definitions are employed (5, 47, 48). The 5.09% estimate of statin intolerance from the ePBRN dataset used the EAS definition, which is a narrower and stringent definition that reflects clinical experiences leading to change in therapy, but does not capture milder symptoms or borderline cases (12, 48). The alignment of the observed prevalence with the lower limit of the Australian range suggests that the prevalence of statin intolerance depends on methodological criteria. However, an additional consideration while interpreting the data is the nocebo/drucebo effect, which is a phenomenon where the adverse symptoms of statins are influenced by patient expectation and/or fear of intolerance (often inflicted through negative representations by the media or close associates) (5, 49). This results in overestimation of intolerance, making it a crucial factor during the identification of statin intolerant patients (50, 51). Moreover, evidence from some studies show that some patients often experience improvements in muscle symptoms while taking statins (52). This highlights the fluctuating a non-specific nature of muscle symptoms during statin use and reinforces the need for structured definitions and validations while identifying statin intolerance using EHR-based data.

The six rule-based phenotyping algorithms used for the identification of statin intolerant patients employed structured data elements such as diagnosis codes, medications, and procedure codes. These algorithms were Minnesota CRB algorithm (41), Japan - SIMs (42), USA - SIMs (16), Singapore - SIMs (algorithms A, B, C, and D) (43), Japan - SAMT (44), and NHS - UK - Statin intolerance pathway (45). Our findings suggest that none of the algorithms performed consistently well across both screening and confirmation domains. Rule-based phenotyping algorithms are used for cohort identification using EHR data for a

wide variety of diseases (53, 54). This is because clinicians may provide implicit or discrepant data in the EHR system, while the rule-based algorithms can help in identifying the diseased patients by using the correct information (30, 31). The rule-based algorithms could use the recording of a condition with one or more relevant medical codes (such as International Classification of Diseases [ICD] terminology), or it could also use multiple domains, such as observations, measurements, procedures, clinical history, lab test results (such as CK elevations and liver enzymes), and/or genetic data in addition to patient conditions (31, 55, 56). Most of this complexity and robustness can be addressed using rule-based algorithms for EHR data which explains the heterogeneity and missingness of data and provides more flexibility to correct implicit biases (30, 31, 57).

The Minnesota CRB algorithm relied on chart-verified clinical events (such as dose adjustments, therapy changes and discontinuations) in a primary care setting. In a GP setting, this algorithm can be moderately adapted by mapping intolerance signals to structured events motivated by symptoms and sustained CK elevations (5). Although the Japan SIMs algorithm is a hospital-based study, the algorithm focuses on muscle-related adverse effects and can be adapted in the GP setting, where history of prescription, symptoms, and CK values can be obtained and used to determine the presence of statin-related myalgia or CK elevation (47). The USA SIMs algorithm used a definition that combined symptoms, adverse events, and therapy changes, across all settings, for which reason it may be easier to adapt this algorithm in the context of GP data (48). The Singapore SIMs-B algorithm was used for hospital-based cohorts in the tertiary care settings but since it follows the EAS definition of statin intolerance, its application may be moderately feasible in the context of GP setting (43). The Japan SAMT algorithm was primarily designed in the context of trials, which makes it a less feasible algorithm to implement it in the GP settings, although the concepts could be translated with adaptation (44). The NHS-UK algorithm is a management focused pathway which provides standardized definitions of complete versus partial intolerance with explicit rechallenge steps and therapy alternatives but is not focused on patient identification. This is compatible with GP data structures, as clinicians routinely document prescriptions, symptoms, and objective CK results, enabling alignment with pathway criteria to determine intolerance status and management steps (58).

Applying the Singapore SIMs-B algorithm in this study showed the highest accuracy (57.05%), sensitivity (92.95%), NPV (43.43%) and F1 (71.51%) scores, while applying the

Japan SAMT algorithm showed the highest specificity (99.13%), PPV (76.19%) and correlation coefficient (MCC and Pearson's correlation coefficient; 0.05%). The high sensitivity and NPV means that the Singapore SIMs-B algorithm was robust and the most correct in identifying and classifying truly statin intolerant patients (46, 59). This helps in ruling out intolerance which is expected to give clinicians more confidence to re-challenge, optimize dose, consider alternative therapy or escalate therapy without unnecessary avoidance (60, 61). The high specificity in the Japan SAMT algorithm means that more truly statin tolerant patients were identified and classified by this algorithm (46, 59). A high specificity and high PPV can inform a clinician to decide if it is necessary for a patient to discontinue or switch therapy when intolerance is strongly suspected, which reduces exposure to ineffective regimens (60, 61). Due to the low prevalence and the varying definitions of statin intolerance found in this study, the NPV scores are more reassuring since PPV scores increase with the increase in prevalence of statin intolerance, which affects the interpretation of positive intolerance assessments (61, 62). However, in the context where the prevalence of statin intolerance is around 5%, a high MCC suggests substantial agreement beyond random chance, reinforcing the reliability of the Japan SAMT algorithm for confirming patients who are intolerant (60, 62).

