Validation of knee osteoarthritis case identification algorithms in a large electronic health record database

Michelle S. Yau\textsuperscript{a,b}, Maureen Dubreuil\textsuperscript{c}, Shanshan Li\textsuperscript{c}, Vibha Inamdar\textsuperscript{c}, Christine Peloquin\textsuperscript{c}, David T. Felson\textsuperscript{c,d,}\textsuperscript{*}

\textsuperscript{a} Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA, USA
\textsuperscript{b} Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA
\textsuperscript{c} Section of Rheumatology, Department of Medicine, Boston University School of Medicine, Boston, MA, USA
\textsuperscript{d} University of Manchester, Manchester, UK

ARTICLE INFO

Keywords:
Knee osteoarthritis
Electronic health records
Validation
Algorithms

ABSTRACT

Purpose: To facilitate studies of knee osteoarthritis (OA) in large databases, case finding algorithms with high levels of diagnostic performance are needed.

Methods: From a UK general practitioner (GP) practice derived database, we selected adults ages 40–90 years meeting algorithms that included various combinations of codes for knee OA or knee pain and imaging. The GP for each patient was mailed a questionnaire to assess the cause of knee pain and provide knee x-ray and/or MRI findings. We considered knee pain with x-ray and/or MRI findings consistent with OA the gold standard. We calculated positive predictive values (PPV) and sensitivity for case identification algorithms.

Results: Of 100 questionnaires sent, 93 were returned; we excluded 8 subjects who had other rheumatic disorders or total knee replacements. Among those with one code for OA, the PPV was 64% (95% CI = 49%–79%) and it increased to 92% (95% CI = 76%–100%) when two or more OA codes over six months were required. The increase in PPV was accompanied by a drop in sensitivity from 44% (95% CI = 31%–57%) to 19% (95% CI = 9%–30%). Use of one pain code yielded similar results to use of one OA code. Requiring two or more knee pain codes over six months yielded a PPV of 68% (95% CI = 49%–88%) and sensitivity of 26% (95% CI = 15%–38%).

Discussion: A case identification algorithm requiring two or more knee OA codes yielded the highest PPV at the cost of reduced sensitivity. Tradeoffs between PPV and sensitivity will need to be weighed alongside study goals when selecting a case identification algorithm.

1. Introduction

Osteoarthritis (OA) is the most common form of arthritis worldwide. For knee OA, one of the most common and symptomatic sites for disease, there is no medical treatment that delays disease progression and patients often undergo knee replacement surgery. Rates of total joint replacement surgeries are increasing exponentially in the United States with estimates that over 1.6 million knee replacements will be performed annually by the year 2030 [1]. The increasing rates of joint replacement surgeries for OA and the costs of disability from OA are growing causes of concern.

Large observational studies, including the Multicenter Osteoarthritis Study (MOST) and Osteoarthritis Initiative (OAI) have provided many new insights into OA, but are limited to a well-defined cohort, often selected for OA risk factors. To further accelerate the discovery of novel OA risk factors and insights into management strategies, even larger studies will be needed. Many newly developed large-scale resources like the United Kingdom (UK) Biobank and All of Us are population-based studies that are linked to electronic health records (EHRs), potentially providing information on thousands of persons with OA. Additionally, EHR data may complement other data sources by providing an assessment of real-world medication utilization and effectiveness. These resources provide much larger sample sizes than can be reasonably recruited in an observational study or clinical trial.

To the best of our knowledge, there has been only one published study examining an algorithm to identify knee OA in claims or EHR...
databases. The one study with validation of knee OA against physician diagnosis ICD10 code was unusual in that it studied older men and women who, as part of a cohort study, had been asked about frequent knee pain and obtained knee radiographs, gold standards for OA diagnosis [2]. The positive predictive value (PPV) for a single physician diagnosis of OA in this cohort was 88%. There has also been one study of OA that combined knee, hip, and hand OA [3] and another focusing on hip OA [4]. In the study combining hand, knee and hip OA, positive predictive values (PPVs) were >80% for a single report of OA but even higher PPVs if two or more codes for OA are used [3]. The high PPVs from these studies may have been because for joints studied, most did not require imaging evidence of OA to diagnose disease. Another reason for the high PPV is that only OA codes were tested even though some patients with OA are treated without such codes ever used.

Previous studies of knee OA in EHR and claims databases have often only included one diagnostic code to identify persons with OA [5–7]. Whether this is sufficient to accurately identify OA is unclear. Single ICD-9 or ICD-10 codes for other forms of arthritis have had variable predictive value. For example, in both rheumatoid arthritis (RA) and ankylosing spondylitis, a single ICD-9 code has a low PPV [8,9]. Validation studies have suggested that for RA, cases need at least one diagnostic code plus documentation of RA treatment, or at least two diagnostic codes separated by a period of time.

