Received: 20 March 2024 | Accepted: 15 May 2024

DOI: 10.1111/jep.14045

ORIGINAL PAPER

Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research
WILEY

"Application of five different strategies to define a cohort of patients with knee osteoarthritis in a large primary care database"

Aqila Taqi PhD[1,2] | Sonia Gran PhD[3] | Roger David Knaggs PhD[1,4,5]

[1]Division of Pharmacy Practice and Policy, School of Pharmacy, University Park Campus, University of Nottingham, Nottingham, UK

[2]Department of Pharmacy, Sultan Qaboos University, Sultan Qaboos University Hospital, Muscat, Oman

[3]Division of Rheumatology, Orthopaedics and Dermatology, Centre of Evidence Based Dermatology, School of Medicine, University of Nottingham, Nottingham, UK

[4]Pain Centre versus Arthritis, University of Nottingham, Nottingham, UK

[5]Primary Integrated Pain Services, Nottingham, UK

Correspondence
Aqila Taqi, PhD, Sultan Qaboos University, Sultan Qaboos University Hospital, Musat, Oman.
Email: Aqeela@squ.edu.om

Abstract

Background: Electronic health records (EHR) are frequently used for epidemiological research including drug utilisation studies in a defined population such as the population with knee osteoarthritis (KOA). We sought to describe the process of defining a cohort of patients with KOA from a large UK primary care database and estimate the annual incidence of diagnosed KOA between 2000 and 2015.

Method: This was a retrospective study using data from the clinical practice research datalink (CPRD). CPRD is a large primary care longitudinal electronic medical records' database that contains anonymous records of patients from general practices across United Kingdom. Five different cohort definition strategies were applied including symptoms-based or diagnosis-based strategies or a combination of both. To validate results, the annual incidence of KOA was estimated and compared to published data.

Results: The study defined 898,690 patients when symptoms-based strategy was applied, 137,541 patients when diagnosis based and 83,294 when a combination of both strategies were applied. The final cohort was defined using a diagnosis-based strategy that avoided overestimation (with symptoms-based definition) or underestimation (with a combination of symptoms and diagnosis). The incidence of KOA ranged from 1.33 per 1000 CPRD registrants in 2000, 1.76 in 2008 and 1.45 patients in 2015.

Conclusion: This study logically/sensibly defined a cohort of patients with diagnosed KOA through the application of several strategies. This was an essential step to avoid subsequent over or underestimation of the prevalence of drug utilisation and the associated adverse clinical outcomes within primary care patients with KAO.

KEYWORDS
cohort definition strategy, knee osteoarthritis incidence, osteoarthritis, read codes

1 | INTRODUCTION

Electronic healthcare records are frequently used for epidemiologic studies of osteoarthritis (OA) generally and knee osteoarthritis (KOA) specifically. Such studies are conducted for several purposes, including the estimation of incidence and risk factors for KOA,[1] the association between socioeconomic status and the risk of KOA,[2] the study of population trends in the incidence of OA (including KOA) and initial pharmacological management[3] and the study of drug utilisation in patients with OA[4] or KOA.[5] Cases of OA and KOA were defined through the application of International Classification of Diseases codes in studies conducted in the United States,[6] and Europe,[1,2,4] or by application of Read codes in UK studies using primary care data.[3]

The Read code system allows clinicians to label a presenting complaint using symptom-based or diagnosis-based Read codes. Accordingly, OA may be recorded in electronic health records (EHRs) as peripheral OA-relevant joint symptoms, i.e. peripheral joint pain, arthralgia of the joint (e.g., knee pain, arthralgia of the knee) or as a disease diagnosis, i.e. joint-specific OA diagnoses (e.g., KOA).[7] Unlike conditions such as rheumatoid arthritis, where disease-modifying antirheumatic drugs may aid cohort definition, there aren't any pharmacological treatments specific to OA, hence defining the population of patients with OA in EHR records relies on the diagnostic or symptom coding.[8] Epidemiologic studies have either adopted symptom-based definitions (also known as clinical OA definition) or diagnosed OA definitions. However, the estimated prevalence and incidence rates of OA or KOA are strongly influenced by the applied definition as summarised in Table 1.

