

Validity of Using Prescription Medications to Classify Disease – A Retrospective Observational Study Using Routinely Collected Electronic Health Records from the UK

Christian Schnier ¹, John Busby¹, Aziz Sheikh², Jennifer K Quint ³, David B Price ^{4–6}, Liam G Heaney^{1,7}

¹School of Medicine, Dentistry and Biomedical Sciences, Queen's University, Belfast, UK; ²Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK; ³School of Public Health, Imperial College London, London, UK; ⁴Optimum Patient Care Global Ltd, Cambridge, UK; ⁵Observational and Pragmatic Research Institute, Singapore, Republic of Singapore; ⁶Centre of Academic Primary Care, Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK; ⁷Belfast Health & Social Care NHS Trust, Belfast, UK

Correspondence: Christian Schnier, School of Medicine, Dentistry and Biomedical Sciences Wellcome Wolfson Institute for Experimental Medicine, Queen's University Belfast, 97 Lisburn Road, Belfast, United Kingdom, Email c.schnier@qub.ac.uk

Background: Epidemiological studies rely on valid classifications of patients' disease status. However, in the absence of perfect information on every patients' health status, researchers use proxy information with variable and often unknown validity.

Methods: To investigate the validity of using prescription records for disease classification, we conducted a retrospective observational study on a UK-wide database of medical prescriptions and clinical records (Optimum Patient Care Research Database). We used electronic health records of 25,000 randomly selected patients for each year between 2004 and 2020 (total N=425,000) and compared disease classification of 18 different chronic conditions based on clinical records for a period of three years (gold standard) with disease classification based on prescription records for the same period. We then used logistic regression to analyse if positive and negative predicted values (PPV and NPV) were associated with known predictors of disease.

Results: Results showed large variations in PPV ranging from 8% (heart failure) to 94% (all type diabetes) and smaller variations in NPV ranging from 96% (anxiety) to 100% (Type 1 diabetes). Age, sex, ethnicity, and year but not socio-economic status were associated with variations in validity, especially in classifying dementia, diabetes, and depression.

Discussion: Varying validity can partly be explained by different (stratum-specific) prevalence of disease. Additionally, conditions like heart failure can be treated with medication that can also be prescribed for other conditions or can be treated without medication. However, varying validity can also be attributed to imperfect clinical records, which we used as gold standard. As a consequence of low validity, the apparent prevalence based on using prescription records was between 1.3 times lower (all-type diabetes) and up-to 11 times higher (heart failure) than the true prevalence based on the clinical records.

Conclusion: Studies using prescription data to classify disease status run a substantial risk of misclassification bias.

Plain Language Summary: In medical research, we often group people into those who have a disease and those who are healthy. However, in large studies we usually cannot medically examine every person, so we use readily available information, medical health records, to estimate the disease status. The problem with estimating the disease status from these records, though, is that the information is incomplete. To study how well we can estimate disease status from prescription records, we compared disease status of 425,000 patients based on clinical records with disease status of the same patients based on prescription records. For the 18 health conditions we included, we found wide variability. For most health conditions we found that if we did not find a prescription record, we could be reasonably confident that the patient indeed did not have the health condition. However, we also found that if we did find a prescription for a health condition, this did not necessarily mean that this patient indeed had a clinical diagnosis. We found that some of these failures to group patients based on prescription records can be explained by the frequency of disease and by different treatment options. We found that using prescription records to predict disease could lead to an estimate of up-to 11 times higher burden of



disease than actually true. Finally, we demonstrated that comparisons of disease burden between different patient groups (eg, between males and females) can go wrong if prescription records are used to estimate the disease status.

Keywords: electronic health records, misclassification bias, validation study, diagnostic code lists, Optimum Patient Care Research Database

Introduction

Worldwide use of linked routinely collected electronic health records (EHRs) in medical research has increased considerably. In the UK, EHRs have extensively been used throughout the COVID-19 pandemic to conduct research into the health of the nation.¹ EHRs are primarily used, and the system is optimised, to support patient care;² with research making secondary use of the data to improve public health, policymaking, care planning and safety monitoring.³ In the UK, EHRs have additionally been used since 2004 to calculate payments to General Practitioners under the Quality and Outcomes Framework (QOF) (abolished in Scotland in 2016). QOF is a financial incentive to improve health service quality by linking up to 25% of general practitioners' income to achievement of publicly reported quality targets for several chronic conditions.⁴ By linking EHRs to payments, the introduction of QOF might have improved recording procedures at practice level for conditions included under QOF, with little effect on conditions that were outside of QOF.⁵

Advantages and limitations of using linked EHRs for medical health research have been previously studied in-depth, with concerns raised about the quality of data and potential for bias. A common feature of EHR studies is a requirement to identify patients as diseased or non-diseased either to determine outcome, exposure, or study covariates. Misclassification in studies using EHR relates to the problem that, based on information from the EHRs, patients in the study population are wrongly identified as diseased or non-diseased (outcome misclassification) or as exposed or non-exposed to a specific risk factor, for example, the smoking status of a patient (exposure misclassification). In a recent review, Young et al identified eight sources for misclassification frequently encountered in studies using EHRs and investigators were urged to carefully evaluate and rigorously address this potential source of bias in studies.⁶ According to Young et al, one reason for misclassification is that, in the absence of a comprehensive data linkage including all possible health-related data sets, information that is potentially crucial for valid classification is missing.⁶ In a study using EHRs of all medical prescriptions to classify the health status of patients, patients would be misclassified for whom medications were contra-indicated or who bought the medication without a prescription. Another point of concern raised by Young et al was that the performance of EHR-based clinical prediction algorithms may vary widely between different health systems and that temporal changes in EHR data elements recorded may produce systematic differences in classification and/or missingness over time.⁶ Misclassification can severely bias medical health research and estimates of validity should be included and referenced in any publication using EHRs.⁷ A large increase in apparent prevalence using only prescription records for classification compared to the true prevalence using medical records has been shown in a study from New Zealand, where the prevalence of multimorbidity was 7.9% using past hospital discharge data, and 27.9% using past pharmaceutical dispensing data.⁸

Validity in classification can be expressed in several ways. For medical research, arguably the most important indicators are the positive and negative predictive values (PPV and NPV, respectively), which are the proportions of positive and negative classifications using the EHRs that are truly positive and truly negative, respectively. However, validity of classification in medical studies using EHRs is often not reported and complicated to assess. To address this gap, we have analysed GP diagnosis and GP prescription records in a UK-wide dataset over a 20-year period for a wide range of conditions. The primary aim of this study was to assess the validity of using prescription records for classifying disease and additionally to identify predictors for misclassification. GP diagnostic records were used as the gold standard and analysed the validity of using prescription records to classify disease in the study population.

Materials and Methods

Study Design

We undertook a retrospective validation study, following the design of an evaluation of medical tests for classification and prediction.