The identification of statin intolerant patients depends on several factors, and the understanding these factors may be the key to identifying these patients. Furthermore, addressing statin intolerance can affect statin adherence. Thus, the accurate identification of intolerant patients is critical for the strategies to improve statin adherence or adherence to alternative treatment regimens involving other lipid-lowering medications (63). In our previous study, we looked at inflammatory markers of CVD in patients prescribed statins from the ePBRN dataset, whereas in this study we looked at the identification of statin intolerant patients (33). While the metrics found in this study suggest that the Singapore SIMs-B algorithm can be used to screen for potential statin intolerant patients who are eligible for statin re-challenge, and the Japan SAMT algorithm can be used to confirm statin intolerance where the clinical decision is to discontinue or switch therapy (46, 59), the other algorithms may also have potential and could be considered for adaptation in different contexts as well.

Limitations

This study was limited to primary prevention cohort. Due to the lack of a standardized taxonomy, statin intolerance has various definitions, including SAMT and SIMs, making it a complex subject matter. This study evaluated the application of established electronic phenotyping algorithms. The reference standard was developed through the manual application of the EAS definition, which provided a clinically informed reference for validation. As such, the reference standard is a pragmatic benchmark rather than a definitive standard, as it depends on routine clinical documentation and variability in documentations. The different algorithms in this study used different definitions which may have influenced the low prevalence of statin intolerance found in this study. Moreover, the varying sample sizes reported in the studies may be significant. Since PPV and NPV are prevalence dependent metrics, the limitation in estimating the accurate prevalence of statin intolerance might have affected the PPV and NPV values. The low prevalence found in this study meant that the PPV may be high even for an algorithm with high specificity. The lack of a universal reference standard for statin intolerance in EHR data can introduce misclassification bias. Furthermore, CK distributions are known to differ by ethnicity along with the gender and presence or absence of muscle symptoms. In this study the variation in CK due to ethnicity was not considered which may have also introduced misclassification bias during the interpretation of the operational definition of statin intolerance. Due to these biases and the various population and definitions of statin intolerance, the results obtained in this study may not generalised. The study was conducted with a EHR derived data from a population in Southwestern Sydney, and the findings may not be directly transferable to other healthcare systems or populations. Thus, the results are interpreted as context-specific estimates, reflecting local data structure, prescribing practices, and documentation patterns. The generalizability could be tested under alternative reference standards and in different populations.

5 CONCLUSION

This study found that the prevalence of statin intolerance in the ePBRN dataset was found to be 5.09%. The six rule-based phenotyping algorithms for identifying statin intolerance were Minnesota CRB algorithm, Japan - SIMs, USA - SIMs, Singapore - SIMs (algorithms A, B, C, and D), Japan - SAMT, and NHS - UK - Statin intolerance pathway. None of these

algorithms performed consistently well across both screening and confirmation domains. The Singapore SIMs-B algorithm displayed the highest accuracy (57.05%), sensitivity (92.95%), NPV (43.43%) and F1 (71.51%) scores, while the Japan SAMT algorithm displayed the highest specificity (99.13%), PPV (76.19%) and correlation coefficient values (MCC and Pearson's correlation coefficient; 0.05%). Based on these metrics, it can be inferred that in practice it might be logical for use a two-stage approach of screening with Singapore SIMs-B to flag potential intolerance patients and subsequently confirming intolerant patients with Japan SAMT, before making definitive management changes. This will align with a diagnostic workflow that emphasizes minimizing missed cases and avoiding unnecessary treatment changes. However, in case of low statin intolerance prevalence, if clinicians aim to minimize missed true intolerance while preserving safety, then the Singapore SIMs-B could be employed as a screening tool to identify candidates for re-challenge or dose optimization. In case of high statin intolerance prevalence, if clinicians aim to confirm intolerance with high confidence to justify therapy change, then they could rely on the Japan SAMT algorithm to convey robust agreement with the reference standard and to use as a decisional anchor for discontinuation or switching therapies. Despite this, based on the study settings and objectives, the other algorithms, especially the NHS-UK algorithm, can also align well with the ePBRN dataset. The selection of an algorithm, or a combination of algorithms for the identification of statin intolerant patients depends significantly on several factors, and the understanding these factors may be the key to identifying these patients. A pragmatic workflow may integrate genetic and phenotypic data to refine predictions and enhance predictive values, potentially improving both screening and confirmation phases and enabling personalized intolerance management in statin therapy. Therefore, the use or adaptation of any single algorithm or a combination of algorithms to identify statin intolerant patients depends on the context, such as the setting in which the algorithm is being adapted and the intended target patients for the algorithm. This study supports guideline recommendations that suggest reassessment, dose modification, or rechallenge before labelling patients as statin intolerant. Moreover, the results emphasise that the EHR-detectable phenotypes should be used as decision-support aids rather than as definitive diagnostic tools and that clinical judgement and patient engagement is necessary for the management of suspected statin intolerance.

