BJGP OPEN

Exploring the coding of migration status in English primary care from 2011 to 2025 using OpenCodeCounts

Boukari, Yamina; Hiam, Lucinda; Scuffell, James; Tamborska, Arina; Burns, Rachel; Wiedemann, Milan; Campos-Matos, Ines; Aldridge, Robert W; Walsh, Peter; Hargreaves, Sally; Pathak, Neha; Goldacre, Ben; Hulme, William J

When citing this article please include the DOI provided above.

Author Accepted Manuscript

Title: Exploring the coding of migration status in English primary care from 2011 to 2025 using OpenCodeCounts

Authors:

Yamina Boukari: Bennett Institute Junior Research Fellow in Applied Health Data Science, Bennett Institute of Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford
yamina.boukari@jesus.ox.ac.uk

Lucinda Hiam: Clarendon Scholar, School of Geography and the Environment, University of Oxford.

Jamie Scuffell: NIHR In-Practice Fellow, Department of Population Health Sciences, King's College London

Arina Tamborska: Epidemiologist, Bennett Institute of Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford

Rachel Burns: Senior Research Fellow, UCL Collaborative Centre for Inclusion Health, Research Department Epidemiology and Public Health, University College London

Milan Wiedemann: Senior Researcher, Bennett Institute of Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford

Ines Campos-Matos: Deputy Director, Office for Health Improvement and Disparities, Department of Health and Social Care

Robert W Aldridge: Professor of Health Metrics Sciences, Institute for Health Metrics and Evaluation, University of Washington

Peter Walsh: Senior Researcher, The Migration Observatory, University of Oxford

Sally Hargreaves: Professor of Clinical Public Health, City St George's, University of London

Neha Pathak: NIHR Advanced Fellow & Principal Clinical Research Fellow, Institute for Women's Health, University College London; Honorary Consultant in Sexual and Reproductive Health, Central North West London NHS Foundation Trust

Ben Goldacre: Professor and Director of Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford

William Hulme: Epidemiology Team Lead and Statistical Epidemiologist, Bennett Institute of Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford

Abstract

Background: The migration status of the 9.8 million migrants living in England is not consistently recorded in primary care electronic health records (EHRs). Codelist approaches enable creation of cohorts of individuals who have had a predefined, optional migration-related code (e.g. "refugee") added to their EHR.

Aims: We aimed to explore the use of migration-related SNOMED CT codes to inform future research using primary care data.

Design and Setting: We used our OpenCodeCounts tool to explore data published by NHS England on SNOMED CT code usage in English primary care.

Method: We created migration-related codelists and described their use from 1st August 2011 to 31st July 2025. To understand code usage in the context of known information on migrants in England, we compared code usage to trends in migration-related statistics from the Home Office and the 2021 Census.

Results: There were 34.2 million uses of 1,119 migration-related codes from 2011 to 2025. Migration-related coding increased over time, generally exceeding the increase observed for coding overall, with a sharp increase from 2020, particularly for country-of-birth and language. Language-related coding represented 65% of code usage and where country of birth was recorded, there was mixed agreement with the Census. Coding of immigration legal statuses was low and overwhelmingly about asylum/refugee status.

Conclusion: Utilising OpenCodeCounts, we demonstrate the feasibility of using migration-related SNOMED CT codelists within primary care EHRs and highlight some of the potential biases that cohorts created based on these codelists may have to inform future research.

Keywords: migration, SNOMED CT, England

How this fits in

This study offers population-wide insights into migration-related SNOMED CT coding in primary care in England from 2011 to 2025 using our new open-source tool, OpenCodeCounts. Here, we show that it would be feasible to use SNOMED CT migration-related codelists to create migrant cohorts in English primary care data, but that these cohorts may be biased towards individuals with current or previous English language barriers. We also show that migration-related coding has increased, particularly after the start of the pandemic and for country-of-birth and language codes. The increased use of these code types offers the opportunity for GP practices to better identify patients requiring language support and potential screening and service needs based on their country of birth and paves the way for future needed improvements in migration-related coding.

Introduction

International migrants represent 17.4% or 9.8 million individuals living in England [1]. Migration is a determinant of health, and conditions before, during and after their journey can affect health outcomes [2,3], highlighting the importance of understanding their health status and unmet health needs [4,5]. All individuals in England are legally entitled to access and use primary care free of charge, regardless of their immigration status [6]. Despite this, international migrants (defined here as individuals who are born outside of the UK, an oft-used definition in UK-based migrant health research [7]) face documented barriers to accessing care [8]. For individuals who do manage to access primary care, international migrant status is not systematically or routinely recorded in primary care electronic health records (EHRs).

Having accurate information on migration status in EHRs, with appropriate risk mitigation strategies to prevent misuse, is essential for monitoring inequalities in morbidity, mortality and risk factors for ill health between migrant and non-migrant populations [2]. In the absence of routine, compulsory recording of migration status, several methods can be used to study migrants' health within EHR data, such as data linkage [4,9], natural language processing [10] and codelists approaches [5,11,12]. The codelist approach involves compiling a list of predefined, optional-to-use clinical codes that relate to migration (for example, "refugee" or "interpreter needed") and then searching the EHRs for patients who have at least one of these codes added to their record by primary care staff. Individuals with a relevant migration-related code then comprise a cohort. If migration information is instead written in free text, such as "patient is a refugee", this individual can be missed. The results and interpretation of research based on codelists are dependent on the codes available, coding practices of individuals inputting the codes and policy initiatives that encourage or financially incentivise certain coding practices [13,14].