Data Resource

We extracted EHRs relating to diagnosis and prescriptions from the Optimum Patient Care Research Database (OPCRD). OPCRD is a primary care EHR database, which, in 2025, was holding 24 million patient-records from 1014 practices from England (922 practices), Northern Ireland (1 practice), Scotland (57 practices), and Wales (34 practices).⁹ The database is broadly representative of the UK population in terms of age, sex, ethnicity, and socio-economic status.¹⁰ Data for OPCRD were extracted from all major clinical software systems used within the UK and across all four coding systems (Read version 2, Read CTV3, SNOMED DM+D and SNOMED CT).

Study Population and Study Period

The study population was a random selection of 25,000 patients for each year from 2004 to 2020, who were registered with a GP in the OPCRD research database ([Supplementary Figure 1](#)). Selected patients were only included once over the study period. We chose 2004 as a starting point because of changes in data quality associated with the introduction of QOF in the UK in 2004. We chose to select 25,000 patients/year as a fixed proportion of the general adult population of England (approximately 0.55%) rather than a complete random selection to increase precision of the estimates and to reduce bias from annual variations in GP visits (eg, due to COVID 19). For each patient in the study population, we created a random index date that allowed a 3-year observation window ([Figure 1](#)); thereby reducing the potential effect of seasonality in morbidity, prescription and recording accuracy. We chose the 3-year observation window to compare the outcome of a hypothetical RCT that uses a lookback of diagnostic records to classify disease status at the start of the study (eg,¹¹) with the outcome of a hypothetical RCT that uses a lookback of prescription records; to study the effect of a shorter period, we included a sensitivity analysis. We excluded children under the age of 16 at index date and patients with flags for poor data quality in the data set (eg no valid year of birth, no joined/leaving dates available).

Patient Comorbidities

Comorbidities included in this validation study were diseases that are common, non-communicable, are pharmacologically treated and where medications are reasonably specific to treating the condition. These included diseases of the respiratory tract (asthma and chronic obstructive pulmonary disease (COPD)); neurological and psychiatric conditions (anxiety, depression, migraine, dementia, and epilepsy), (congestive) heart failure, diabetes, gastro-oesophageal reflux disease (GORD), hypothyroidism, and bone health (osteoporosis and osteopenia). Several conditions were additionally combined into disease categories: Type 1 and Type 2 diabetes into general diabetes; osteoporosis and osteopenia into reduced bone density; anxiety and depression into mood and anxiety disorders; and asthma and COPD into obstructive lung diseases.

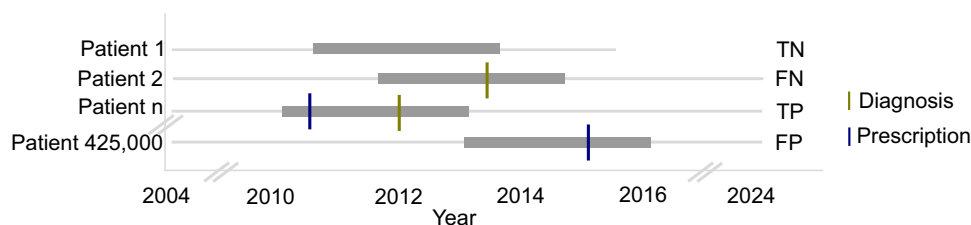


Figure 1 Study schematic for selection and classification of the study population.

Abbreviations: TN, True Negative; FN, False Negative; TP, True Positive; FP, False positive.

Classification – Index Test

The index test was a code in the GP prescription records indicative of the comorbidity (eg, a code for a prescription of an anti-depressant in the classification of clinical depression). We selected prescription codes using a list of codes from the British National Formulary (BNF) translated to Read and SNOMED CT code together with lists of codes from Health Data Research UK (HDRUK), Bennett Institute for Applied Data Science (Open Safely) and other trusted sources ([Supplementary Table 1](#)). We classified patients as test positive if their patient record included at least one code indicative of a relevant prescription during the period. We classified patients as test negative if their patient record did not include any code indicative of a relevant prescription during the period.

Classification – Gold Standard

The gold standard was a code in the GP records that was indicative of the comorbidity (eg, a code for a diagnosis of Agitated depression in the classification of clinical depression). Code lists mainly consisted of diagnostic codes but also included relevant codes for disease history, examination, procedures and administration (eg, a code for “On depression register”); codes for prescribed medications were not included. We selected relevant codes (Read version 2, Read CTV3, and SNOMED CT) from several sources including the Phenotype Library from HDRUK and Open Safely; except for osteopenia, for which we could not find any published coding lists, these coding lists have previously been applied in peer-reviewed research studies. ([Supplementary Table 1](#)). We classified patients as gold standard positive if their patient record included at least one code indicative of a diagnosis during the 3-year period. We classified patients as gold standard negative if their patient record did not include any code indicative of a diagnosis during the period. By design, there was no temporal order between prescription and the relevant diagnostic code; both codes could be up-to 3 years apart. We included a sensitivity analysis to study the effect of a longer period for the gold standard classification.

Covariates

Covariates included in the analysis included age at index date; sex; year of index date; ethnicity; deprivation; GP practice; a data quality index provided by OPCR, clinical commissioning group and data provider. For most diseases, age at index date was grouped into young (16–29), mid (30–69) and elderly (>69). In the validation of dementia diagnoses, we re-grouped age into young (16–69), mid (70–89) and elderly (>89) to avoid problems with statistical disclosure. To classify ethnicity, we interrogated the complete OPCR dataset for GP records associated with the ethnicity of the patient in the study population. We derived a list of ethnicity-related codes from the Office of National Statistics.¹² To avoid problems with statistical disclosure, ethnicity was grouped into “White”, “Other” and “Missing”. To classify deprivation, we used country-specific Multiple Deprivation deciles (IMD)¹³ for the location of the GP practice, grouped into high (1–3), mid (4–7) and low (8–10).

Analysis

For each disease we calculated Sensitivity (SE), Specificity (SP), PPV, NPV, Area under the receiver operating characteristic curve (ROC), true period prevalence (TPP) and apparent period prevalence (APP). TPP was defined as the proportion of patients with at least one disease-associated diagnostic code during the three-year period (ie gold standard test) and APP as the proportion of patients with at least one prescription-associated code during the same period (ie index test). ROC was calculated as the mean value of SE and SP.¹⁴ To identify predictors for misclassification, we calculated test validity for different strata of the study population based on the covariates; for example, we separately calculated the validity in males and in females. To further study associations of PPV and NPV with the covariates, we fitted two multilevel logistic regression models for every condition in two strata of the study population. (1) Using the population of patients with at least one disease-related prescription, we fitted a model for the likelihood of finding at least one diagnostic code (PPV). (2) Using the population of patients with no disease-related prescription, we fitted a model for the likelihood of finding no diagnostic code (NPV). To analyse statistical significance of the association of covariates with NPV and PPV, we conducted likelihood ratio tests. We included the practice identifier in every model as random effect because observations from patients of the same GP practice were most likely correlated. We additionally conducted

two sensitivity analyses: (a) To analyse the effect of using a shorter period, we studied the validity of prescription records to classify COPD and dementia in a 1-year study period. (b) To study the effect of GPs' coding a diagnosis prior to the randomly selected index date, we compared the patients' prescription records for COPD and dementia in the three-year period post index date to the patients' diagnostic records in the period from GP registration to three-year post index date.