For example, a study in the United Kingdom used clinical practice research datalink (CPRD) data to estimate the incidence of KOA using diagnosis-based and symptom-based case definitions. The incidence rates were reported as 47.7 per 1000 person-years (95% confidence interval [CI]: 47.4, 47.9) with the clinical OA definition, compared to 7.9 (95% CI: 7.8, 8.0) with diagnosed OA in 2013.[3]

Additionally, with the application of OA relevant codes in EHRs, cases were defined with restrictive or less-restrictive algorithms based on the number of codes (single or >1 within a given time span) and type (outpatient or hospitalisation records).[9–11]

In epidemiologic studies focussing on drug utilisation in patients with OA/KOA using EHR data, cases were defined by researchers across health systems, mostly using OA- or KOA-specific diagnostic codes rather than symptom-based codes, however, a number of studies have applied both (symptom-based and diagnosis-based definitions) to provide separate prevalence and incidence estimates (Table 2).

Guided by the definitions applied in previous research in patients with OA or KOA using EHRs, the present study aimed to describe the process of selection of an appropriate cohort of patients with KOA on which further analyses of drug utilisation were carried out.[12,13] Additionally, to validate the cohort selection method and resulting numbers, the study aimed to estimate the annual incidence of KOA diagnosis among primary care patients using CPRD data.

TABLE 1 Variation in incidence estimates with the applied case definitions.

Type of KOA case definition	Description of case definition	Resulting estimates	Application/usefulness	Advantage
Clinical case definition	Peripheral joint symptom codes in the records	Potential overestimation of the true incidence	To estimate the burden of the condition & for service planning	High sensitivity
Diagnosed case definition	KOA diagnostic codes in the records	Potential underestimation of the true incidence	When high specificity is required, no false positive cases to be included	High specificity

Abbreviation: KOA, knee osteoarthritis.

TABLE 2 Examples of case definitions applied in drug utilisation studies.

Author, year	Study aim	OA/KOA case definition codes	Records
Gore, 2011	To examine comorbidities, pain-related pharmacotherapy and direct medical costs of patients with OA (including KOA)	Diagnosis of OA using ICD-9-CM codes (715.XX) corresponding to osteoarthritis and allied disorders	Single record
Wilson, 2014	To examine the prevalence of drug use in patients with OA (including KOA) in Spain.	ICD diagnostic codes of OA including M17 corresponding to osteoarthritis of the knee	Single record
Wright, 2014	To investigate the use of opioids in older adults aged ≥65 years with KOA in 2003, 2006 and 2009 in the United States	KOA cases were defined as those with ICD-9 diagnostic codes for OA of the knee (715.x6) OR knee pain (719.46) plus unspecific OA codes 715.x8, 715.x9 or 715.x0	Single record OR >1 record
Yu, 2017	To determine trends in the rate and pharmacological management of new cases of OA (including KOA) in the United Kingdom.	Two separate case definitions: **Clinical OA**: using joint pain codes including knee pain codes. **Diagnosed OA:** including diagnostic codes for KOA	Single record
Thorlund, 2019	To quantify opioid use in patients with KOA and hip OA in Sweden	ICD-10-diagnostic codes: M17 corresponding to osteoarthritis of the knee	Single record

Abbreviations: ICD, International Classification of Diseases; KOA, knee osteoarthritis.