Whilst NHS England does not publish reference codelists on migration [15], in previous work, a migration codelist using the Read thesaurus of clinical codes [16] was developed and validated with data from individuals registered at GP practices using the Vision EHR software (~4% of the population [17]) between 1997 and 2018 [5]. This study reported improvements in migration-related coding over time and better representation of younger migrants. Cohorts generated using this Read codelist were used to investigate all-cause [11] and sexual and reproductive health-related primary care consultations and contraceptive prescriptions [18]. Another study investigated the incidence of long-term conditions in Latin American migrants using a codelist constructed from the SNOMED CT structured clinical vocabulary [19]) in a London borough [12].

Given the recent availability of population-wide primary care EHRs, such as OpenSAFELY [20], and shifts in clinical coding from Read to the more detailed SNOMED CT terminology from 2018 [19,21], it is important to investigate migration-related SNOMED CT coding at the population level. We aimed to use a new tool developed by our team, OpenCodeCounts [22], to explore the usage of migration-related SNOMED CT codes in English primary care EHRs. We then cross-referenced the findings against general migration trends and characteristics using publicly available immigration datasets to inform the validity of future research on migrant health using codelists applied to primary care EHRs.

Methods

OpenCodeCounts

OpenCodeCounts is a new tool that allows users to explore primary and secondary care coding trends interactively (using the web tool) or programmatically (using the R package) [22]. It is underpinned by publicly available data published by NHS England on annual SNOMED CT code usage in primary care [23] and hospital activity [24] from 1st August 2011 to 31st July 2025. This particular analysis focused on SNOMED CT coding with the underlying datasets representing the number of times a SNOMED CT concept was added to a GP patient record in England during the specific yearly period, with code usage rounded to the nearest 10. Calculating distinct patient counts for a given code or set of codes is not possible as a patient may have a code added to their record multiple times during the reporting period. SNOMED CT coding data was submitted from 2019. Prior to 2019, data was submitted in other formats (Read v2 or CTV3) but has been mapped to SNOMED CT.

Codelists

We created lists of migration-related SNOMED CT codes using OpenCodelists.org, an open-access website for codelist curation and sharing. To do this we implemented a previously reported search strategy to find migration-related codes [5], which were reviewed to include or exclude codes based on their relevance to the category definition (Table 1). The codelists were reviewed with two general practitioners (JS and LH). All codelists are published and openly available for reuse via OpenCodelists.org. Final codelists (comprising codes used at least once) for this analysis are included in the supplementary appendix (Supplementary Tables 1-6).

Table 1: SNOMED CT codelist definitions

Codelist*	Definition	Search terms used in OpenCodelists.org	Number of codes included†	OpenCodelist.org reference
All migration-related codes	Any code indicating that an individual could be an international migrant (e.g. related to visa or legal status, country of birth, a main language that is not English, requiring an interpreter or unable to speak/understan	abroad, asylum, born in, countr, english, exploit, forced, humanitarian, illegal, interpreter, language, leave to remain, migrant, migrat, refugee, servitude, slav,	2,820	[25]

	d English or being a victim of trafficking)	traffick, victim, visa, overseas		
Country-of-birth codes	Any code indicating a country of birth that is not the UK or its devolved nations	born in	719	[26]
Immigration legal status codes	Any code indicative of a specific legal status (e.g indefinite leave to remain, asylum seeker awaiting decision)	asylum, citizenship, immigra, indefinite leave, refugee, spouse visa, visa, work permit	40	[27]
Asylum or refugee status codes	Any code related to refugee or asylum status, including codes related to presence in an immigration removal centre	asylum, refugee, illegal, initial health assessment	30	[28]
Language-related codes	Any code indicating a first language that is not English	language, interpreter	1,984	[29]
Interpreter need codes	Any code indicating that an individual requires an interpreter for a non-English language, (codes specifically relating to sign language were excluded)	language, interpreter	505	[30]

*Includes inactive codes and codes with 0 usage during the study period.

†Examples of individual SNOMED CT codes can be found in Table 3.

Comparator data source: Migration statistics

In order to contextualise the trends in migration-related SNOMED CT coding with trends and characteristics of known migrants in England we compared migration-related code usage to

trends in annual migration data published by the Office for National Statistics (ONS) and the Home Office (HO; Table 2), using the latest data released at the time of publishing. We also compared the usage of country-of-birth codes to data on non-UK countries of birth from the Census 2021 [31].