Data management and analysis was conducted using Microsoft SQL Server and Stata (V18), respectively. All random selections were created using MS SQL random number generation. Classification of both index test and gold standard were conducted unblinded by the main author (cs).

Results

The study population comprised 425,000 randomly selected patients from 340 GP practices in OPCR, of whom 51% were female, 20% were over the age of 70, approximately 36% lived in areas of higher deprivation and more than 95% were resident in England (Table 1). Eighty-five percent of the study population for whom information on ethnicity was available were classified as white.

Table 1 Summary Statistics

Variable	N (%)
Patients (total N)	425,000 (100%)
GP practices (total N)	340 (100%)
Sex	
Female	216,706 (51.0%)
Male	208,285 (49.0%)
Year	
2004–200	150,000 (35.3%)
2010–2015	150,000 (35.3%)
2016–2020	125,000 (29.4%)
Age	
16–29	117,674 (27.7%)
30–69	221,395 (52.1%)
70+	85,931 (20.2%)
Ethnicity	
White	257,981 (60.7%)
Other	43,868 (10.3%)
Missing	123,151 (29.0%)
Deprivation (IMD Deciles)	
High (1–3)	154,551 (36.6%)
Mid (4–7)	143,045 (33.9%)
Low (8–10)	124,815 (29.5%)
Region	
East of England commissioning region	89,637 (21.1%)
London commissioning region	16,725 (3.9%)
Midlands commissioning region	46,543 (11.0%)
North East and Yorkshire commissioning region	86,111 (20.3%)
North West commissioning region	47,593 (11.2%)
South East commissioning region	90,485 (21.3%)
South West commissioning region	36,176 (8.5%)
Scotland	7,734 (1.8%)
Wales	3,027 (0.7%)
Data Quality Index (RA)	
Low	103,303 (24.3%)
High	321,697 (75.7%)

Diseases with the highest period prevalence (>5%) over the three-year observation period based on the diagnostic codes included obstructive lung diseases, mental health conditions and all type diabetes (Figure 2). Due to low validity, the APP based on medical prescriptions varied from the TPP. Diseases with large increased APP (>500% increase in relative value) included COPD, epilepsy, Type 1 DM, the combined reduce bone density diseases, heart failure and GORD. Diseases with lower APP compared to TPP included dementia (200% decrease) and all Type diabetes (30% decrease).

As expected, estimates for SE, SP, PPV and NPV varied between the different diseases (Figure 2). Diseases with >90% SE and SP, meaning that there's a >90% chance that a patient with a diagnostic record has at least one relevant prescription (SE>90%) and a >90% chance that a person without a diagnostic record has no relevant prescription (SP>90%), included obstructive lung diseases, asthma, epilepsy, Type 1 DM and hypothyroidism. Those with lower SE (<90%) but still high (>90%) SP included anxiety, dementia, migraine, all type DM, Type 2 DM and all disease associated with reduced bone density. Diseases with high SE (>90%) but lower SP included depression, heart failure,

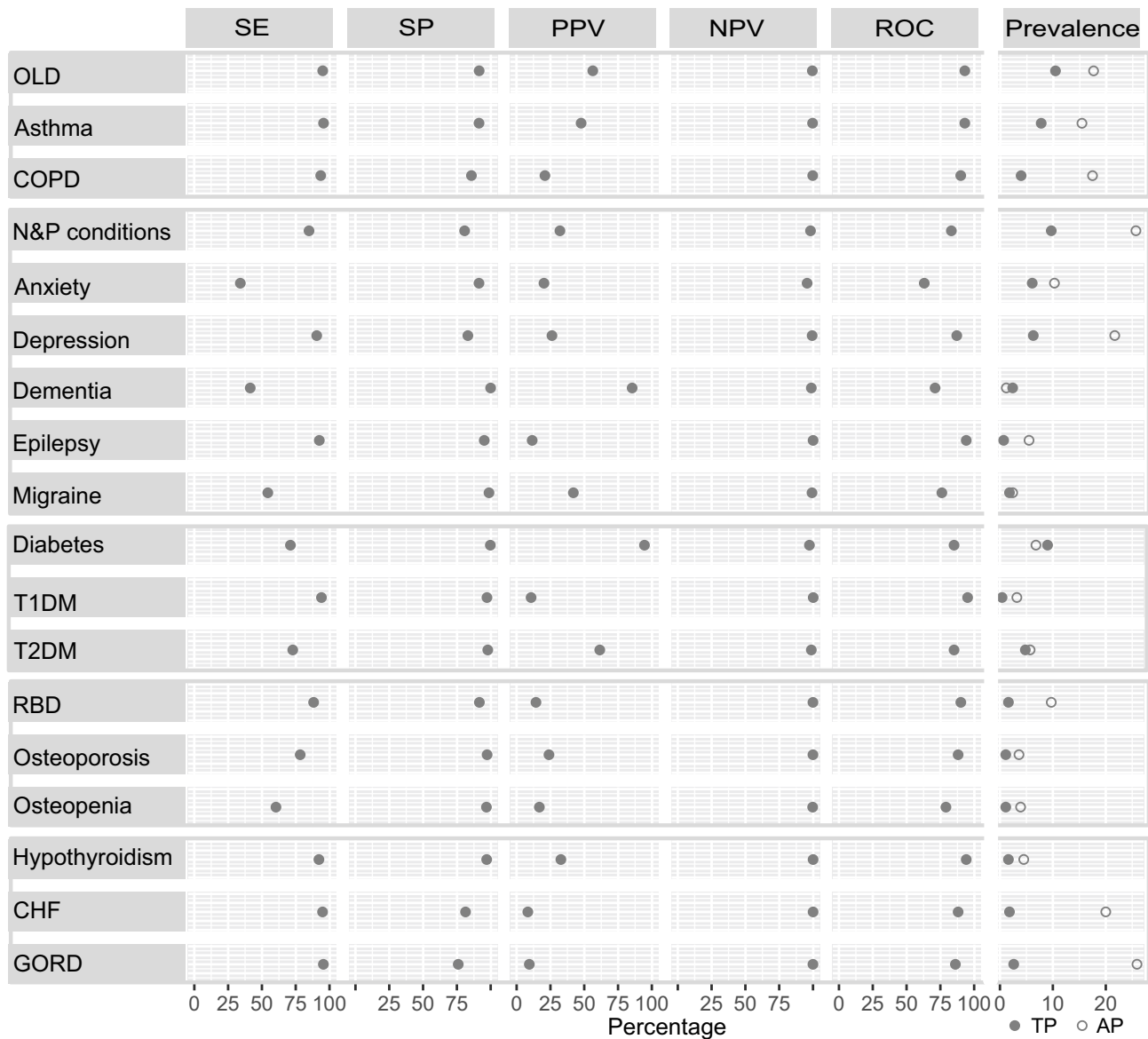


Figure 2 Estimates for the validity of using prescription records to classify disease.