Acknowledgements: We would like to thank the ePBRN Primary Care Health Informatics Working Group of the Secure Research Environment for Digital Health (SREDH) Consortium (www.sredhconsortium.org, accessed on 1 November 2025) for their assistance with access to the ePBRN dataset to investigate the findings from this review. We would also like to acknowledge Preetham Kadappu and Kwok Leung Ong for their preliminary work.

Funding: The study was funded by the Australian National Health and Medical Research Council (Grant Number: GNT1192469) and Medical Research Future Fund (Grant Number: GA414054). JJ also acknowledges the funding support received through the 296 Research Technology Services at UNSW Sydney, Google Cloud Research (Award Number: 297 GCP19980904) and NVIDIA Academic Hardware grant programs.

Competing Interests: Joel Rhee received honoraria from Merck Sharp & Dohme (MSD) and Pfizer for presenting and assisting with the organisation of clinician education events. Jitendra Jonnagaddala has served in a consulting or advisory capacity for WHO and UNICEF and he also received speakers' fees from the Ministry of Health, Indonesia. The Authors declare that they have no other competing interests.

Author contributions: SR: Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing; JR: Conceptualization, Resources, Supervision, Validation, Writing – review & editing; STL: Resources, Supervision, Validation, Writing – review & editing; KAR: Resources, Supervision, Validation, Writing – review & editing; JJ: Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing – review & editing

Data sharing statement: The corresponding author will make the data available up on request.

REFERENCES

1. Kang DO, Park Y, Seo JH, Jeong MH, Chae SC, Ahn TH, et al. Time-dependent prognostic effect of high sensitivity C-reactive protein with statin therapy in acute myocardial infarction. Journal of cardiology. 2019;74(1):74-83.

2. Amiri M. Worldwide statins prescription pattern: is it similar. Biom Biostat Int J. 2020;9(6):194.

3. MedicineWise NPS. Statins revisited–appropriate patient selection and management are key.

4. Hilmer S, Gnjidic D. Statins in older adults. Australian Prescriber. 2013;36(3).

5. Bytyçi I, Penson PE, Mikhailidis DP, Wong ND, Hernandez AV, Sahebkar A, et al. Prevalence of statin intolerance: a meta-analysis. European heart journal. 2022;43(34):3213-23.

6. Randall MD, Kendall D, Alexander S. chapter 13 Cardiovascuar drugs. FASTtrack: Pharmacology. 2nd edition ed: Pharmaceutical Press; 2012.

7. Fitchett DH, Hegele RA, Verma S. Statin intolerance. Circulation. 2015;131(13):e389-e91.

8. Toth PP, Patti AM, Giglio RV, Nikolic D, Castellino G, Rizzo M, et al. Management of statin intolerance in 2018: still more questions than answers. American Journal of Cardiovascular Drugs. 2018;18:157-73.

9. Schulman KL, Lamerato LE, Dalal MR, Sung J, Jhaveri M, Koren A, et al. Development and validation of algorithms to identify statin intolerance in a US administrative database. Value in Health. 2016;19(6):852-60.

10. Banach M, Rizzo M, Toth PP, Farnier M, Davidson MH, Al-Rasadi K, et al. Statin intolerance–an attempt at a unified definition. Position paper from an International Lipid Expert Panel: This paper is also published in parallel in Archives of Medical Science [Banach M, Rizzo M, Toth PP, et al. Statin intolerance–an attempt at a unified

definition. Position paper from an International Lipid Expert Panel. Arch Med Sci 2015; 11 (1): 1–23]. Expert opinion on drug safety. 2015;14(6):935-55.

11. Guyton JR, Bays HE, Grundy SM, Jacobson TA. An assessment by the Statin Intolerance Panel: 2014 update. Journal of clinical lipidology. 2014;8(3):S72-S81.

12. Stroes ES, Thompson PD, Corsini A, Vladutiu GD, Raal FJ, Ray KK, et al. Statin-associated muscle symptoms: impact on statin therapy—European Atherosclerosis Society consensus panel statement on assessment, aetiology and management. European heart journal. 2015;36(17):1012-22.

13. Mancini GJ, Baker S, Bergeron J, Fitchett D, Frohlich J, Genest J, et al. Diagnosis, prevention, and management of statin adverse effects and intolerance: Canadian Consensus Working Group Update (2016). Canadian Journal of Cardiology. 2016;32(7):S35-S65.

14. Sposito AC, Faria Neto JR, Carvalho LSFd, Lorenzatti A, Cafferata A, Elikir G, et al. Statin-associated muscle symptoms: position paper from the Luso-Latin American Consortium. Current Medical Research and Opinion. 2017;33(2):239-51.

15. Cheeley MK, Saseen JJ, Agarwala A, Ravilla S, Ciffone N, Jacobson TA, et al. NLA scientific statement on statin intolerance: a new definition and key considerations for ASCVD risk reduction in the statin intolerant patient. Journal of Clinical Lipidology. 2022;16(4):361-75.

16. Alfirevic A, Neely D, Armitage J, Chinoy H, Cooper RG, Laaksonen R, et al. Phenotype standardization for statin□induced myotoxicity. Clinical Pharmacology & Therapeutics. 2014;96(4):470-6.