Table 2: Migration datasets used for comparison against migration-related SNOMED CT coding

Category	Definition	Source
Long-term international migration	Defined according to the UN definition of an international migrant as "a person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of destination effectively becomes his or her new country of usual residence". Data represent long-term immigration to the UK for all nationalities excluding British nationality and represent annual figures ending in June of each respective year.	ONS: Long-term international migration: Year Ending June 2012 to Year Ending June 2024, Table 1 [32]
Individuals seeking asylum, on humanitarian schemes, and refugees (referred to herein as "Asylum and humanitarian status")	All individuals applying for asylum at port or in-country and individuals admitted out-of-country to a refugee resettlement scheme (including British Nationals Overseas [BNO] visas for Hong Kong Nationals and Ukraine Visa Schemes).	Asylum claims: HO table Data_Asy_D01 Refugee resettlement (excluding BNO and Ukraine schemes): HO table Data - Res_D02 BNO and Ukraine Visa Schemes: HO table Hum_01 [33]
Individuals granted work visas (referred to herein as "Work visa status")	All individuals issued a work-related visa out-of-country (Visa type group = "Work"; Case outcome = "Issued"). Includes main applications and dependants.	Entry clearance visa applications and outcomes detailed datasets, year ending December 2024, Data_Vis_D02 [33]
Individuals granted study-related visas (referred to as "Study visa status")	All individuals issued a study visa out-of-country (Visa type group = "Study"; Case outcome = "Issued"). Includes main applications and dependants.	HO: Entry clearance visa applications and outcomes detailed datasets, year ending December 2024, Data_Vis_D02 [33]

Individuals granted Family-related visas (referred to as "Family visa status")	All individuals issued a family visa out-of-country (Visa type group = "Work"; Case outcome = "Issued"). Includes main applications and dependants.	HO: Entry clearance visa applications and outcomes detailed datasets, year ending December 2024, Data_Vis_D02 [33]

HO, Home Office; ONS, Office for National Statistics

Analysis

The total numbers of recorded events for all codes that were used at least once during the study period in each codelist were plotted annually and the five most frequently used codes were tabulated. Migration-related coding as a percentage of overall SNOMED CT coding was calculated for each analysis year. The percentage increases in migration-related code usage and overall code usage during the study period were calculated as firstly, the difference in the total code usage (either migration or overall) in 2011/12 and total usage in 2024/25 over the total usage at the beginning of the period (2011/12) or secondly as annual percentage increases. Immigration data for each category of migrant were plotted and descriptively compared to the code usage data. For each of the 10 most common non-UK countries of birth according to the Census 2021, the respective percentage of code use was calculated (Census data: numerator represents the number of individuals with country of birth, denominator represents the total number of individuals with a non-UK country of birth. SNOMED CT data: numerator represents the cumulative number of the respective country's SNOMED CT codes from 2011/12 to 2020/21; denominator represents the total recorded instances of any country-of-birth codes from 2011/12 to 2020/21). As this was a descriptive analysis of data representing near population coverage, statistical significance testing was not required.

Tools

Data analysis was carried out using R version 4.4.2 and the *opencodecounts* R package [34]. Analysis code was written by YB, reviewed by MW and is available on GitHub [34].

Results

SNOMED CT coding trends

From the 1st August 2011 to 31st July 2025, there were 34.2 million (34,224,670) instances of 1,119 different migration-related codes recorded in primary care EHRs in England. Migration-related coding increased over time, with a dip at the beginning of the COVID-19 pandemic, followed by recovery and a sharp increase from 2020, particularly for country-of-birth and language-related codes, that continued until July 2025 (Figure 1). A similar increasing trend was observed for overall SNOMED CT coding (grey bar chart, Figure 1). Migration-related coding as a percentage of overall SNOMED CT coding has increased from 0.06 to 0.08% over the study period (Supplementary Figures 1-3).

Language codes represented the majority of migration-related codes (814/1,119 codes and 22,366,320/34,224,670 [65%] of recorded instances; Supplementary Table 7), with "interpreter needed" (7% of all migration-related code instances), "main spoken language Polish" (4%), "main spoken language Romanian" (4%), "main spoken language Urdu" (4%) and "interpreter present" (4%) being the most commonly used migration-related codes (Table 3). There were 182 codes indicating interpreter needs with 4,998,335 recorded instances (Supplementary Table 7).

Country-of-birth codes were the next most common code type (representing 256 unique codes and 9,821,900 recorded instances; Supplementary Table 7). Amongst all uses of country-of-birth codes, the five most common codes used were being born in India, Pakistan, Romania, Nigeria and China. Of the 702,750 codes relating to a specific immigration legal status, the majority (670,460/702,750; 95%; Supplementary Table 7) indicated that an individual was an asylum seeker or refugee; only 2% indicated that an individual had a student visa.