Abbreviations: OLD, Obstructive lung disease; COPD, Chronic obstructive pulmonary disease; N&P conditions, neurological and psychiatric conditions; RBD, Reduced bone density; CHF, Heart failure; GORD, Gastro-oesophageal reflux disease; AP, Apparent prevalence; TP, True prevalence.

COPD and GORD. Finally, only the combined mental health group had lower SE and SP (<90%). The ROC value was above 90% for only seven of the conditions, it was lowest for anxiety (63%).

Most diseases had a <50% chance that a patient with at least one relevant prescription record has had a diagnostic record (PPV<50%), combined with a >98% chance that a patient without any relevant prescription diagnostic record also had no diagnostic record (NPV>98%) (Figure 2). Diseases with >50% PPV combined with high NPV (>98%) included obstructive lung diseases, dementia and Type 2 DM. Anxiety was the only disease associated with low PPV (20%) and low NPV (96%).

For all diseases, estimates for PPV and NPV were significantly associated with gender, ethnicity, age and year at the start of the period; associations of PPV and NPV with socio-economic status were mostly not statistically significant (Figures 3–7). Diseases with large associations include dementia, with higher validity in the year 2016–2020 (PPV=93% and NPV=98%) and lower validity in the year 2004–2009 (PPV=67% and NPV=99%) and in people older than 90 (PPV=86% and NPV=87%). For the classification of anxiety validity was comparably low in all strata, with estimates

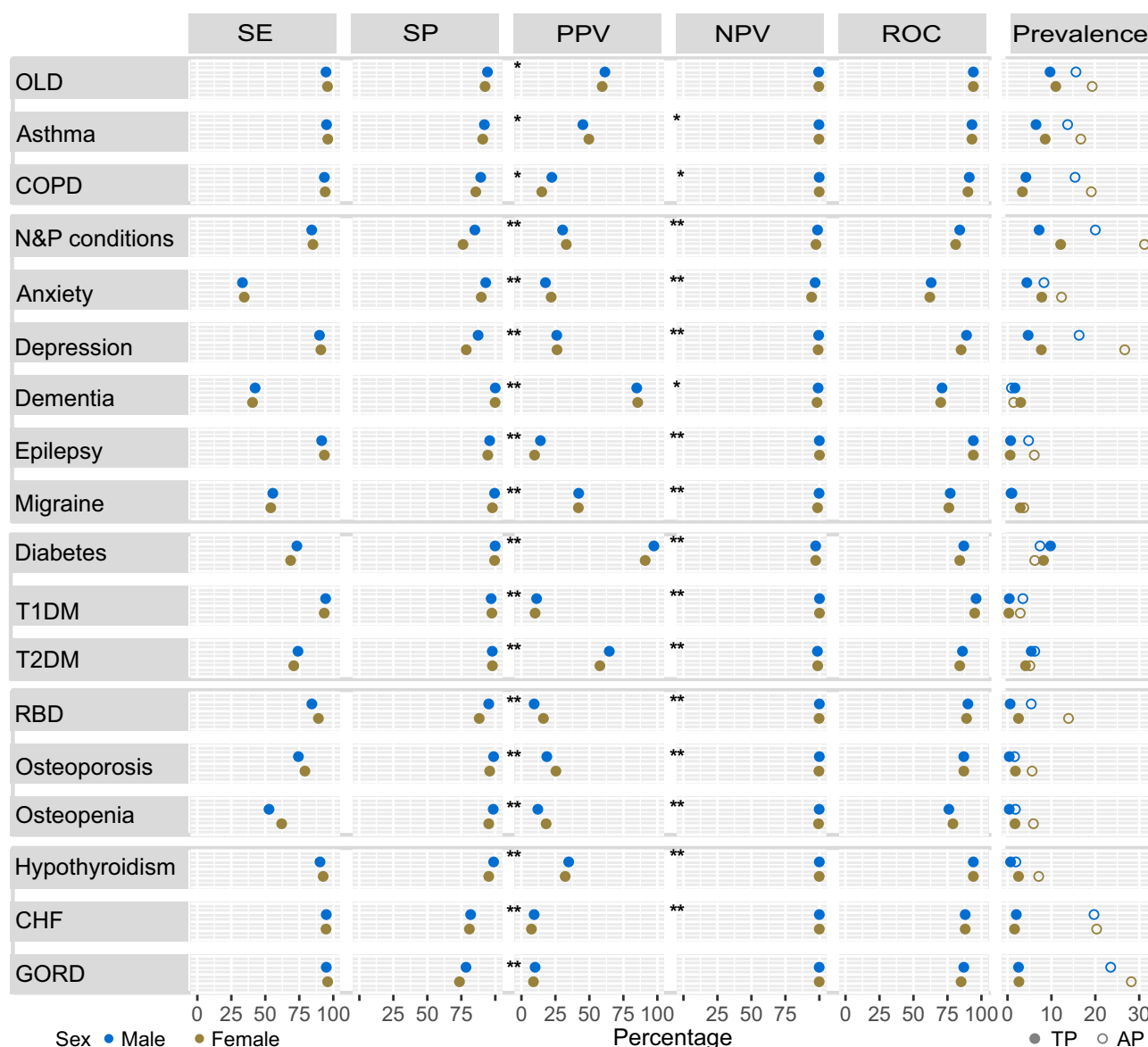


Figure 3 Estimates for the validity of using prescription records to classify disease by sex. *P<0.05; **P<0.01. **Abbreviations:** OLD, Obstructive lung disease; COPD, Chronic obstructive pulmonary disease; N&P conditions, neurological and psychiatric conditions; RBD, Reduced bone density; CHF, Heart failure; GORD, Gastro-oesophageal reflux disease; AP, Apparent prevalence; TP, True prevalence.