17. Jacobson TA, Edelman SV, Galipeau N, Shields AL, Mallya UG, Koren A, et al. Development and content validity of the statin experience assessment questionnaire (SEAQ)©. The Patient-Patient-Centered Outcomes Research. 2017;10:321-34.

18. Rosenson RS, Gandra SR, McKendrick J, Dent R, Wieffer H, Cheng L-I, et al. Identification and management of statin-associated symptoms in clinical practice:

extension of a clinician survey to 12 further countries. Cardiovascular Drugs and Therapy. 2017;31:187-95.

19. Kraut R, Wierenga F, Molstad E, Korownyk C, Perry D, Dennett L, et al. Intolerance upon statin rechallenge: A systematic review and meta-analysis of randomized controlled trials. Plos one. 2023;18(12):e0295857.

20. Herrett E, Williamson E, Brack K, Beaumont D, Perkins A, Thayne A, et al. Statin treatment and muscle symptoms: series of randomised, placebo controlled n-of-1 trials. bmj. 2021;372.

21. Taylor BA, Lorson L, White CM, Thompson PD. A randomized trial of coenzyme Q10 in patients with confirmed statin myopathy. Atherosclerosis. 2015;238(2):329-35.

22. Raju SB, Varghese K, Madhu K. Management of statin intolerance. Indian journal of endocrinology and metabolism. 2013;17(6):977.

23. Rakhshanda S, Liaw S-T, Rhee J, Rye K-A, Jonnagaddala J. Strategies to Address Statin Medication Intolerance Among Patients at Risk of Cardiovascular Disease Identified Through Electronic Health Records: A Literature Review and Pooled Analysis. Digital Health and Informatics Innovations for Sustainable Health Care Systems. 2024:132-6.

24. Nielsen SF, Nordestgaard BG. Negative statin-related news stories decrease statin persistence and increase myocardial infarction and cardiovascular mortality: a nationwide prospective cohort study. European heart journal. 2016;37(11):908-16.

25. Banda JM, Seneviratne M, Hernandez-Boussard T, Shah NH. Advances in electronic phenotyping: from rule-based definitions to machine learning models. Annual review of biomedical data science. 2018;1(1):53-68.

26. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics. 2017;22(5):1589-604.

27. Lee S, Doktorchik C, Martin EA, D'Souza AG, Eastwood C, Shaheen AA, et al. Electronic medical record–based case phenotyping for the Charlson conditions: scoping review. JMIR medical informatics. 2021;9(2):e23934.

28. Papez V, Moinat M, Payralbe S, Asselbergs FW, Lumbers RT, Hemingway H, et al. Transforming and evaluating electronic health record disease phenotyping algorithms using the OMOP common data model: a case study in heart failure. JAMIA open. 2021;4(3):ooab001.

29. Ning W, Chan S, Beam A, Yu M, Geva A, Liao K, et al. Feature extraction for phenotyping from semantic and knowledge resources. Journal of biomedical informatics. 2019;91:103122.

30. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. Journal of the American Medical Informatics Association. 2013;20(1):117-21.

31. Newbury A, Elhussein A, Gürsoy G. Multi-domain rule-based phenotyping algorithms enable improved GWAS signal. npj Digital Medicine. 2025;8(1):499.

32. MedicineWise NPS. Statin-associated muscle symptoms (SAMS) 2017 [Available from: https://www.nps.org.au/professionals/managing-lipids/statin-associated-muscle-symptoms-sams.

33. Secure Research Environment for Digital Health SREDH. ePBRN 2019 GP + Hospital linked Dataset 2024 [Available from: https://www.sredhconsortium.org/sredh-datasets/epbrn-2019-gp-hospital-linked-dataset.

34. Kadappu P, Jonnagaddala J, Liaw S-T, Cochran BJ, Rye K-A, Ong KL. Statin prescription patterns and associations with subclinical inflammation. Medicina. 2022;58(8):1096.

35. Jonnagaddala J, Chen A, Batongbacal S, Nekkantti C. The OpenDeID corpus for patient de-identification. Scientific Reports. 2021;11(1):19973.

36. Artstein R, Poesio M. Inter-coder agreement for computational linguistics. Computational linguistics. 2008;34(4):555-96.

37. Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. Journal of the American medical informatics association. 2005;12(3):296-8.

38. Seaman K, Sanfilippo F, Bulsara M, Roughead L, Kemp-Casey A, Bulsara C, et al. Predictors of ceasing or reducing statin medication following a large increase in the consumer copayment for medications: a retrospective observational study. Public Health Research and Practice. 2020;30(1).

39. Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. Accreditation and quality assurance. 2007;12(1):45-7.

40. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Frontiers in public health. 2017;5:307.

41. Sun B, Yew PY, Chi C-L, Song M, Loth M, Zhang R, et al. Development and application of pharmacological statin-associated muscle symptoms phenotyping algorithms using structured and unstructured electronic health records data. JAMIA open. 2023;6(4):ooad087.