Figure 1: Number of migration-related SNOMED CT codes (lines) and all SNOMED CT codes (bars) recorded in primary care from 2011/12 to 2024/25
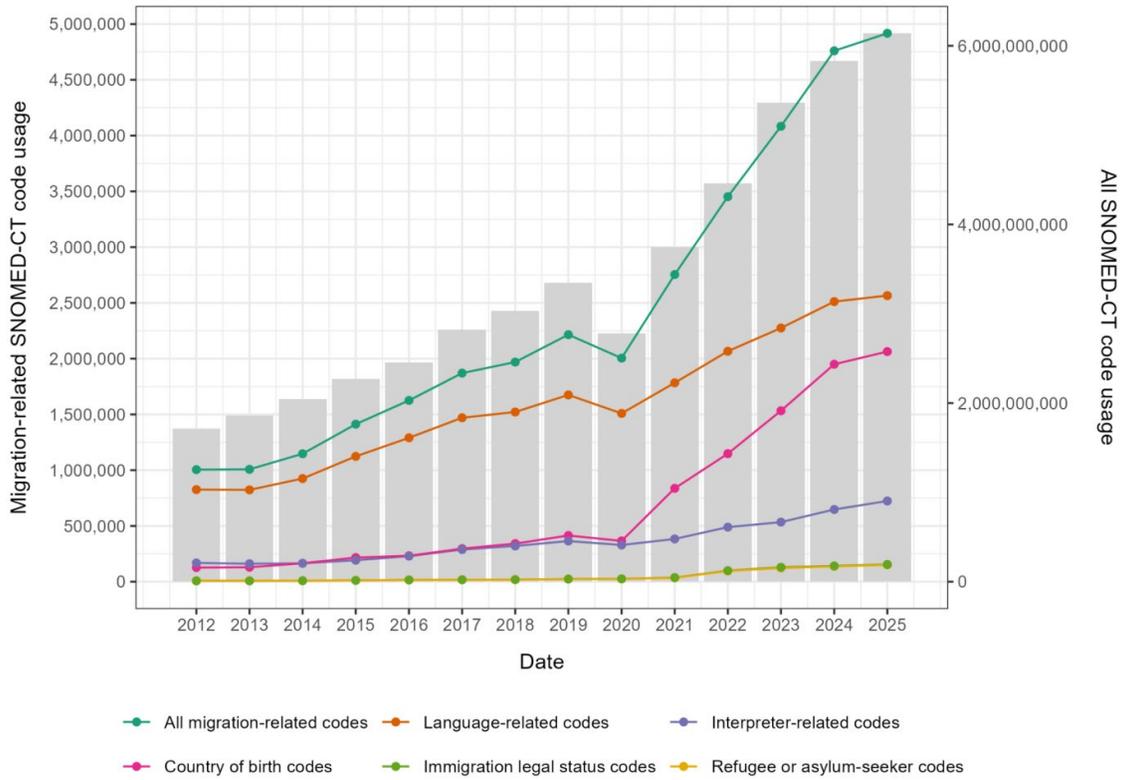
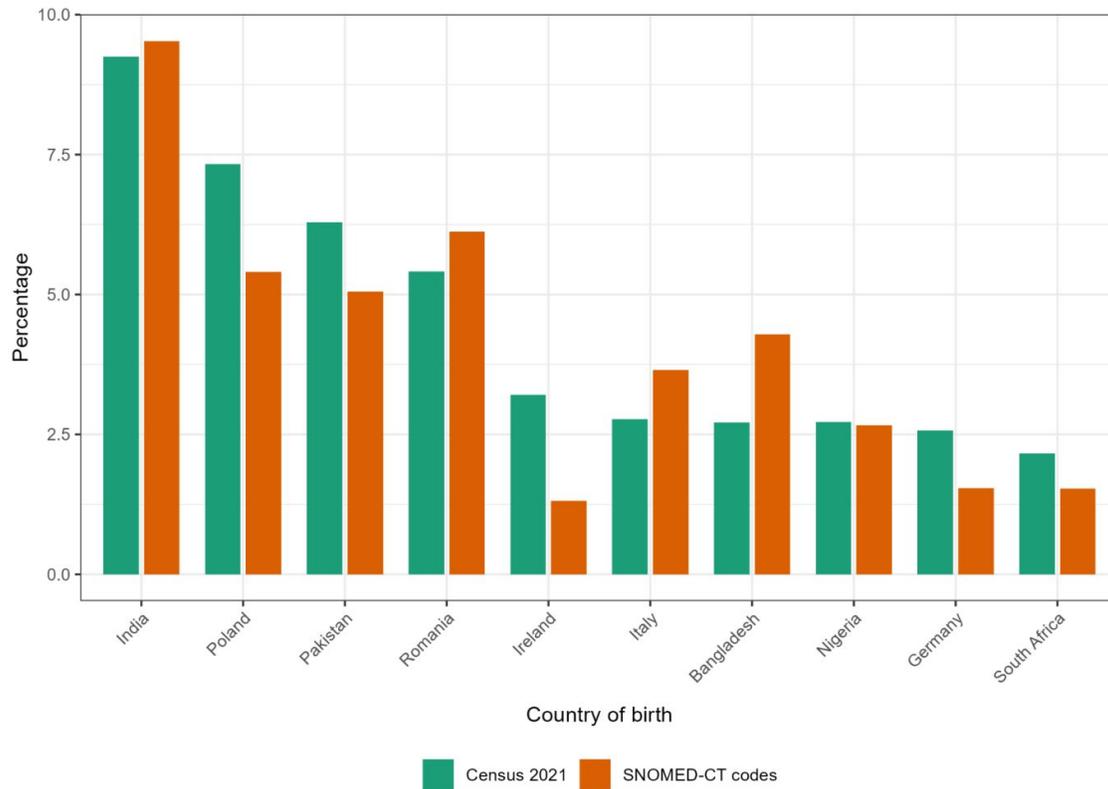**Table 3: Five most commonly recorded migration-related SNOMED CT codes**