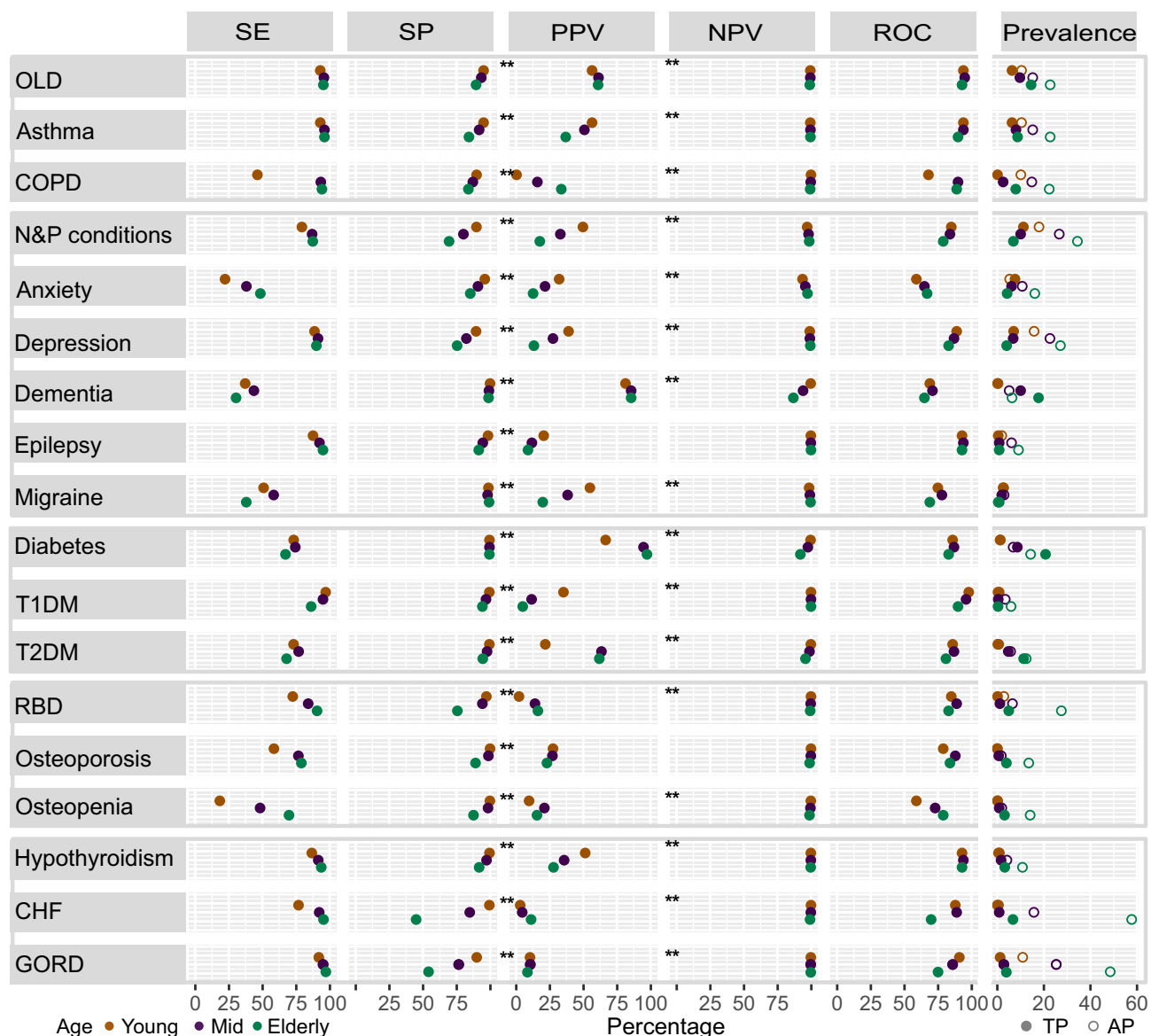


Figure 4 Estimates for the validity of using prescription records to classify disease by age group. * $P<0.05$; ** $P<0.01$.

Abbreviations: OLD, Obstructive lung disease; COPD, Chronic obstructive pulmonary disease; N&P conditions, neurological and psychiatric conditions; RBD, Reduced bone density; CHF, Heart failure; GORD, Gastro-oesophageal reflux disease; AP, Apparent prevalence; TP, True prevalence.

ranging from PPV=32% and NPV=94% in patients aged 16–40 to PPV=13% and NPV=97% in patients older than 70. For the classification of COPD estimates for PPV ranged from 0.3% in patients aged 16–40 to 34% in patients older than 70; the NPV was 99.5–100% for all strata.

The sensitivity analysis showed that estimates of validity of using prescription medications to classify disease varied with the length of the period (Table 2). Shorter periods were associated with a decrease in PPV and an increase in NPV. Replacing the 3-year period with an “ever” diagnosis as a gold standard was associated with increased PPV and a decreased NPV.

Discussion

In our study, population, for every condition the NPV was above 95%, indicating a reasonable validity of using the absence of a relevant prescription to classify patients without disease. However, in only four out of 18 conditions (dementia, Type 2 DM, any diabetes and obstructive lung diseases) was the PPV above 50%, indicating low validity of