42. Sai K, Hanatani T, Azuma Y, Segawa K, Tohkin M, Omatsu H, et al. Development of a detection algorithm for statin□induced myopathy using electronic medical records. Journal of clinical pharmacy and therapeutics. 2013;38(3):230-5.

43. Chan S, Tham M, Tan S, Loke C, Foo B, Fan Y, et al. Development and validation of algorithms for the detection of statin myopathy signals from electronic medical records. Clinical Pharmacology & Therapeutics. 2017;101(5):667-74.

44. Chang C-H, Kusama M, Ono S, Sugiyama Y, Orii T, Akazawa M. Assessment of statin-associated muscle toxicity in Japan: a cohort study conducted using claims database and laboratory information. BMJ open. 2013;3(4):e002040.

45. England NHS. Statin Intolerance Pathway 2020 2020 [Available from: https://www.england.nhs.uk/aac/publication/statin-intolerance-pathway/.

46. Li JJ, Dou KF, Zhou ZG, Zhao D, Ye P, Chen H, et al. Chinese expert consensus on the clinical diagnosis and management of statin intolerance. Clinical Pharmacology & Therapeutics. 2024;115(5):954-64.

47. Hai J-J, Wong Y-K, Wong C-K, Un K-C, Chan P-H, Siu C-W, et al. Prognostic implications of statin intolerance in stable coronary artery disease patients with different levels of high-sensitive troponin. BMC cardiovascular disorders. 2019;19(1):168.

48. Rosenson RS, Baker S, Banach M, Borow KM, Braun LT, Bruckert E, et al. Optimizing cholesterol treatment in patients with muscle complaints. Journal of the American College of Cardiology. 2017;70(10):1290-301.

49. Penson PE, Bruckert E, Marais D, Reiner Ž, Pirro M, Sahebkar A, et al. Step□by□step diagnosis and management of the nocebo/drucebo effect in statin□associated muscle symptoms patients: a position paper from the International Lipid Expert Panel (ILEP). Journal of cachexia, sarcopenia and muscle. 2022;13(3):1596-622.

50. Penson PE, Banach M. Nocebo/drucebo effect in statin-intolerant patients: an attempt at recommendations. Oxford University Press; 2021.

51. Banach M, Penson PE. Drucebo effect–the challenge we should all definitely face! Archives of Medical Science: AMS. 2021;17(2):542.

52. Warden BA, Guyton JR, Kovacs AC, Durham JA, Jones LK, Dixon DL, et al. Assessment and management of statin-associated muscle symptoms (SAMS): A clinical perspective from the National Lipid Association. Journal of clinical lipidology. 2023;17(1):19-39.

53. Cai N, Revez JA, Adams MJ, Andlauer TF, Breen G, Byrne EM, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nature genetics. 2020;52(4):437-47.

54. Patel RS, Denaxas S, Howe LJ, Eggo RM, Shah AD, Allen NE, et al. Reproducible disease phenotyping at scale: Example of coronary artery disease in UK Biobank. Plos one. 2022;17(4):e0264828.

55. Bastarache L. Using phecodes for research with the electronic health record: from PheWAS to PheRS. Annual review of biomedical data science. 2021;4(1):1-19.

56. Bigossi M, Maroteau C, Dawed AY, Taylor A, Srinivasan S, Melhem AL, et al. A gene risk score using missense variants in SLCO1B1 is associated with earlier onset statin intolerance. European Heart Journal-Cardiovascular Pharmacotherapy. 2023;9(6):536-45.

57. Sun TY, Bhave SA, Altosaar J, Elhadad N, editors. Assessing phenotype definitions for algorithmic fairness. AMIA Annual Symposium Proceedings; 2023.

58. Minissian M, Agarwal M, Shufelt C, Mehta PK, Waldman T, Lentz G, et al. Do women with statin-related myalgias have low vitamin D levels? BMC Research Notes. 2015;8(1):449.

59. Eusebi P, Speybroeck N, Hartnack S, Stærk-Østergaard J, Denwood MJ, Kostoulas P. Addressing misclassification bias in vaccine effectiveness studies with an application to Covid-19. BMC Medical Research Methodology. 2023;23(1):55.

60. Hong CS, Choi SY. Positive and negative predictive values by the TOC curve. Communications for Statistical Applications and Methods. 2020;27(2):211-24.

61. Ben-Haim Y, Dacso CC. Interpreting PPV and NPV of Diagnostic Tests with Uncertain Prevalence. Rambam Maimonides Medical Journal. 2024;15(3):e0013.

62. Manski CF. Bounding the accuracy of diagnostic tests, with application to COVID-19 antibody tests. Epidemiology. 2021;32(2):162-7.

63. Osula D, Wu B, Schesing K, Das SR, Moss E, Alvarez K, et al. Comparison of pharmacy refill data with chemical adherence testing in assessing medication nonadherence in a safety net hospital setting. Journal of the American Heart Association. 2022;11(19):e027099.