SNOMED CT code	Description	Number of times used	Percentage of times used within the codelist
All migration-related codes (1,119 codes with 34,224,670 recorded instances)			
315594003	Interpreter needed (finding)	2,292,860	7
315579002	Main spoken language Polish (finding)	1,455,920	4
698678003	Main spoken language Romanian (finding)	1,394,510	4
315588006	Main spoken language Urdu (finding)	1,331,550	4
314431000	Interpreter present (finding)	1,242,420	4

Comparison of migration-related SNOMED CT coding and migration data

Long-term immigration to the UK remained relatively stable from the year ending June 2012 (565,000 individuals) to June 2021 (675,000), with a sharp increase to 1,173,000 in the year ending June 2024 (Figure 2). The number of individuals immigrating annually was numerically lower than the annual numbers of migration-related codes recorded in GP records. Of the different immigration legal status types, study (278,100 visas issued in 2012 increasing to 419,312 in 2024) and work (145,110 to 369,419) visas (for main applicants and their dependants) were the most common in the immigration data, compared to asylum- or refugee-related statuses in the coding data.

The percentage of "Born in India" and "Born in Nigeria" codes (as a percentage of all country-of-birth codes) used in GP data were generally consistent with the percentages of individuals born in these countries according to the 2021 Census (Figure 3). Romania, Italy and Bangladesh were overrepresented in the coding data in comparison to the Census data. Poland, Pakistan, Ireland, Germany and South Africa were underrepresented.

Figure 2: Immigration per calendar year by legal status types from 2012 to 2024.

Figure 2 notes: Long-term immigration [32] to the UK excludes individuals with British nationality. Study, work and family visas [33] represent grants of the respective visa types for main applicants and dependants. Asylum and humanitarian status represents asylum claims, decisions (grants, refusals and withdrawn applications), grants of out-of-country resettled refugee schemes, British Nationals Overseas [BNO] visas for Hong Kong Nationals and Ukraine Visa Schemes [33].

Figure 3: Top 10 non-UK countries of birth as a percentage of the total non-UK born population in England from the Census 2021 compared to the use of the respective country-of-birth SNOMED CT codes as a percentage of all country-of-birth SNOMED CT codes from 2011/12 to 2020/21

Census 2021   SNOMED-CT codes

Discussion

Summary

We found over 34 million records of migration-related codes used in primary care EHRs from 2011 to 2025 using the OpenCodeCounts tool. Migration-related coding has increased over time, particularly from 2020 and in the case of country-of-birth and language-related codes. Overall, language-related codes were the most commonly recorded type. Annual instances of migration-related codes are greater than the annual immigration flows to the UK, which likely reflects that multiple migration-related codes may be used in an individual's EHR. Where countries of birth were recorded, there was mixed alignment with non-UK countries of birth reported in the census. Specific immigration legal statuses were less commonly recorded and when used, predominantly related to asylum or refugee status, which did not align with immigration statistics in which work and study statuses were more common.

Strengths and limitations

The strength of this analysis is that it utilises data representing all uses of SNOMED CT codes across the majority of GP practices in England (n=6,624 practices covering 62,804,225 registered patients at the latest release of the data [23]) and uses external data sources on known migrants, HO immigration statistics and Census data, to qualitatively compare trends. Additionally, it is the first output from the OpenCodeCounts tool, which was

designed to facilitate research in primary care data resources such as OpenSAFELY, and uses a selection of curated migration-related SNOMED CT codelists that are openly accessible for review and reuse at OpenCodelists.org.

The main limitation of this study is that data represent individual instances of code use, not individual patients (i.e. patients may have more than one migration-related code). To understand how many individuals have a migration-related code recorded, analysis of patient-level data is required, for example via OpenSAFELY, plans for which are underway. Additionally, we were unable to determine the drivers of migration coding for primary care staff, which is important in order to understand the potential biases in cohorts generated based on this coding. We were also unable to capture unstructured, free-text migration-related information, which may be where some staff chose to record migration status. Although natural language processing methods have begun to be used to explore free-text recording of migration status in secondary care records, there is currently, to our knowledge, no means to do so using large-scale primary care data resources [10]. Finally, this analysis represents instances of migration coding in the EHRs of individuals who have been able to register with a GP and therefore excludes individuals who may not be able to register or attend services [35,36].

Comparison with existing literature

The gradual increase in migration coding from 2011 to 2019 is consistent with findings from the analysis of the previously reported migration Read codelist [5]. At the start of the pandemic, there was a decrease in coding, consistent with reported dips in coded primary care activity at this time [37], followed by sharp increases in migration-related coding over that seen for overall coding. This sharp increase was also seen in the annual immigration data for all groups apart from family visas. Potential reasons for the increase in coding could be the push to ensure GP registration of groups such as migrants in vulnerable situations at the beginning of the pandemic, the use of EHR templates that included country of birth and main language spoken [38], a heightened awareness of migration due to the use of risk assessments related to recent travel, and the increase in all types of migration from 2021, as shown in the immigration statistics. Additionally, it may be that specific refugee resettlement schemes were provided with an initial healthcare assessment where migration-related codes were captured [39,40].