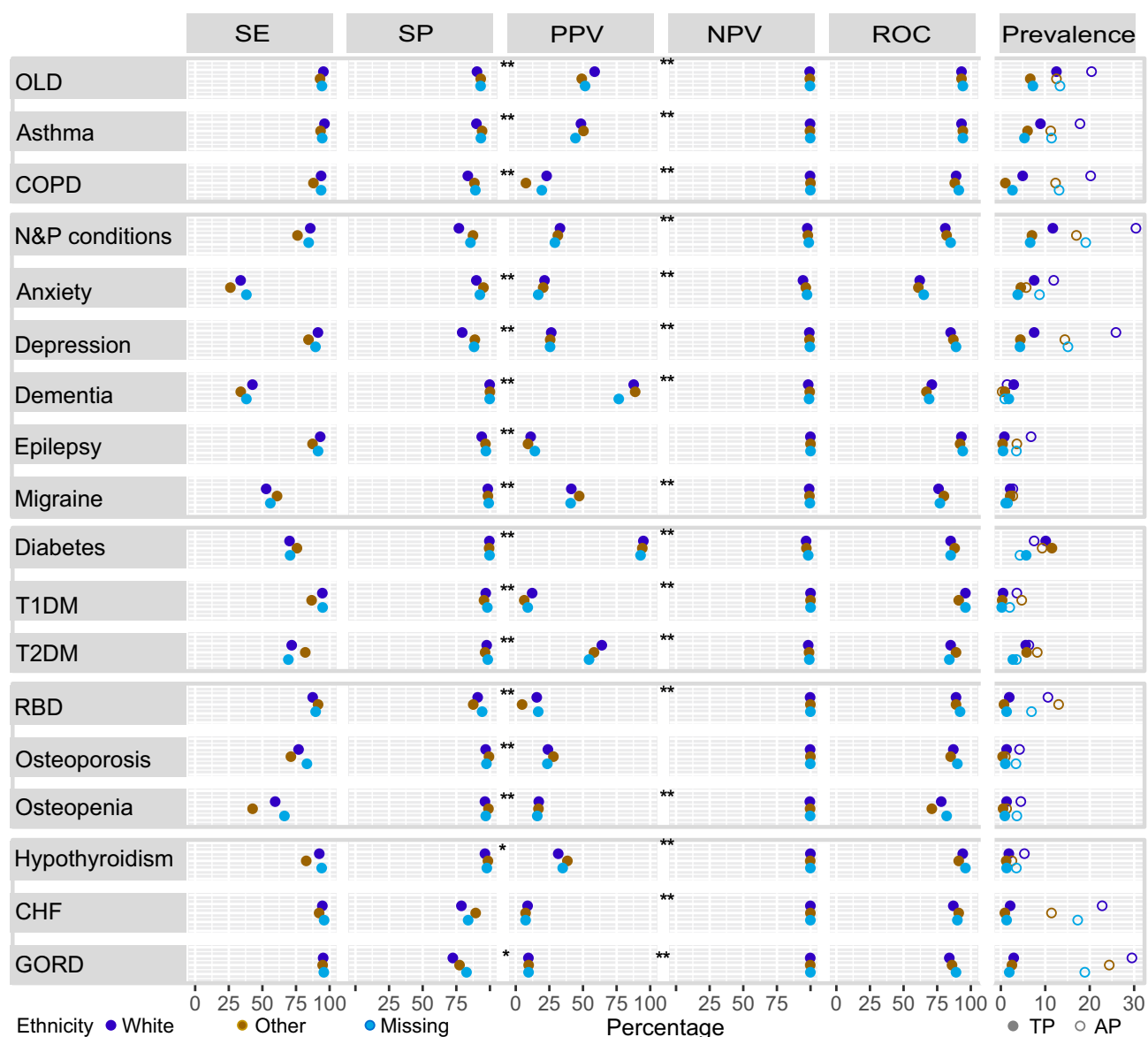


Figure 5 Estimates for the validity of using prescription records to classify disease by ethnicity. * $P < 0.05$; ** $P < 0.01$.

Abbreviations: OLD, Obstructive lung disease; COPD, Chronic obstructive pulmonary disease; N&P conditions, neurological and psychiatric conditions; RBD, Reduced bone density; CHF, Heart failure; GORD, Gastro-oesophageal reflux disease; AP, Apparent prevalence; TP, True prevalence.

using the presence of a relevant prescription to classify patients with a specific disease. The low and heterogeneous validity of using prescription records to classify diseases was expected. For example, the low PPV of using anti-epileptic drugs to classify epilepsy (13%) observed in our study was similar to results from other studies reviewed by Mbizvo.¹⁵ Combining diseases with similar prescription code lists, eg, combining asthma and COPD into an obstructive lung disease group generally increased validity of classification; indeed, in the absence of information on dosage and frequency, asthma (PPV: 48%) and COPD (PPV: 21%) could not reliably be classified using prescriptions.

The observed low PPV and high NPV can partly be explained by the low TPP of the conditions in the study population, which, for all conditions, was below 10%. This is because predictive values are a function of prevalence, with increasing prevalence associated with higher PPV and lower NPV. For example, the association of PPV with ethnicity in the classification of COPD was mostly driven by variations in prevalence, the 3.5 times higher PPV in patients of white ethnicity was not associated with variations in SE and SP. Similarly, in the classification of anxiety the higher SE and lower SP in patients of white ethnicity was not associated with variations in PPV and NPV. However, for some

Table 2 Results from the Sensitivity Analysis

		SE [%]	SP [%]	PPV [%]	NPV [%]	TP [%]
COPD	Original analysis	93.4	85.7	21.0	99.7	3.9
	1 year FU	92.2	89.9	19.7	99.8	2.6
	Ever diagnosis	81.3	86.1	24.1	98.8	5.2
Dementia						
	Original analysis	41.3	99.8	85.3	98.6	2.4
	1 year FU	41.9	99.7	62.0	99.4	1.1
	Ever diagnosis	37.8	99.9	94.0	98.2	2.9

Notes: Original analysis: Analysis using a 3-year period to classify both index test and gold standard; 1 year FU: Analysis using a 1-year period to classify both index test and gold standard; Ever diagnosis: Analysis using a period from the start of GP registration to 3 year post index date to classify gold standard while using a 3-year period to classify index test.

Abbreviations: COPD, Chronic obstructive pulmonary disease; NPV, Negative predictive value; SE, Sensitivity; SP, Specificity; PPV, Positive predictive value; TP, True prevalence.

conditions, for example, in the classification of heart failure and GORD (PPV<10%), the low observed PPV can also be explained by low SP and high SE. These diseases are treated with medications that are not specific for the conditions and for which medications could be given preventatively.

As highlighted by Young et al, temporal changes in EHR data elements recorded may produce systematic differences in classification and/or missingness over time.⁶ Indeed, we observed variations in validity over time, but no overall trend of increasing or diminishing validity (Figure 6). For some part of the study period asthma, COPD, dementia, depression, diabetes, epilepsy, heart failure, hypothyroidism and osteoporosis have been included under QOF, which has been associated with more reliable and complete recording by the GP.¹⁶ Furthermore, during the study period different coding systems have been implemented (READ V2, READ CTV3 and SNOMED CT), which made it necessary to include several code lists from different sources with varying validity to classify both gold standard and index test (Table 2). Finally, all health services in the UK were disrupted by the COVID-19 pandemic from March 2020 onwards.

We were able to take advantage of a long study period and a large study population. The demography in terms of age and sex was broadly representative of the UK population with a 6% over representation of patients living in areas of high deprivation and a 10% over representation of patients living in England (ONS statistics). In terms of ethnicity, the study population compared well with results from the English and Welsh Census in 2011 (86%) and 2021 (81.7%). However, classification of ethnicity in GP records has been described as unreliable, and almost 30% of the patients in the study population could not be classified. Similarly, classification of socio-economic status of patients using the postcode area of the GP practice was unreliable, which could explain the unexpected low association of IMD with prevalence. IMD was also not significantly associated with validity; however, this can also be explained by correlations between the random practice-id variable and IMD, which was the same for all patients of the same practice. In our study, we could also take advantage of the large data repositories for health-codes used in research and in the management of QOF. We could demonstrate that while using the absence of a prescription record to classify freedom of disease was reasonably valid for medical research, using the presence of a prescription record to classify disease was more problematic. We could demonstrate that because the validity of the classification was stratum-specific, studies using incomplete information for disease classification run a substantial risk of misclassification bias.

We were limited in our study by uncertainty and heterogeneity of the validity of the gold standard. One assumption in our study was that patients can reliably be classified using General Practice diagnostic codes in the patient record. However, particularly in chronic conditions, such as dementia, patient records will not necessarily carry a diagnostic code with every repeat prescription. This has also been demonstrated in other settings, eg, in secondary care records, where a study showed that a large proportion of patients with chronic conditions such as chronic kidney disease identified via laboratory measurements did not have a corresponding diagnostic code.¹⁷ Therefore, the randomly selected 3-year observation period might have included repeat prescriptions, with the initial diagnosis having been coded prior to the



Figure 6 Estimates for the validity of using prescription records to classify disease by year. * $P < 0.05$; ** $P < 0.01$.