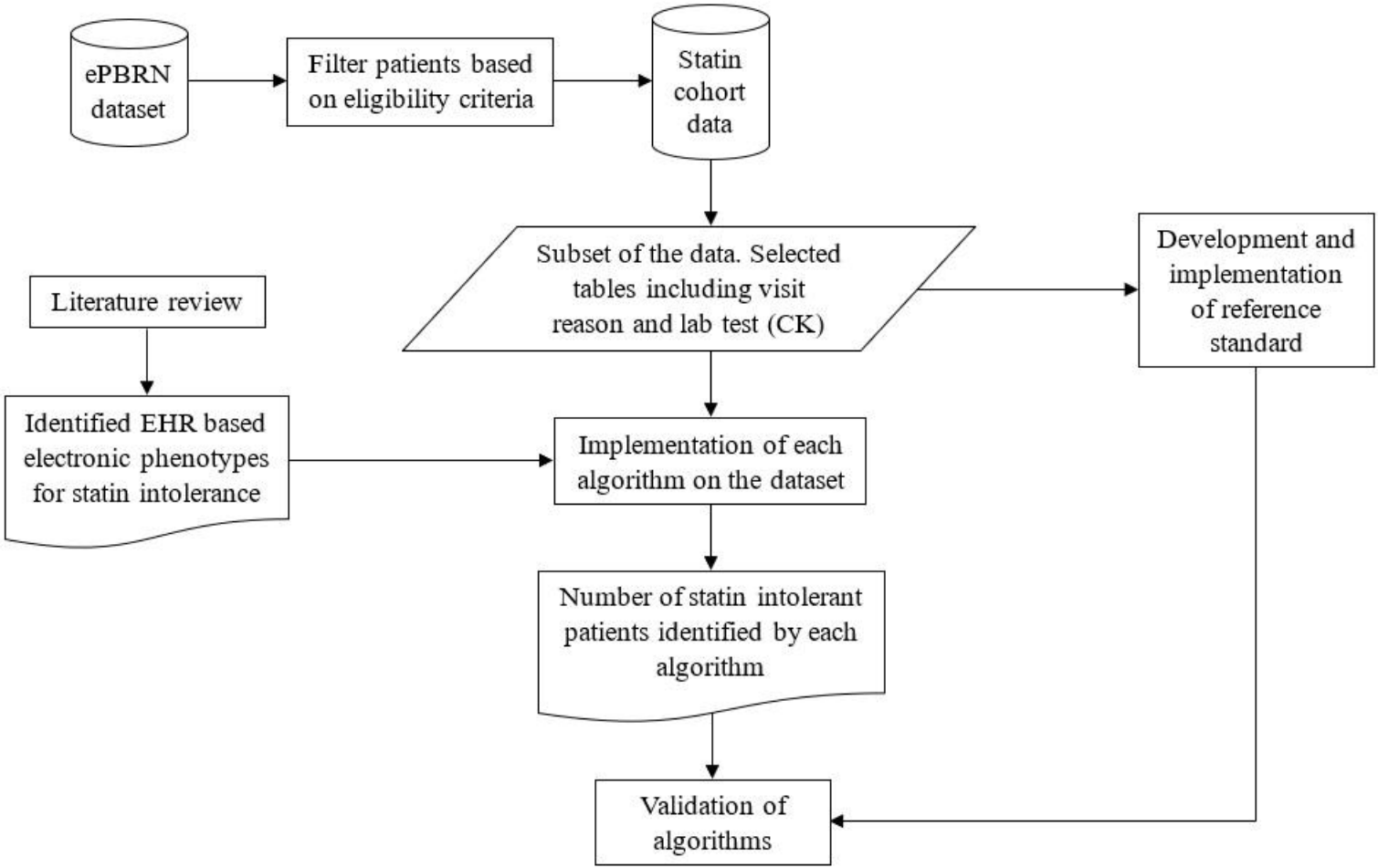

Figure 1: Overview of the study methodology

ePBRN: Electronic Practice Based Research Network; CK: creatine kinase; EHR: electronic health record