Knee OA cases may be more challenging to identify in large databases than rheumatic diseases for which widely used effective drug therapies can be used to identify cases. Further, chronic knee pain in middle aged and older persons is treated the same way as knee OA and imaging may not be obtained. It is unclear whether those with chronic knee pain should be labeled as having knee OA. Most but not all criteria for knee OA require the presence of knee pain and some structural evidence of OA, usually evidenced by imaging.

We sought to evaluate case identification algorithms to identify persons with knee OA within The Health Improvement Network (THIN), a large EHR database from the UK.

1.1. Methods

The Health Improvement Network (THIN) (now called IQVIA Medical Research Data (MRD)) is an anonymized EHR database collected from general practitioners (GPs) in the United Kingdom (UK) including data on over 11 million patients, with over 45 million patient-years of data (http://www.epic-uk.org/). THIN data represent routine medical practice in a population-based setting. The dataset includes demographics, details from GP visits, specialists’ reports and hospital admissions, test results, height, weight, blood pressure, and smoking status.

We randomly selected 100 patients from THIN who were ages 40–90 years, enrolled with a GP for at least 1 year between January 1, 2000 to May 31, 2015, and had their first diagnostic code for knee OA or knee pain within the assessment period (Supplementary Table 1). To test common strategies for case identification, subjects also had to meet one of three criteria: (1) one or more diagnostic code(s) for knee OA, (2) one or more diagnostic codes for knee OA and a knee x-ray or magnetic resonance images (MRI) within 2 years of knee OA diagnosis, (3) one or more diagnostic codes for knee pain with a knee x-ray or MRI within 2 years of knee pain diagnosis. If the patient had 2 or more codes for OA and/or knee pain, these needed to be separated by at least 7 days. The final selection included 25 subjects who had OA with imaging, 25 subjects who had OA without separate codes for imaging, and 50 subjects who had knee pain with imaging. Subjects were excluded from analyses if they had diagnostic codes for rheumatoid arthritis, psoriatic arthritis, or inflammatory bowel disease. To test the validity of the algorithms used to identify knee OA, we used a gold standard comprising the presence or absence of person-level knee OA. We calculated the positive predictive value (PPV) and sensitivity and corresponding 95% confidence intervals for each algorithm using the ‘senspec’ option in PROC FREQ (PPV = % true positives/test positives; sensitivity is: of those with disease, % who test positive). All analyses were conducted using SAS 9.4 (Cary, NC).

1.2. Results

Among the 100 subjects whose GPs received questionnaires, 93 questionnaires were completed and returned. Of these, four subjects were excluded for having gout, two subjects were excluded for having rheumatoid arthritis, and one subject was excluded for having a total knee replacement prior to the first OA or pain code. Additionally, one subject was excluded due to not having enough data needed to classify the subject according to the gold standard, leaving 85 subjects for analyses. Mean (Standard deviation (SD)) age was 63.8 (11.1) and about 53% were women (Table 1). There were 60% with one or more OA codes, 78% with one or more pain codes, and 73% with one or more imaging codes. The frequency of knee OA as defined by the gold standard was 67%.

Among subjects who had only one OA or pain code, the PPV was 64.1% (95% CI = 49.0%–79.2%) and 65.9% (95% CI = 51.9%–79.9%).
respectively (Table 2). Sensitivities for one OA or pain code were 43.9% (95% CI = 31.0%–56.7%) and 50.9% (95% CI = 37.9%–63.9%), respectively. When we required two or more OA codes within 6 months, the PPV was 91.7% (95% CI = 76.0%–100.0%) and the sensitivity was 19.3% (95% CI = 9.1%–29.5%). When we required two or more pain codes within 6 months, the PPV was 68.2% (95% CI = 48.7%–87.6%) and the sensitivity was 26.3% (95% CI = 14.9%–37.7%). When we required one or more OA codes and one or more pain codes within 6 months, the PPV was 71.9% (95% CI = 56.3%–87.5%) and the sensitivity was 40.4% (95% CI = 27.6%–53.1%). Extending the time-period between diagnostic codes to 12 months did not substantially change results.