2 | METHODS

2.1 | Study design and data source

This observational study applied a cross-sectional design between 2000 and 2015 using data from the CPRD. CPRD is a large primary care longitudinal electronic medical records database that contains anonymous records of patients from general practices across the United Kingdom (including England, Scotland, Wales and Northern Ireland) who have agreed at a practice level to provide data on a monthly basis.[14] Around the study period, 674 practices contributed data to CPRD GOLD and records of more than 4.4 million active patients (alive and contribute data to CPRD) were included. Recording of diagnosis is mandatory for every consultation, and there is no limit on the number of diagnoses entered. The database contains information on symptoms, diagnoses, prescriptions, referrals, tests, immunisation, lifestyle factors, and information on medical staff.[14]

Substantial research has been undertaken to investigate the validity and completeness of CPRD data and has provided satisfactory results.[15] Further information on CPRD can be obtained from https://cprd.com/primary-care.

2.2 | Knee OA code list development

Code list development process started with the generation of a list of osteoarthritis-diagnosis-related codes using different Read terms,[16] including: *arthritis*, *osteoarthritis* and *arthrosis* through a search of the medical browser dictionary of CPRD. Additionally, a separate search for Read codes starting with N05* (osteoarthritis and allied disorders) was also performed using the browse function of CPRD. The results of both searches were combined to generate a single list containing all OA relevant codes, while the codes for other arthritis conditions (e.g., rheumatoid arthritis) were saved separately.

Alongside, a list of Read codes for OA (including KOA) was compiled from a free clinical codes repository established by the University of Manchester (www.clinicalcodes.org) and from the supplementary code files of relevant publications.[17] A list of OA Read codes was then generated by merging the codes derived from both sources and removing any duplicates. Knee joint pain codes were compiled using a similar method (Table S1).

2.3 | Cohort definition strategies

The present study applied five different cohort definition strategies to inform the justification of the final cohort selection. These strategies differed in their sensitivity to define cases and subsequently to select cohorts of patients with KOA. The strategies were the result of the application of KOA diagnostic codes, symptoms codes or both, as summarised in Table 3 and Figure 1.

The set of Read/medcodes corresponding to each strategy was then applied in the define tool of the CPRD.

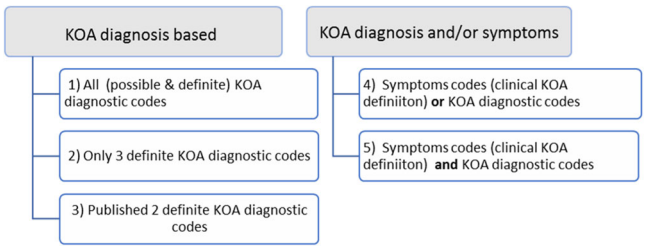

FIGURE 1 Summary of the five cohort selection strategies.

TABLE 3 Cohort selection strategies applied for final cohort selection.

Strategy	Code type	Description	Read terms used	No. records[a]
1	KOA diagnostic	Nine codes (definite and possible KOA cases)	– Localised OA of the lower leg (3 codes) – Arthrosis (2 codes) – Knee OA – OA of Knee, NOS – Patellofemoral OA – Tibiofibular OA	Single
2	KOA diagnostic	Three KOA specific codes (definite codes only)	– Knee OA – OA of Knee, NOS – Patellofemoral OA	Single
3	KOA diagnostic	Two KOA specific codes (definite codes only) obtained from published work	– Knee OA – OA of Knee, NOS	Single
4	KOA symptoms **or** KOA diagnostic	This definition ensured maximum sensitivity	Symptom-based Read terms were: knee pain, arthralgia of the knee, knee joint pain **or** Read terms from strategy 1	Single
5	KOA symptoms **and** KOA diagnostic	This definition ensured maximum specificity	Read terms from strategy 1 **and** 4	>1 record

Abbreviations: KOA, knee osteoarthritis; NOS, not otherwise specified.
[a]Number of records of relevant KOA codes required for a case definition.

2.4 | Inclusion criteria

Patients were eligible for inclusion in this study if:

1. They had a recorded diagnostic/or symptoms based Read Code of KOA in their clinical or referral or consultation records of CPRD.
2. The diagnosis/symptom was recorded between 1st January 2000 and 31st December 2015 and recorded at the age of 18 years or over.
3. The diagnosis/symptom occurred at least 12 months after registration and was recorded on or after the practices' up to standard (uts) date (date at which the practice data is deemed to be of research quality).
4. The diagnosis/symptom of KOA must also have been recorded before the earliest of; end of the study date (31st December 2015), transfer out date (date the patient transferred out of the practice) or death date.