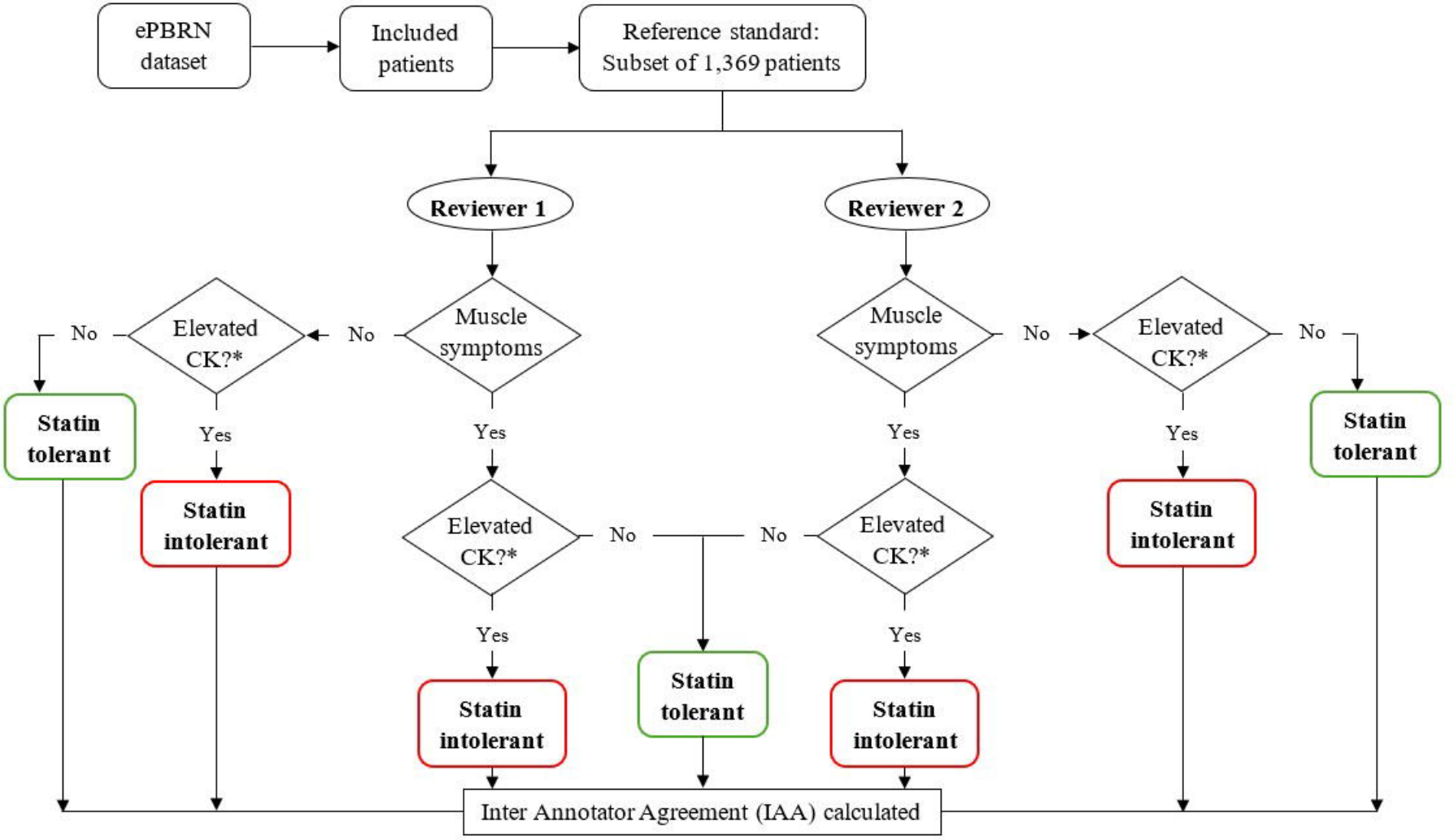

Figure 2: Development and implementation of reference standard

ePBRN: Electronic Practice Based Research Network; CK: creatine kinase; IAA: Inter Annotator Agreement

* The threshold for creatine kinase (CK) was different for patients with and without muscle symptoms