Language codes were the most commonly used migration-related code, consistent with findings showing that in the previously reported Read migration codelist, 57% of migration-related Read codes concerned language [5]. Whilst there are no formal incentives to record language, this finding could reflect the practical benefit to practice staff in recording information related to language needs. Additionally, NHS England's guidance for commissioners and the Office for Health Improvement and Disparities' Migrant Health Guide actively encourage primary care staff to record a patient's preferred spoken and written languages and interpreter requirements [41,42].

Compared to the Census, Romania, Italy and Bangladesh were overrepresented as countries of birth in the coding data, whereas Poland, Pakistan, Ireland, Germany and South Africa were underrepresented. The reasons for these discrepancies could reflect different

health-seeking behaviours, English language skills, or coding practices of healthcare professionals. In the case of Polish and Pakistani migrants, it is possible that Polish and Urdu language codes, which were more frequently recorded than country of birth codes, were used instead.

Only 2% of all migration codes related to immigration legal statuses, which were dominated by asylum- and refugee-related statuses, contrary to immigration statistics showing that asylum and refugee statuses are one of the smaller groups of international migrants in England (29,031 asylum applications and resettled refugee status grants in 2012 and 154,690 in 2024). Potential explanations are payments for asylum seeker services being tied to recording the 'Asylum seeker' SNOMED CT code [40,43]. Secondly, amidst policy pushes towards inclusion health and under-served groups [44], GPs may code asylum and refugee statuses more as they consider them more relevant to the patient's health compared to other visa statuses. Thirdly, it could reflect who primary care staff think of as migrants considering that the general public believe that 62% of all migrants come to the UK to seek asylum versus come to study, work, or join families [45].

Implications for research and practice

The presence of over 34 million migration-related codes over the study period demonstrates that it would be feasible to generate cohorts of a substantial size using EHR data. The next steps would be to apply the codelist to primary care data and validate the resultant cohorts by comparing their characteristics to that of migrants in other publicly available datasets such as the census of country of birth estimates, both published by the ONS. Plans for this are currently underway within the OpenSAFELY platform, dependent on obtaining the relevant approvals. The findings also show that cohorts created using the migration codelist may be biased towards individuals who have a current or previous English language barrier, which should be highlighted in any future discussion of outcomes derived from these cohorts. It would also be important to include a description of the migrant cohort and the code types that have caused individuals to be included in the cohort (e.g. language, country of birth etc.) in future research, particularly given that coding practices may change over time. Further qualitative and quantitative explorations of coding are needed to identify potential biases that coding practices could introduce into epidemiological results, including the timing of migration coding and the concurrent use of different migration-related codes. In studies focusing on individuals who have interpreter needs, further consideration should be given to individuals' changing language skills over time and how this may or may not be represented in coding practices. More broadly, the findings demonstrate the usefulness of OpenCodeCounts for conducting rapid feasibility checks to support the use of new codelists in primary care EHR data and show that OpenCodeCounts can also be used to generate insights on potential biases, which can then be examined in patient-level data in future.

Through showing that migration-related coding has increased since 2020, particularly in the case of language and country of birth, we can see that there is potential for making improvements to the coding of migration status in English primary care EHRs, which would benefit both research and direct patient care. From a direct care perspective, language coding can be used to prospectively identify individuals who would benefit from double appointments to accommodate interpreter usage, although further consideration around how

to reflect an individual's improved language skills over time is needed. Information on country of birth can help in determining screening and vaccination requirements, and chronic disease follow-up as set out by UK migrant health guidelines [46]. Furthermore, Primary Care Networks are required to use data-driven approaches to improve population health outcomes in alignment with the CORE20PLUS5 approach to address health inequalities [47,48]. Improving migration-related coding is important for meeting these requirements and aligns with recommendations for proactive coding of other characteristics such as homelessness within primary care [49].

Further qualitative work is needed to understand the drivers behind coding practices, reasons behind the improvements in migration-related coding, and how to promote improved coding practices across primary care providers. Particular attention should be given to the potential role that administrative staff can play in collecting this data given that individuals may be asked questions regarding their country of birth and length of time in the UK at registration, which may not always be coded in the EHR. The use of clinical decision support systems that prompt the coding of country of birth to improve disease detection and screening guidance for migrants should be further explored [46], as should financial incentives as a way to encourage the routine recording of migration-related information. It is also crucial to explore risk mitigation strategies for preventing misuse of migration status data given the rise in anti-migrant rhetoric, historical data sharing between NHS Digital and the Home Office for immigration enforcement purposes [50], and the impact that coding could have on secondary care charging.

To conclude, utilising OpenCodeCounts we demonstrate the feasibility of using migration-related SNOMED CT codelists to generate migrant cohorts within primary care EHRs and highlight some of the potential biases that cohorts created based on these codelists may have.

Competing interests

Funding

Ethics statement

Ethical approval was not required for this study.

References

1   Country of birth (extended) and ethnic group - Office for National Statistics.
     https://www.ons.gov.uk/datasets/create/filter-outputs/eaea29b3-7f24-4799-a324-
     2790a6c7d9e1 (accessed 6 February 2025)

2   Abubakar I, Aldridge RW, Devakumar D, *et al.* The UCL-Lancet Commission on
     Migration and Health: the health of a world on the move. *Lancet*. 2018;392:2606–54.