2.5 | Exclusion criteria

Patients who were <18 years at KOA diagnosis/symptom and those who were registered for less than a year before study initiation were not included. Patients with a diagnosis of inflammatory arthritis, such as rheumatoid arthritis and systemic lupus erythromatosus (SLE), identified by appropriate Read codes, were excluded.

Patients with KOA who had a recorded cancer diagnosis at any time during follow-up were selected for the purpose of exclusion from drug utilisation analyses. Rheumatoid arthritis, (SLE) and cancer diagnoses codes were obtained from the University of Cambridge[18] and from the Quality and Outcome Framework[19] (Table S1).

2.6 | Study outcomes

This study measured the following outcomes:

1. The total number of patients with KOA during the study period.
2. The annual incidence of KOA between the years 2000 and 2015.
3. The length of follow-up in years and the proportion of patients with relatively long and short follow-up periods.

2.7 | Data extraction

The finalised KOA-related Read codes were then applied in the 'Define' tool of CPRD GOLD, and subsequently a cohort of eligible patients was selected, i.e. a list of patients resulted from the define function.

The list of patients generated in the define tool was then used in the 'Extract' tool to retrieve all the data associated with these patients, including multiple text files extracted from clinical, consultation, immunisation, patient, practice, referral, staff, test, therapy and additional files in the CPRD. Except immunisation files, all the aforementioned records were extracted from the September 2016 build of data.

2.8 | Statistical analysis

Analyses were carried out using Stata 15.1 (StataCorp. 2017). This research was approved by the Independent Scientific Advisory Committee (protocol number 18_170R). CPRD data files were downloaded in a zipped text format from the CPRD Gold interface. Data cleaning involved data inspection for missing information or outliers. Patient records with missing year of birth (yob) were excluded. Patients who consulted more than once with that code were only counted once. The annual incidence of KOA diagnosis was calculated with an estimated at-risk population from the CPRD population during the respective year.

3 | RESULTS

3.1 | Number of selected cases through cohort selection strategies

The numbers of selected cohorts varied with the applied strategies, as illustrated in Figures 2 and 3, Figure S1 and Table S2.

Figures 2 and 3 illustrate the difference in the number of patients obtained by the application of various cohort selection strategies. The selection of the highest or lowest estimates for the final KOA cohort may result in the over- or under-estimation of the true numbers. Hence, strategies 4 and 5 were not considered for final cohort selection. In between were the diagnosis-based strategies (Figure S1), which were judged to be most appropriate for the final cohort selection of this research.

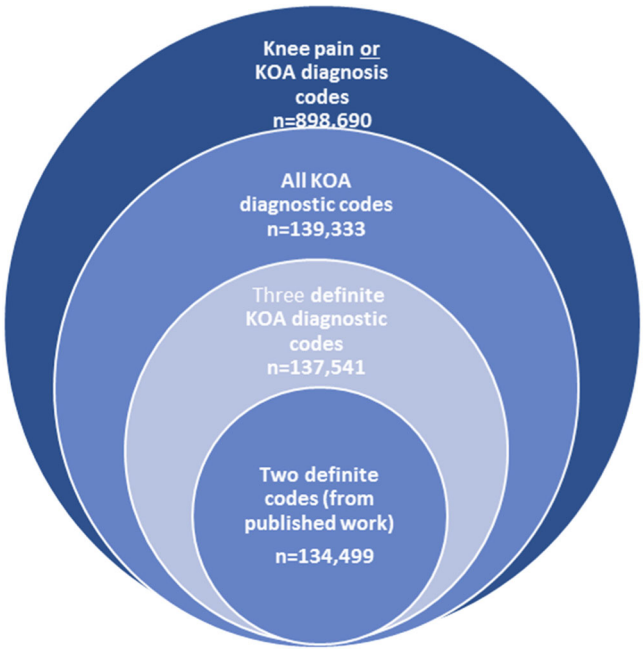

FIGURE 2 Number of patients selected by the application of strategies based on diagnostic codes **or** symptoms codes.