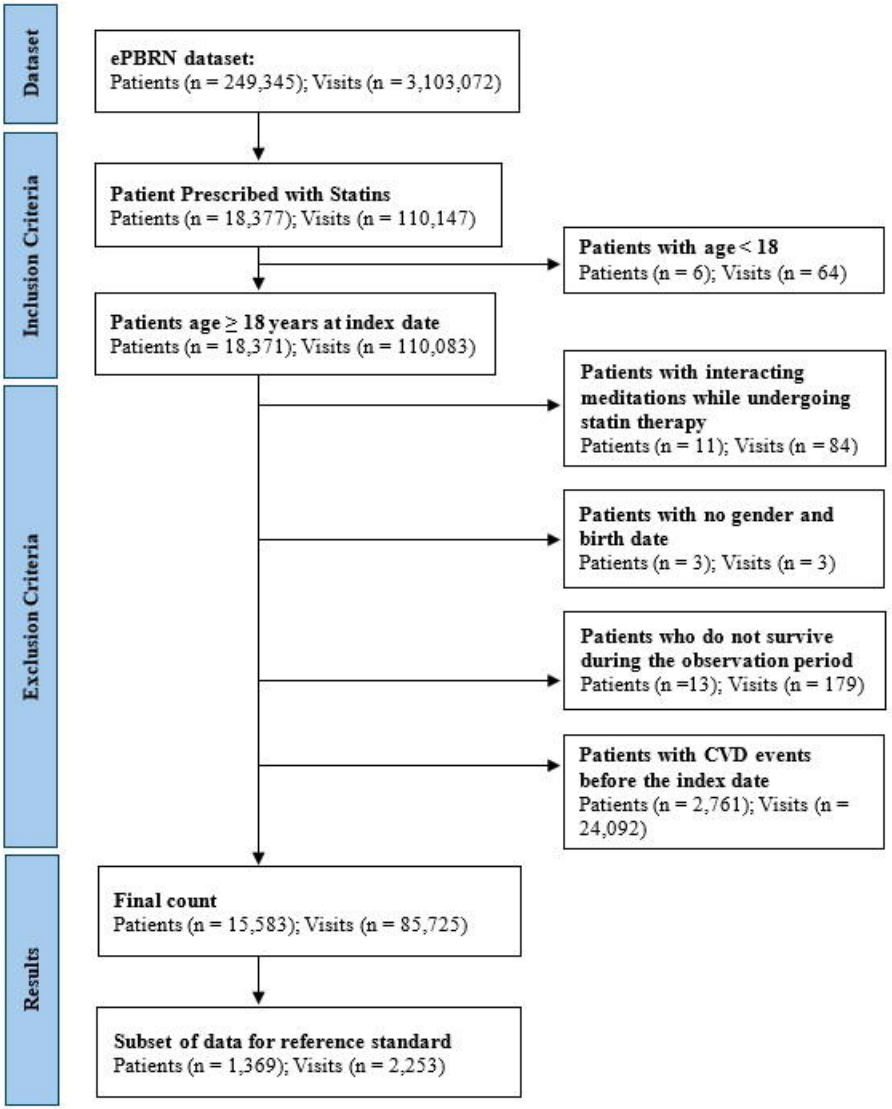

Figure 3: Identification of eligible patients from the ePBRN dataset

ePBRN: Electronic Practice Based Research Network; CVD: Cardiovascular disease

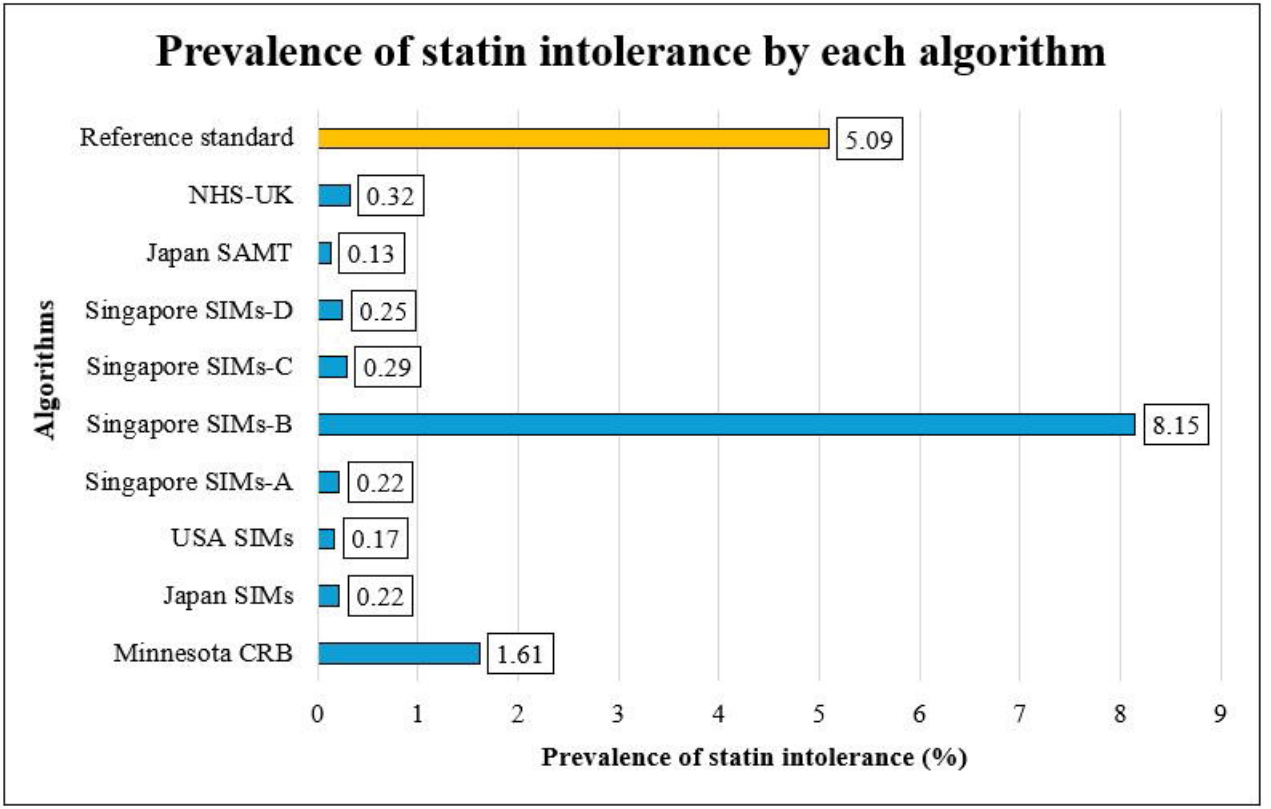

Figure 4: Prevalence of statin intolerance, as identified by each algorithm (N = 15,583)

NHS-UK: National Health Service-United Kingdom; SAMT: Statin-associated muscle toxicity; SIMs: Statin induced myopathy; USA: United States of America; CRB: Combined rule-based