We repeated analyses among individuals who had documented imaging within 2 years of a knee OA or knee pain diagnostic code. Of the 85 subjects included in the analyses, 62 subjects had separate codes that identified knee x-ray or MRI imaging. Among these individuals who had only one OA or pain code, the PPV was 73.7% (95% CI = 53.9%–93.5%) and 67.6% (95% CI = 52.5%–82.7%), respectively (Table 3). Sensitivities for one OA or pain code were 32.6% (95% CI = 18.6%–46.6%) and 58.1% (95% CI = 43.4%–72.9%), respectively. When we required two or more codes within 6 months or 12 months among individuals with knee imaging, the estimates for PPV and sensitivity remained largely similar to estimates obtained among all subjects regardless of whether they had imaging or not.

Our patient selection included 25 individuals who had no codes for imaging. For more than half of these individuals, imaging results were either present in GP notes or reported in the GP questionnaires that were sent. While numbers were small, the PPVs for OA and pain diagnostic codes among these patients were lower for those with one code (mean 56%) compared to those with two or more codes (71%), similar to findings among those who had codes for imaging.

### Table 1
Sample characteristics for individuals with the gold standard assessment (n = 85).

|                | Mean age in years (SD) | % Women | % with one or more OA codes | % with one or more pain codes | % with one or more imaging codes | % OA according to gold standard |
|----------------|------------------------|---------|----------------------------|-------------------------------|--------------------------------|-------------------------------|
|                | 63.8 (11.1)            | 53      | 60                          | 78                            | 73                             | 67                            |

### Table 2
Estimates of positive predictive value and sensitivity among individuals with the gold standard assessment, regardless of imaging status (n = 85).

| Definition                  | Number satisfying definition | Number with OA by gold standard | PPV (95% CI) | Sensitivity (95% CI) |
|-----------------------------|------------------------------|--------------------------------|--------------|----------------------|
| One code\(^a\)             |                              |                                |              |                      |
| OA code                     | 39                            | 25                            | 64.1 (49.0, 79.2) | 43.9 (31.0, 56.7)    |
| knee pain code\(^b\)       | 44                            | 29                            | 65.9 (51.9, 79.8) | 50.9 (37.9, 63.9)    |
| ≥2 codes within 6 months   |                              |                                |              |                      |
| ≥2 OA codes                | 12                            | 11                            | 91.7 (76.0, 100.0)| 19.3 (9.1, 29.5)     |
| ≥2 knee pain codes\(^a\)  | 22                            | 15                            | 68.2 (48.7, 87.6) | 26.3 (14.9, 37.7)    |
| ≥1 OA and ≥1 knee pain code| 32                            | 23                            | 71.9 (56.3, 87.5)  | 40.4 (27.6, 53.1)    |
| ≥2 codes within 12 months  |                              |                                |              |                      |
| ≥2 OA codes                | 14                            | 12                            | 85.7 (67.4, 100.0)| 21.1 (10.5, 31.6)    |
| ≥2 knee pain codes\(^a\)  | 30                            | 20                            | 66.7 (49.8, 83.5)  | 35.1 (22.7, 47.5)    |
| ≥1 OA and ≥1 knee pain code| 32                            | 23                            | 71.9 (56.3, 87.2)  | 40.4 (27.6, 53.1)    |

\(^a\) Exactly one code.

\(^b\) All those with knee pain codes only had to have imaging ordered.

### Table 3
Estimates of positive predictive value and sensitivity among individuals with the gold standard assessment and who also were selected because they had imaging (n = 62).\(^c\)

| Definition                  | Number satisfying definition | Number with OA by gold standard | PPV (95% CI) | Sensitivity (95% CI) |
|-----------------------------|------------------------------|--------------------------------|--------------|----------------------|
| One code\(^c\)             |                              |                                |              |                      |
| OA code                     | 19                            | 14                            | 73.7 (53.9, 93.5) | 32.6 (18.6, 46.6)    |
| knee Pain code             | 37                            | 25                            | 67.6 (52.5, 82.7) | 58.1 (43.4, 72.9)    |
| ≥2 codes within 6 months   |                              |                                |              |                      |
| ≥2 OA codes                | 9                             | 8                             | 88.9 (68.4, 100.0)| 18.6 (7.0, 30.2)     |
| ≥2 knee pain codes         | 21                            | 14                            | 66.7 (46.5, 86.8)  | 32.6 (18.6, 46.6)    |
| ≥1 OA and ≥1 knee pain code| 24                            | 18                            | 75.0 (57.7, 92.3)  | 41.9 (27.1, 56.6)    |
| ≥2 codes within 12 months  |                              |                                |              |                      |
| ≥2 OA codes                | 11                            | 9                             | 81.8 (59.0, 100.0)| 20.9 (8.8, 33.1)     |
| ≥2 knee pain codes         | 28                            | 19                            | 67.9 (50.6, 85.2)  | 44.2 (29.3, 59.0)    |
| ≥1 OA and ≥1 knee pain code| 24                            | 18                            | 75.0 (57.7, 92.3)  | 41.9 (27.1, 56.6)    |

\(^c\) All those with knee pain codes only had to have imaging ordered.