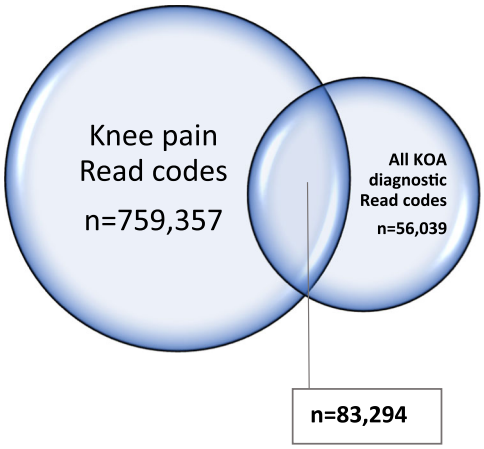

FIGURE 3 Number of patients with KOA selected with the application of diagnostic and symptoms codes. KOA, knee osteoarthritis.

3.2 | Final cohort selection

Among strategies 1–3, the final cohort selection was founded on a balanced approach, i.e. the intermediate one seen in strategy 2, which revealed 137,541 patients (Figures 2 and 3). This avoided a potential under-estimation by excluding true KOA cases (as in strategy 3), and at the same time avoided the unnecessary inclusion of patients with a possible (not definite) diagnosis of KOA (as in strategy 1).

There were 490 patients diagnosed outside the uts date and these were excluded from the selected cohort, leaving a final cohort of 137,051 patients (Figure S2).

In summary, the final study cohort included all CPRD patients who had at least one KOA diagnosis record (corresponding to any of the three selected Read codes) in their clinical files between 1st January 2000 and 31st December 2015. The overall approach for cohort selection was endorsed by the consultant clinician on the research team and by an external assessor.

3.3 | Demographic characteristics of the identified patients with KOA

A total of 125,096 out of the total defined cohort had a new diagnosis of KOA within the study period and their demographic characteristics are presented in Table S3 and Figure S3.

3.4 | Annual incidence of KOA diagnosis

The annual number of patients with an incident diagnosis of KOA during the study years ranged between 5098 in 2000 and 9274 patients in 2008. The number of patients with an incident diagnosis of KOA per 1000 CPRD registrants (Figure S4) ranged from 1.33 per 1000 CPRD registrants in 2000, 1.76 in 2008 and 1.45 patients in 2015 The incidence of KOA remained stable through the years between 2009 and 2015 at 1.73 to 1.45 per 1000 CPRD registrants, respectively.

Compared to males, the incidence of KOA diagnosis was higher in females and among those aged 65–80 years throughout the study period (Figures S5 and S6) (The median follow-up since KOA diagnosis was 5.9 years (interquartile range: 2.9, 9.6) as demonstrated in Figure S7.

3.5 | Selection of patients with cancer diagnoses

Within the selected cohort of patients with KOA (n = 137,051), those with any recorded cancer diagnoses were selected. The number of identified patients with at least one recorded cancer diagnosis was 19,414 (14.1%). The remaining final number of patients with KOA was 117,637 (after excluding patients with cancer diagnoses).

4 | DISCUSSION

4.1 | Main findings

This study selected a total of 137,051 patients with a clinician-recorded diagnosis of KOA between 1st January 2000 and 31st December 2015. The cohort selection was concluded (finalised) after comparing the resultant case counts of several case-finding strategies. The overall incidence ranged between 1.3 and 1.4 per 1000 CPRD registrants in 2000 and 2015, respectively.