Abbreviations: OLD, Obstructive lung disease; COPD, Chronic obstructive pulmonary disease; N&P conditions, neurological and psychiatric conditions; RBD, Reduced bone density; CHF, Heart failure; GORD, Gastro-oesophageal reflux disease; AP, Apparent prevalence; TP, True prevalence.

start of the observation period. Indeed, as expected, when validating the use of prescriptions to predict dementia and COPD using an ever-diagnosis as a gold standard, we found increased PPV in combination with decreased NPV. When using a shorter 1-year period, we found decreased PPV with increased NPV. However, using an ever-diagnosis as gold standard introduced additional uncertainty and heterogeneity in the quality of historical GP records and assume that all diseases in the study were associated with long term/lifelong treatment, which, for example, in calculating validity for migraine classification, was not plausible. We could have improved validity of classification by additional specialist review of published code lists and/or development of different classification algorithms.¹⁸ However, most diseases in our study were classified using limited different codes, which were common in all reviewed published code lists; eg three out of 139 codes used for the classification of COPD identified 52% of cases. For studies with access to linked diagnostic records and prescription records it is not unusual to classify patients' disease status using both records in parallel interpretation (patients are classified disease positive if they either have a diagnostic code or a prescription record). Using this algorithm in our study would, by definition, have increased the PPV to 100%; however, it would assume that all

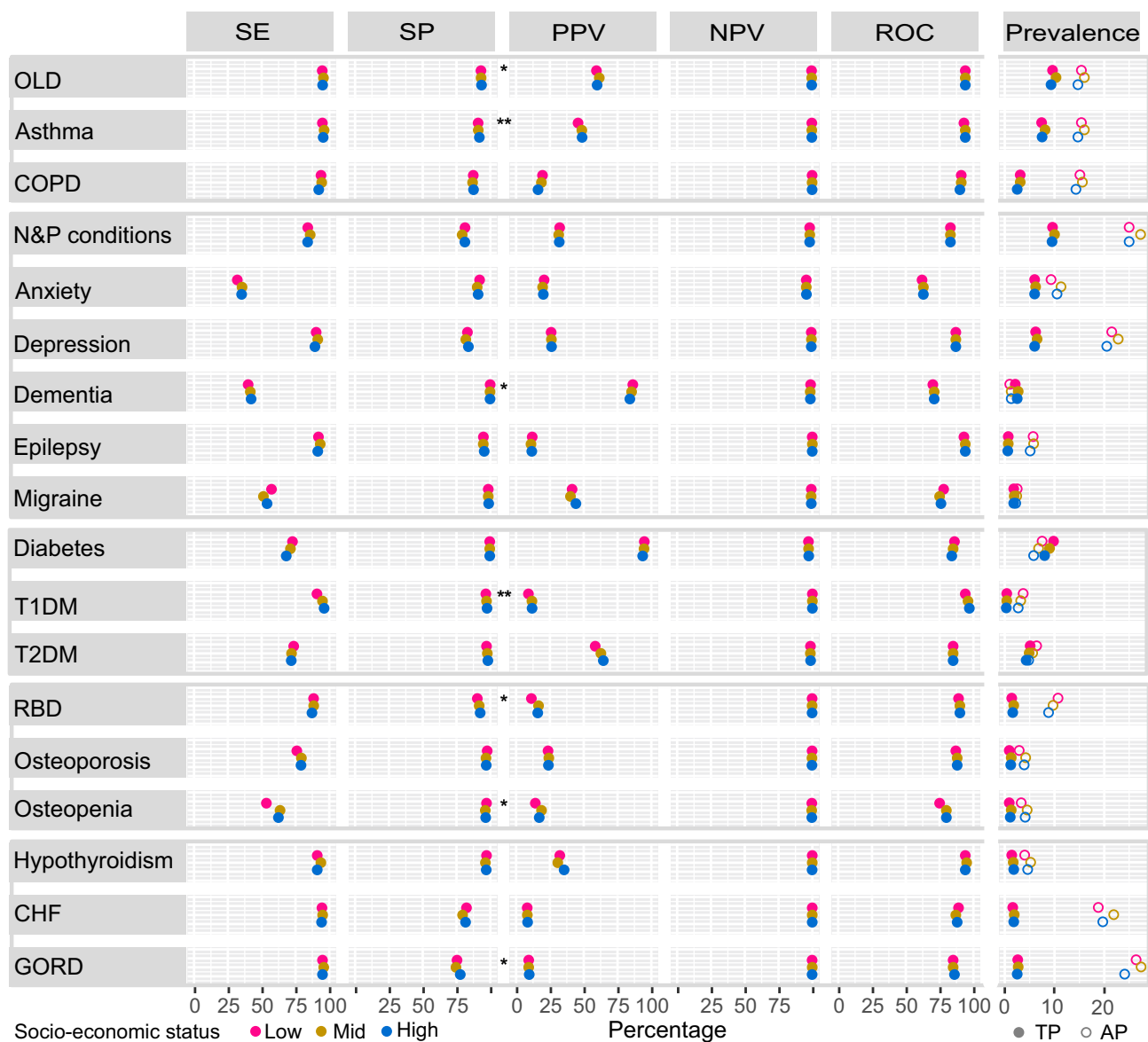


Figure 7 Estimates for the validity of using prescription records to classify disease by socio-economic status. * $P < 0.05$; ** $P < 0.01$.

Abbreviations: OLD, Obstructive lung disease; COPD, Chronic obstructive pulmonary disease; N&P conditions, neurological and psychiatric conditions; RBD, Reduced bone density; CHF, Heart failure; GORD, Gastro-oesophageal reflux disease; AP, Apparent prevalence; TP, True prevalence.

prescriptions are specific for a condition, which, at least for some diseases in our study, is not the case. Another reason for misclassification of the gold standard is caused by fragmentation of health data in the UK, which, in part, is a consequence of fragmented health care.¹⁹ For example, we were not able to link GP data to hospital admissions, and as a consequence, patients diagnosed and treated in hospital were misclassified if the hospital diagnosis was not added to the GP records.

Low and uncertain validity of classification can cause significant problems in medical research. Similar to results from Stanley et al⁸ the difference between AP based on prescription records and TP based on diagnostic records was considerable. This was observed not only between different diseases but also within different strata of the study population. For example, based on diagnostic records the risk ratio of COPD in patients of white ethnic groups compared to other ethnic groups was 5.3, while based on prescription records it was 1.6. Therefore, for studies using prescription records or similar incomplete information for disease classification, there is a considerable risk for bias from differential misclassification.

Conclusions

In summary, researchers should exercise caution when utilising medication records to infer disease status, particularly for conditions with a low prevalence or those where non-pharmacological treatments are available. Although the implications of using medications as a proxy measure will depend on the underlying research question, disease misclassification is likely and may act differentially across time and patient subgroups which could bias both absolute and relative measures of effect.

Abbreviations

APP, apparent period prevalence; BNF, British National Formulary; COPD, chronic obstructive pulmonary disease; EHRs, electronic health records; GORD, gastro-oesophageal reflux disease; HDRUK Health Data Research UK; IMD, index of multiple deprivation; NPV, negative predictive value; OPCR, Optimum Patient Care Research Database; PPV, positive predictive value; QOF, Quality and Outcomes Framework; SE, sensitivity; SP, specificity; TPP, True period prevalence.