3   World report on the health of refugees and migrants. 2017.
     https://www.who.int/publications/i/item/9789240054462 (accessed 18 March 2025)

4   Burns R, Pathak N, Campos-Matos I, *et al.* Million Migrants study of healthcare and
     mortality outcomes in non-EU migrants and refugees to England: Analysis protocol for a
     linked population-based cohort study of 1.5 million migrants. *Wellcome Open Res*.
     2019;4:4.

5   Pathak N, Zhang CX, Boukari Y, *et al.* Development and validation of a primary care
     electronic health record phenotype to study migration and health in the UK. *Int J Environ
     Res Public Health*. 2021;18:13304.

6   Department of Health and Social Care. NHS entitlements: migrant health guide.
     https://www.gov.uk/guidance/nhs-entitlements-migrant-health-guide (accessed 29
     September 2025)

7   Burns R, Zhang CX, Patel P, *et al.* Migration health research in the United Kingdom: A
     scoping review. *J Migr Health*. 2021;4:100061.

8   Asif Z, Kienzler H. Structural barriers to refugee, asylum seeker and undocumented
     migrant healthcare access. Perceptions of Doctors of the World caseworkers in the UK.
     SSM Ment Health. 2022;2:100088.

9   Aldridge RW, Zenner D, White PJ, *et al.* Tuberculosis in migrants moving from high-
     incidence to low-incidence countries: a population-based cohort study of 519 955
     migrants screened before entry to England, Wales, and Northern Ireland. *Lancet*.
     2016;388:2510–8.

10   Crowley G, Roberts A, Stewart R, *et al.* Investigating mental healthcare inequalities
     associated with forced migration: promise and potential pitfalls of electronic health
     records. *Br J Psychiatry*. 2024;225:305–7.

11   Zhang CX, Boukari Y, Pathak N, *et al.* Migrants' primary care utilisation before and
     during the COVID-19 pandemic in England: An interrupted time series analysis. *Lancet
     Reg Health Eur*. 2022;20:100455.

12   Scuffell J, Bailey J, Dodhia H, *et al.* Incidence of long-term conditions in the Latin
     American community of London: A validation and retrospective cohort study of 890,922
     primary care records, 2005-2022. *PLoS One*. 2024;19:e0312311.

13   Donegan K, Fox N, Black N, *et al.* Trends in diagnosis and treatment for people with
     dementia in the UK from 2005 to 2015: a longitudinal retrospective cohort study. *Lancet
     Public Health*. 2017;2:e149–56.

14   Beaney T, Clarke J, Salman D, *et al.* Identifying potential biases in code sequences in
     primary care electronic healthcare records: a retrospective cohort study of the
     determinants of code frequency. *BMJ Open*. 2023;13:e072884.

15   England NHS. Primary Care Domain - reference sets (PCD refsets).
     https://app.powerbi.com/view?r=eyJrIjoiNDRmYjEwMzQtZGE3MS00ZGE5LTgwMTUtNj
     Q2NGE1NTZiYmEzIiwidCI6IjM3YzM1NGIyLTg1YjAtNDdmNS1iMjIyLTA3YjQ4ZDc3NG
     VlMyJ9 (accessed 29 September 2025)

16   Read Codes. NHS England Digital. https://digital.nhs.uk/services/terminology-and-
     classifications/read-codes (accessed 4 April 2025)

17   CPRD GOLD February 2024 dataset. CPRD. https://www.cprd.com/doi/cprd-gold-
     february-2024-dataset (accessed 15 May 2025)

18   Pathak N, Zhang CX, Boukari Y, *et al.* Sexual and reproductive health and rights of
     migrant women attending primary care in England: A population-based cohort study of
     1.2 million individuals of reproductive age (2009-2018). *J Migr Health*. 2024;9:100214.

19   SNOMED CT. NHS England Digital. https://digital.nhs.uk/services/terminology-and-
     classifications/snomed-ct (accessed 5 February 2025)

20   About OpenSAFELY. https://www.opensafely.org/about/ (accessed 15 May 2025)

21   England NHS. NHS England » Clinical coding – SNOMED CT.
     https://www.england.nhs.uk/long-read/clinical-coding-snomed-ct/ (accessed 15 May
     2025)

22   Tamborska AA, Higgins R, Boukari Y, *et al.* OpenCodeCounts : An open-access,
     interactive online tool and R package for analysing clinical code usage in England.
     medRxiv. 2025;2025.10.14.25338005.

23   [MI] SNOMED Code Usage in Primary Care. NHS England Digital.
     https://digital.nhs.uk/data-and-information/publications/statistical/mi-snomed-code-
     usage-in-primary-care (accessed 19 February 2025)

24   Hospital Admitted Patient Care Activity. NHS England Digital. https://digital.nhs.uk/data-
     and-information/publications/statistical/hospital-admitted-patient-care-activity (accessed
     29 September 2025)

25   OpenCodelists: Migrant.
     https://www.opencodelists.org/codelist/user/YaminaB/migration-status/5ba3fcb8/
     (accessed 19 November 2025)

26   OpenCodelists: Born outside the UK.
     https://www.opencodelists.org/codelist/user/YaminaB/born-outside-the-uk/317726d5/
     (accessed 19 November 2025)