4.2 | Justification of final cohort selection strategy

Due to the potential risk of over-estimation of the true case counts, this study did not consider clinical definitions (knee pain/symptoms) for the final cohort selection and used diagnosis-based codes instead. Higher numbers of patients with a clinical definition compared to diagnosed OA was a consistent finding in the literature; with a total of 1,716,253 incident cases of clinical OA compared to 432,163 cases of diagnosed OA in a study using CPRD data.[3] The estimated incidence rates were 47.7 per 1000 person-years (95% CI: 47.4, 47.9) with a clinical OA definition compared to 7.9 (95% CI: 7.8, 8.0) with diagnosed OA in 2013.[3] Similarly, Jordan et al.[7] found 616 diagnosed OA patients compared to 811 with a clinical definition, out of a total of 13,831 patients in their cohort study using data from a regional EHR database.

Among the diagnosis-based strategies (1, 2 and 3), strategy 2 provided a balance across the applied strategies as it avoided both over- and under-estimation of the true counts and was therefore a justified approach for cohort selection for further analyses on drug utilisation and associated outcomes.[12,13]

4.3 | Comparison of final cohort selection method and resulting case counts with other studies

In contrast to the more restrictive case-definition algorithms implemented in studies requiring multiple records of KOA codes to

ascertain diagnosis,[20] the present study required a single record of KOA diagnosis within the CPRD during the study period. A single OA diagnosis code was a standard definition used in several previous epidemiological studies[3,21,22] including drug utilisation studies[4] and evidence showed that only 1.3% of cases defined in this way are subsequently given a different diagnosis.[1] This provided further reassurance on the adopted final cohort selection strategy.

Similar to the case definitions used in previous drug utilisation studies in Europe and the United Kingdom, the present study applied diagnostic codes to define cases with KOA in CPRD. In Spain, Wilson et al.[4] found a total of 96,222 patients with diagnosed KOA over a period of 5 years, which represented 2.9% from the total registered population in the SIDIAP (Sistema de Información para el Desarrollo de la Investigación en Atención Primaria) database and aged 40 years or over ($n = 3,266,862$).[4] In the United Kingdom, 432,163 patients with diagnosed OA were identified between 1992 and 2013 using CPRD data; however, joint specific prevalence estimates were not published.[3]

The present study selected 137,541 patients, who represented 3.4% of the CPRD population (CPRD population $n = 4,000,000$ patients). This overall prevalence over a period of 16 years was comparable to that reported in the Spanish study (2.9%).[4]

4.4 | Comparison of the annual incidence of KOA with other studies

The incidence of KOA reported in the present study was broadly similar to estimates reported in other studies from Europe and Canada, where the incidence of KOA was reported to be 1.5 per 1000 persons in 2001,[21] similar to the estimated incidence of 1.48 per 1000 CPRD registrants in the same year obtained from the present study. However, the reported KOA incidence in the present study was lower than that estimated in a previous UK-based study 2.9 per 1000 persons in the year 2010[22] compared to 1.67 per 1000 CPRD registrants estimated in the present study for the year 2010. Such a difference can be attributed to the fact that the data in the study by Yu et al.[22] were obtained from a local primary care database in North Staffordshire in England which includes 11 general practices within a confined region of England (total practice patient population of 94,955).[2] Therefore, the differences in the included population may have led to the differences in incidence estimates between the published study[22] and the present study.

The present study reported a varied incidence of diagnosed KOA across age ranks, with patients aged between 65 and 80 years being the most prominent group. This finding was in line with findings of OA in the United Kingdom.[22] Similarly, the incidence varied with age and was highest in those aged between 65 and 80 years in a Canadian study population.[20]

Patients with KOA and cancer diagnosis in their entire record were selected. This was for the purpose of excluding these patients and their corresponding prescriptions from further drug utilisation analyses due to variations in drug utilisation particularly opioids[23] leading to potential biased results of drug utilisation prevalence and outcomes.