Data Sharing Statement

The dataset supporting the conclusions of this article was derived from OPCR. The authors do not have permission to give the public access to the study dataset; researchers may request access to OPCR for their own purposes. The scripts for data management and analysis and the results of the regression models can be shared with readers by contacting the Corresponding Author.

Ethics Approval

The study protocol was established prior to data extraction, in accordance with the criteria for the European Network Centers for Pharmacoepidemiology and Pharmacovigilance (ENCePP). The study was approved by the Anonymised Data Ethics Protocols and Transparency (ADEPT) committee (PROTOCOL2349); the study protocol was registered with the HMA-EMA Catalogues of real-world data sources and studies (EUPAS100000359). The dataset was derived from the OPCR, which has ethical approval from the National Health Service Research Authority to hold and process anonymised research data (REC: 15/EM/0150). All data access complied with relevant data protection and privacy regulations.

Acknowledgments

We would like to acknowledge Mr Steven Cooper (BA) of Optimum Patient Care (OPC) for technical help with the queries.

Funding

Health Data Research UK.

Disclosure

Dr John Busby reports grants from Astrazeneca, outside the submitted work. Professor Jennifer Quint reports grants and/or personal fees from AZ, MRC, Chiesi, and Sanofi, outside the submitted work. Dr David Price reports grants and/or personal fees from AstraZeneca, Boehringer Ingelheim, Chiesi, Cipla, GlaxoSmithKline, Viartis, Teva Pharmaceuticals, Novartis, Regeneron Pharmaceuticals, Sanofi Genzyme, UK National Health Service, and Medscape; peer reviewer for grant committees for Efficacy and Mechanism Evaluation programme and Health Technology Assessment, outside the submitted work; In addition, he owns 74% of the social enterprise Optimum Patient Care Ltd (Australia and UK) and 92.61% of Observational and Pragmatic Research Institute Pte Ltd (Singapore). Professor Liam Heaney reports grants, personal fees, and/or non-financial support from Astrazenea, Glaxosmithkline, and Sanofi, outside the submitted work. The authors report no other conflicts of interest in this work.

References

1. Massen GM, Blamires O, Grainger M, et al. UK electronic healthcare records for research: a scientometric analysis of respiratory, cardiovascular, and COVID-19 publications. *Pragmat Obs Res.* 2024;15:151–164. doi:10.2147/POR.S469973
2. NHS England. Purpose of the GP electronic health record. Available from: <https://www.england.nhs.uk/long-read/purpose-of-the-gp-electronic-health-record/>. Accessed February 20, 2025, 2025.
3. Copenhagen Institute for Further Studies. Secondary use of health data: aggregation to improve policies. Available from: <https://cifs.health/backgrounds/secondary-use-of-health-data/>. Accessed February 20, 2025.
4. Roland M. Linking physicians' pay to the quality of care—a major experiment in the United Kingdom. *N Engl J Med.* 2004;351(14):1448–1454. doi:10.1056/NEJHpr041294
5. Doran T, Kontopantelis E, Valderas JM, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK quality and outcomes framework. *BMJ.* 2011;342(1):d3590. doi:10.1136/bmj.d3590
6. Young JC, Conover MM, Funk MJ. Measurement error and misclassification in electronic medical records: methods to mitigate bias. *Curr Epidemiol Rep.* 2018;5(4):343–356. doi:10.1007/s40471-018-0164-x
7. Benchimol EI, Smeeth L, Guttman A, et al. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLoS Med.* 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885
8. Stanley J, Semper K, Millar E, Sarfati D. Epidemiology of multimorbidity in New Zealand: a cross-sectional study using national-level hospital and pharmaceutical data. *BMJ Open.* 2018;8(5):e021689. doi:10.1136/bmjopen-2018-021689
9. Optimum Patient Care Research Database. Our database. Available from: <https://opcrd.optimumpatientcare.org/our-database>. Accessed April 7, 2025.
10. Lynam A, Curtis C, Stanley B, et al. Data-resource profile: United Kingdom optimum patient care research database. *Pragmat Obs Res.* 2023;14:39–49. doi:10.2147/POR.S395632
11. Whittaker HR, Torkpour A, Quint J. Eligibility of patients with chronic obstructive pulmonary disease for inclusion in randomised control trials investigating triple therapy: a study using routinely collected data. *Respir Res.* 2024;25(1):43. doi:10.1186/s12931-024-02672-x
12. Razieh C, C B. Data from: mapping detailed SNOMED ethnicity codes to harmonised census 2021 ethnic categories, England. 2023. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthinequalities/datasets/mappingdetailedsnomedethnicitycodestoharmonisedcensus2021ethniccategoriesengland>. Accessed January 12, 2026.
13. MHCLG, OCSI, NISRA, Government S, van Dijk J, O'Brien O. Index of multiple deprivation (IMD) 2025. Available from: <https://data.geods.ac.uk/dataset/index-of-multiple-deprivation-imd>. Accessed January 12, 2026.
14. Muschelli J. ROC and AUC with a binary predictor: a potentially misleading metric. *J Classif.* 2020;37(3):696–708. doi:10.1007/s00357-019-09345-1
15. Mbizvo GK, Bennett KH, Schnier C, Simpson CR, Duncan SE, Chin RFM. The accuracy of using administrative healthcare data to identify epilepsy cases: a systematic review of validation studies. *Epilepsia.* 2020;61(7):1319–1335. doi:10.1111/epi.16547
16. Simpson CR, Hannaford PC, Lefevre K, Williams D. Effect of the UK incentive-based contract on the management of patients with stroke in primary care. *Stroke.* 2006;37(9):2354–2360. doi:10.1161/01.STR.0000236067.37267.88
17. Fernández-Llaneza D, Hilbrands LB, Vogt L, Engberink R, Klopotoska JE. Identifying a cohort of hospitalized chronic kidney disease patients using electronic health records: lessons learnt and implications for future research and clinical practice guidelines. *Clin Kidney J.* 2025;18(4):sfaf073. doi:10.1093/ckj/sfaf073
18. Georgie May M, Philip WS, Harley HYK, et al. Review of codelists used to define hypertension in electronic health records and development of a codelist for research. *Open Heart.* 2024;11(1):e002640. doi:10.1136/openhrt-2024-002640
19. Price WN. Risk and resilience in health data infrastructure. *Colo Tech L J.* 2017;16(1):65–85.

Pragmatic and Observational Research

Publish your work in this journal

Pragmatic and Observational Research is an international, peer-reviewed, open access journal that publishes data from studies designed to reflect more closely medical interventions in real-world clinical practice compared with classical randomized controlled trials (RCTs). The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/pragmatic-and-observational-research-journal>

Dovepress
Taylor & Francis Group