27   OpenCodelists: Immigration legal status.
     https://www.opencodelists.org/codelist/user/YaminaB/uk-visa/4eb363bd/ (accessed 19
     November 2025)

28   OpenCodelists: Asylum seeker or refugee.
     https://www.opencodelists.org/codelist/user/YaminaB/asylum-seeker-or-
     refugee/35a3f088/ (accessed 19 November 2025)

29   OpenCodelists: English not main language.
     https://www.opencodelists.org/codelist/user/YaminaB/english-not-main-
     language/7f58406a/ (accessed 19 November 2025)

30   OpenCodelists: Interpreter required.

https://www.opencodelists.org/codelist/user/YaminaB/interpreter-required/2c11a6e9/ (accessed 19 November 2025)

31 Office for National Statistics. International migration, England and Wales - Office for National Statistics. 2022. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internat ionalmigration/bulletins/internationalmigrationenglandandwales/census2021 (accessed 27 January 2025)

32 Long-term international immigration, emigration and net migration flows, provisional. 2024. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internat ionalmigration/datasets/longterminternationalimmigrationemigrationandnetmigrationflow sprovisional (accessed 25 February 2025)

33 Home Office. Immigration system statistics data tables. 2023. https://www.gov.uk/government/statistical-data-sets/immigration-system-statistics-data-tables (accessed 20 March 2025)

34 *opencodecounts: Clinical Code Usage Counts in England*. Github .

35 Worthing K, Seta P, Ouwehand I, *et al.* Reluctance of general practice staff to register patients without documentation: a qualitative study in North East London. *Br J Gen Pract*. 2023;73:e276–83.

36 Ciftci Y, Blane DN. Improving GP registration and access for migrant health. *Br J Gen Pract*. 2022;72:56–7.

37 Curtis HJ, MacKenna B, Wiedemann M, *et al.* OpenSAFELY NHS Service Restoration Observatory 2: changes in primary care clinical activity in England during the COVID-19 pandemic. *Br J Gen Pract*. 2023;73:e318–31.

38 New Patient Resources. Ardens EMIS Web. https://support-ew.ardens.org.uk/support/solutions/articles/31000159449-new-patient-resources (accessed 3 April 2025)

39 Knights F, Munir S, Ahmed H, *et al.* Initial health assessments for newly arrived migrants, refugees, and asylum seekers. *BMJ*. 2022;377:e068821.

40 NHS England. Afghan Resettlement – Enhanced Health Assessments Service Specifications. https://www.england.nhs.uk/wp-content/uploads/2021/09/20210827-Afghanistan-Relocations-and-Assistance-Policy-annex-.pdf (accessed 4 April 2025)

41 Language interpreting and translation: migrant health guide. GOV.UK. https://www.gov.uk/guidance/language-interpretation-migrant-health-guide (accessed 3 April 2025)

42 Guidance for commissioners: Interpreting and Translation Services in Primary Care. https://www.england.nhs.uk/wp-content/uploads/2018/09/guidance-for-commissioners-interpreting-and-translation-services-in-primary-care.pdf (accessed 4 April 2025)

43 North Central London Integrated Care Board. Healthcare Provision for People Seeking Asylum Residing in Initial Accommodation Centre sites in NCL. https://gps.northcentrallondon.icb.nhs.uk/enhanced-service/healthcare-provision-for-people-seeking-asylum-residing-in-initial-accommodation-centre-iac-sites-in-ncl (accessed 3 April 2025)

44   Inclusion Health: applying All Our Health. GOV.UK.
       https://www.gov.uk/government/publications/inclusion-health-applying-all-our-
       health/inclusion-health-applying-all-our-health (accessed 1 May 2025)

45   Thinking Behind the Numbers: Understanding Public Opinion on Immigration in Britain.
       Migration Observatory. 2016.
       https://migrationobservatory.ox.ac.uk/resources/reports/thinking-behind-the-numbers-
       understanding-public-opinion-on-immigration-in-britain/ (accessed 3 April 2025)

46   Carter J, Goldsmith LP, Knights F, *et al.* Health Catch-UP!: a realist evaluation of an
       innovative multi-disease screening and vaccination tool in UK primary care for at-risk
       migrant patients. *BMC Med*. 2024;22:497.

47   NHS England. NHS England » Core20PLUS5 (adults) – an approach to reducing
       healthcare inequalities. https://www.england.nhs.uk/about/equality/equality-
       hub/national-healthcare-inequalities-improvement-programme/core20plus5/ (accessed
       30 September 2025)

48   NHS England. Network contract directed enhanced service. 2024.
       https://www.england.nhs.uk/wp-content/uploads/2024/03/PRN01035_iii_Network-
       contract-DES_Part-A-Clinical-and-support-services-Section-8.pdf (accessed 30
       September 2025)

49   Transforming primary care for homeless and inclusion health. Transformation Partners
       in Health and Care. 2024.
       https://www.transformationpartners.nhs.uk/programmes/homeless-health/transforming-
       primary-care-for-homeless-and-inclusion-health/ (accessed 30 September 2025)

50   Papageorgiou V, Wharton-Smith A, Campos-Matos I, *et al.* Patient data-sharing for
       immigration enforcement: a qualitative study of healthcare providers in England. *BMJ
       Open*. 2020;10:e033202.