4.5 | Strengths and limitations

In the present study, a patient was labelled as an incident case with the maximum look-back period. The look-back period is an observation time required to eliminate prevalent KOA cases.[24] The present study has some limitations; the incidence of KOA reported in the present study may potentially be underestimated as it only reflects the estimated consultation incidence (the rate of new cases presenting to primary care). Nevertheless, the primary care setting is the setting where the majority of patients with OA or KOA are diagnosed and managed.

While acknowledging the lack of validation against a gold standard (e.g., chart review), this study had mitigated the potential associated limitations. The incidence of KOA diagnosis was generally comparable and consistent with the estimates reported from other countries using EHRs. The demographic characteristics and the variation of incidence of KOA diagnosis with gender and age were also in line with estimates published in the literature, which strengthened the validity of the chosen cohort.

The study had transparently reported a cohort selection process that was based on prior research and relevant to the purpose of drug utilisation research using CPRD. The validation and validity of diagnoses in CPRD was investigated through performing a systematic literature review that investigated 183 different diagnoses including musculoskeletal disorders. The median proportion of cases with a confirmed diagnosis was 89% (range from 24% to 100%) and 85% utilised data from outside the CPRD to validate diagnoses.[15] Leveraging CPRD data allowed efficient and systematic selection of eligible patients which enhanced the credibility of the selected cohort. The impact of alternative cohort selection criteria (i.e., application of knee pain related medcodes) was assessed by assessing results across different selection approaches which further strengthened confidence in the validity of the chosen cohort. Finally, the cohort selection process was endorsed by experts who were familiar with the database and were involved in OA research.[25]

5 | CONCLUSION

The present study selected a valid cohort of patients with a clinician-recorded KOA diagnosis from CPRD data ($n = 137,051$). The final cohort selection was informed by the case counts obtained from several case-finding strategies, which led to minimised KOA misclassification in the final cohort. To validate the identified cohort, the annual incidence was estimated between 2000 and 2015, and it was found to be in general agreement with published data.

AUTHOR CONTRIBUTIONS

Aqila Taqi initiated and developed the research questions, accessed the research data, conducted data management, data analysis, and led on drafting the manuscript. Roger David Knaggs and Sonia Gran advised on the study design and data analysis. All of the authors contributed to the interpretation of the data, critically revised the manuscript, and approved the final version submitted for publication.

ACKNOWLEDGEMENTS

AT was funded by a PhD scholarship from the Ministry of Higher Education, Oman. This article was funded by the University of Nottingham, UK. Funders had no role in the design and execution of the study.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are owned by the CPRD and may be available subject to data sharing regulations of the CPRD.