\(^d\) As noted in text, many selected without imaging turned out to have imaging with results described in chart review or x-ray reports.

### 1.3. Discussion

In a validation study testing diagnostic algorithms for knee OA, we found that, like other rheumatic diseases, an algorithm that required two diagnostic codes at least 7 days apart for knee OA had a higher PPV than algorithms requiring only a single diagnostic code. If we added those with only knee pain to those with knee OA, we captured more persons with OA, increasing sensitivity, but the PPV did not change.

While diagnostic test performance may differ by setting and country (see below), our study raises questions about the validity of studies that have used EHR and claims databases to evaluate risk factors for OA or OA outcomes, especially those relying on one OA code for diagnosis [5–7]. Low PPVs for one OA code or one pain code indicate that many identified OA cases do not actually have OA or at least have not had imaging studies that would document their OA. The distinction is important because many EHR-based studies will be used for health services research, epidemiologic or genetic studies that seek to identify relevant biology that can be used to identify new treatment targets. Our results are similar to those for other rheumatic disorders including RA and spondyloarthritis, which have reported that two diagnostic codes for the disease have a higher predictive value than one code. PPVs over 70% seen in this study are higher than has been reported in studies of other rheumatic diseases using algorithms that are widely used. It is likely that PPVs for OA and other musculoskeletal disorders vary from country to country and setting to setting depending on how and where diagnosis coding is carried out. In Sweden where the caring physician enters the code and this is often a physician specializing in a specific set of diseases, PPVs may be high whereas in a system where an administrative clerk enters the code, PPVs may be lower; they may be lower in THIN where only general
M.S. Yau et al. Osteoarthritis and Cartilage Open 4 (2022) 100229

validity of analyses related to OA. However, if potential misclassi-

port higher PPVs than our study likely because the gold standard used

Given the insensitivity of x-ray changes in early OA, these studies re-

100%, the latter present when more than one diagnostic code was used.

Among important limitations was the small size of our sample. This

was driven by the cost of acquiring additional data on persons from the

THIN data which limited our numbers. High PPVs may be driven by the

high prevalence of OA in our sample and may be different in other

samples with a lower prevalence of OA. Despite these limitations, we

provide new data from THIN that provides a gold standard assessment of

OA needed to assess the use of EHR codes for identifying OA. There are

few EHR data sources that provide a gold standard assessment of OA,

which is time intensive, requiring clinical chart review or ascertainment

of additional clinical data as was done in this study. Larger studies will

be needed to confirm our results, but we expect such findings will be similar
given the consistency of our results with other rheumatic disorders.

Another possible limitation is that we required knee pain and pres-
ence of knee x-ray and/or MRI findings consistent with OA to meet the
gold standard definition. Importantly, OA may be diagnosed and treated
without imaging. In fact, imaging is not recommended in current
guidelines mostly because treatment of chronic knee pain and treatment
of OA are similar. Further, many persons with chronic knee pain who
failed to meet our criteria, which required imaging confirmation of knee
OA, may have OA [11]. This raises challenging questions about whether
to require imaging evidence of OA in large-scale studies. In secondary
analyses, we showed that the addition of imaging findings did not sub-
stantially change PPVs compared to the same case identification algo-
rithms used without the requirement of imaging; some of these
individuals had imaging but it was not required in the case finding al-
gorithm. Knee pain symptoms can be transient or mild and may reflect
other disorders and not OA, which is why we required the presence of
imaging evidence in our gold standard definition. The decision to require
imaging evidence of OA in a claims or EHR based OA study depends on
the question posed and whether it requires that cases have unequivocal
disease. If so, imaging evidence of disease would be preferred.

Our study provides new information that should inform studies of
claims and EHR-based studies of OA. Using one claim or diagnosis code
for OA may leave many persons misclassified and may compromise the
validity of analyses related to OA. However, if potential misclassification
is not a concern and a study aims to identify as many OA cases as possible,
a non-imaging OA pain or OA code may be used since it has higher sensitivity.
Whether to prioritize PPV or sensitivity depends on the study question.
A two-phase approach may be used, where the first phase prioritizes
sample size over misclassification, yielding findings that are brought
forward to a second phase that prioritizes PPV over sample size. This may
help yield meaningful results.