ORCID

Aqila Taqi http://orcid.org/0000-0003-0933-9003

REFERENCES

1. Prieto-Alhambra D, Judge A, Javaid MK, Cooper C, Diez-Perez A, Arden NK. Incidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: influences of age, gender and osteoarthritis affecting other joints. *Ann Rheum Dis*. 2014;73(9):1659-1664.
2. Reyes C, Garcia-Gil M, Elorza JM, et al. Socio-economic status and the risk of developing hand, hip or knee osteoarthritis: a region-wide ecological study. *Osteoarthritis Cartilage*. 2015;23(8):1323-1329.
3. Yu D, Jordan KP, Bedson J, et al. Population trends in the incidence and initial management of osteoarthritis: age-period-cohort analysis of the clinical practice research datalink, 1992-2013. *Rheumatology*. 2017;56(11):1902-1917.
4. Wilson N, Sanchez-Riera L, Morros R, et al. Drug utilization in patients with OA: a population-based study. *Rheumatology*. 2015;54(5):860-867.
5. Thorlund JB, Turkiewicz A, Prieto-Alhambra D, Englund M. Opioid use in knee or hip osteoarthritis: a region-wide population-based cohort study. *Osteoarthritis Cartilage*. 2019;27:871-877.
6. Wright EA, Katz JN, Abrams S, Solomon DH, Losina E. Trends in prescription of opioids from 2003–2009 in persons with knee osteoarthritis. *Arthritis Care Res (Hoboken)*. 2014;66(10):1489-1495.
7. Jordan KP, Tan V, Edwards JJ, et al. Influences on the decision to use an osteoarthritis diagnosis in primary care: a cohort study with linked survey and electronic health record data. *Osteoarthritis Cartilage*. 2016;24(5):786-793.
8. Yu D, Jordan K, Peat G. Underrecording of osteoarthritis in United Kingdom primary care electronic health record data. *Clin Epidemiol*. 2018;10:1195-1201.
9. Shrestha S, Dave AJ, Losina E, Katz JN. Diagnostic accuracy of administrative data algorithms in the diagnosis of osteoarthritis: a systematic review. *BMC Med Inform Decis Mak*. 2016;16:82.
10. Coleman N, Halas G, Peeler W, Casaclang N, Williamson T, Katz A. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. *BMC Fam Pract*. 2015;16:11.
11. Williamson T, Green ME, Birtwhistle R, et al. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Ann Fam Med*. 2014;12(4):367-372.
12. Taqi A, Gran S, Knaggs RD. Current use of analgesics and the risk of falls in people with knee osteoarthritis: a population-based cohort study using primary care and hospital records. *Osteoarthr Cartil Open*. 2021;3(2):100165.
13. Taqi A, Gran S, Knaggs RD. Analgesic utilization in people with knee osteoarthritis: a population-based study using primary care data. *Pain Practice*. 2023;23(5):523-534.
14. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827-836.
15. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the general practice research database: a systematic review. *Br J Clin Pharmacol*. 2010;69(1):4-14.
16. Jordan K, Clarke AM, Symmons DP, et al. Measuring disease prevalence: a comparison of musculoskeletal disease using four general practice consultation databases. *Br J Gen Pract*. 2007;57(534):7-14.
17. Kontopantelis E, Olier I, Planner C, et al. Primary care consultation rates among people with and without severe mental illness: a UK cohort study using the clinical practice research datalink. *BMJ Open*. 2015;5(12):e008650.
18. CPRD at Cambridge/CPRD @ Cambridge – Code Lists/CPRD @ Cambridge – Codes Lists Version 1.1 – October 2018. [Internet]. 2018. Available from: https://www.phpc.cam.ac.uk/pcu/cprd_cam/codelists/v11/
19. NHS. *Quality and Outcomes Framework QOF business rules v36.0 UK2017*. Available from: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/quality-and-outcomes-framework-qof/quality-and-outcome-framework-qof-business-rules/quality-and-outcomes-framework-qof-business-rules-v36-0
20. Rahman MM, Kopec JA, Goldsmith CH, Anis AH, Cibere J. Validation of administrative osteoarthritis diagnosis using a clinical and radiological population-based cohort. *Int J Rheumatol*. 2016;2016:1-7.
21. van der Waal JM. The incidences of and consultation rate for lower extremity complaints in general practice. *Ann Rheum Dis*. 2006;65(6):809-815.
22. Yu D, Peat G, Bedson J, Jordan KP. Annual consultation incidence of osteoarthritis estimated from population-based health care data in England. *Rheumatology*. 2015;54(11):2051-2060.
23. Zin CS, Chen LC, Knaggs RD. Changes in trends and pattern of strong opioid prescribing in primary care. *Eur J Pain*. 2014;18(9):1343-1351.
24. Kopec JA, Rahman MM, Sayre EC, et al. Trends in physician-diagnosed osteoarthritis incidence in an administrative database in British Columbia, Canada, 1996–1997 through 2003–2004. *Arthritis Rheum*. 2008;59(7):929-934.
25. Ferguson RJ, Prieto-Alhambra D, Walker C, et al. Validation of hip osteoarthritis diagnosis recording in the UK clinical practice research datalink. *Pharmacoepidemiol Drug Saf*. 2019;28(2):187-193.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Taqi A, Gran S, Knaggs RD. "Application of five different strategies to define a cohort of patients with knee osteoarthritis in a large primary care database". *J Eval Clin Pract*. 2024;1-7. doi:10.1111/jep.14045