In conclusion, large data base analyses targeting OA should base their
case algorithm on PPV and the sensitivity of the case finding strategy.
In settings where the PPV for one diagnostic code is low, use of two
diagnoses of OA separated by at least 7 days may be needed to ensure a
high PPV. Adding knee pain as diagnostic code has tradeoffs, with
improved sensitivity and number of cases but this may lead to a drop in
PPV. More studies on the optimal strategy to identify knee OA cases in
large datasets are needed including whether and how these are affected
by the way the data are coded and whether imaging is critical to the case
definition.

Contributions

Michelle S. Yau: design of study, drafting of article, approval of final
version. Maureen Dubreuil: design of study, drafting of article, approval
of final version. Shanshan Li: design of study, reviewing of article for
content, approval of final version. Vibha Inamdar: acquired data,
reviewing article for content, approval of final version. Christine Pelo-
quin: analysis of data, reviewing of article, approval of final version.
David T. Felson: conception and design of study, reviewing article for
content approval of final version.

Role of the funding sources

None of the funding sources had any role in study design, writing or
decision to submit results of this study.

Funding

U.S. National Institutes of Health Grants P30 AR072571 (Felson), R01
AR075356 (Yau), Pilot funding from the Boston University CTSI
(Dubreuil) and K23 AR069127 (Dubreuil).

Declaration of competing interest

The authors have no relevant conflicts of interest to disclose.

Acknowledgements

We would like to thank the general practitioners who provide re-
sponses to our survey and Dr. Kelsey I. Thomas for data entry.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://
doi.org/10.1016/j.jocarto.2021.100229.

References

[1] M. Sloan, A. Premkumar, N.P. Sheth, Projected volume of primary total joint
arthroplasty in the U.S., 2014 to 2030, J. Bone Joint Surg. 100 (17) (2018), https://
https://doi.org/10.2106/JBJS.17.01617.
[2] A. Turkiewicz, I.F. Petersson, J. Björk, et al., Current and future impact of
osteoarthritis on health care: a population-based study with projections to year
2032, Osteoarthritis Cartilage 22 (11) (2014) 1826–1832, https://doi.org/
10.1016/j.joca.2014.07.015.
[3] M.M. Rahman, J.A. Kopec, C.J. Goldsmith, A.H. Anis, J. Cibere, Validation of
administrative osteoarthritis diagnosis using a clinical and radiological population-
based cohort, International Journal of Rheumatology 2016 (2016), https://doi.org/
10.1155/2016/6475318.
[4] R.J. Ferguson, D. Prieto-Alhambra, C. Walker, et al., Validation of hip osteoarthritis
diagnosis recording in the UK clinical practice research Datalink, Pharmacoepidemiol.
Drug Saf. 28 (2) (2019) 187–193, https://doi.org/10.1002/pds.4673.
[5] G. Hiday, K.E. Åkesson, O. Melander, et al., Cardiometabolic polygenic risk scores
and osteoarthritis outcomes: a mendelian randomization study using data from the
malmo diet and cancer study and the UK Biobank, Arthritis and Rheumatology 71
(6) (2019) 925–934, https://doi.org/10.1002/art.40812.
[6] T. Funck-Brentano, M. Nethander, S. Movéré-Skritic, P. Richette, C. Ohlsson,
Causal factors for knee, hip, and hand osteoarthritis: a mendelian randomization
study in the UK Biobank, Arthritis and Rheumatology 71 (10) (2019) 1634–1641,
https://doi.org/10.1002/art.40928.
[7] U.T. Kadam, M. Blagojevic, J. Belcher, Statin use and clinical osteoarthritis in the
general population: a longitudinal study, J. Gen. Intern. Med. 28 (7) (2013)
943–949, https://doi.org/10.1007/s11606-013-2382-8.
[8] M. Dubreuil, C. Peloquin, Y. Zhang, H.K. Choi, R.D. Inman, T. Neogi, Validity of ankylosing spondylitis diagnoses in the health improvement network, Pharmacoepidemiol. Drug Saf. 25 (4) (2016), https://doi.org/10.1002/pds.3952.

[9] S.L. Thomas, C.J. Edwards, L. Smeeth, C. Cooper, A.J. Hall, How accurate are diagnoses for rheumatoid arthritis and juvenile idiopathic arthritis in the general practice research database? Arthritis Care Res. 59 (9) (2008) 1314–1321, https://doi.org/10.1002/art.24015.

[10] R. Altman, E. Asch, D. Bloch, et al., Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee, Arthritis Rheum. 29 (8) (1986) 1039–1049, https://doi.org/10.1002/art.1780290816.

[11] J. Cibere, H. Zhang, A. Thorne, et al., Association of clinical findings with pre-radiographic and radiographic knee osteoarthritis in a population-based study, Arthritis Care Res. 62 (12) (2010) 1691–1698, https://doi.org/10.1002/acr.20314.