Prediction models to estimate the future risk of osteoarthritis in the general population: a systematic review.

APPLEYARD, Tom, THOMAS, Martin J, ANTCLIFF, Deborah and PEAT, George <http://orcid.org/0000-0002-9008-0184>

Arthritis Care & Research
Vol. 0, No. 0, Month 2023, pp 1–13
DOI 10.1002/acr.25035

AMERICAN COLLEGE of RHEUMATOLOGY
Empowering Rheumatology Professionals

REVIEW

Prediction Models to Estimate the Future Risk of Osteoarthritis in the General Population: A Systematic Review

Tom Appleyard,[1] Martin J. Thomas,[2] Deborah Antcliff,[3] and George Peat[4]

Objective. To evaluate the performance and applicability of multivariable prediction models for osteoarthritis (OA).

Methods. This was a systematic review and narrative synthesis using 3 databases (EMBASE, PubMed, and Web of Science) from inception to December 2021. We included general population longitudinal studies reporting derivation, comparison, or validation of multivariable models to predict individual risk of OA incidence, defined by recognized clinical or imaging criteria. We excluded studies reporting prevalent OA and joint arthroplasty outcome. Paired reviewers independently performed article selection, data extraction, and risk-of-bias assessment. Model performance, calibration, and retained predictors were summarized.

Results. A total of 26 studies were included, reporting 31 final multivariable prediction models for incident knee (23), hip (4), hand (3) and any-site OA (1), with a median of 121.5 (range 27–12,803) outcome events, a median prediction horizon of 8 years (range 2–41), and a median of 6 predictors (range 3–24). Age, body mass index, previous injury, and occupational exposures were among the most commonly included predictors. Model discrimination after validation was generally acceptable to excellent (area under the curve = 0.70–0.85). Either internal or external validation processes were used in most models, although the risk of bias was often judged to be high with limited applicability to mass application in diverse populations.

Conclusion. Despite growing interest in multivariable prediction models for incident OA, focus remains predominantly on the knee, with reliance on data from a small pool of appropriate cohort data sets, and concerns over general population applicability.

INTRODUCTION

In recent years, risk prediction models have grown increasingly popular and are used to estimate the likelihood of the incidence of health-related outcomes. These models assist clinicians, complementing clinical decision-making and aiding the provision of information to patients. The models also contribute to public health, identifying future health care needs for the wider at-risk population (1). With older people constituting a growing proportion of the global population, disease burden is increasingly associated with noncommunicable diseases, for example cardiovascular, cancer, diabetes mellitus, and musculoskeletal disorders (2). Models predicting an individual's future risk of developing these conditions, permitting the modification of risk factors while patients remain free of disease, may contribute to their prevention. Preventing noncommunicable disease is a global priority. In 2015 the UN Sustainable Development Goals program outlined this prevention as a 15-year aim (3). While risk prediction models have been derived, validated, and implemented in clinical medicine and public health screening programs to predict the incidence of cardiovascular disease (4–6), their use in musculoskeletal disorders, such as osteoarthritis (OA),

The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Drs. Thomas and Antcliff's work was supported by the NIHR (Development and Skills Enhancement awards NIHR 300818 and NIHR 301005, respectively).

[1]Tom Appleyard, MBChB, MPhil: Keele University, Staffordshire, UK; [2]Martin J. Thomas, PhD, MCSP: Keele University and Midlands Partnership NHS Foundation Trust, Staffordshire, and Haywood Hospital, Burslem, UK; [3]Deborah Antcliff, PhD: Keele University, Staffordshire, Northern Care Alliance NHS Foundation Trust, Bury Care Organisation, Manchester, and University of Leeds, Leeds, UK; [4]George Peat, PhD: Keele University, Staffordshire, and Sheffield Hallam University, Sheffield, UK.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.25035&file=acr25035-sup-0001-Disclosureform.pdf.

Address correspondence via email to Tom Appleyard, MBChB, MPhil, at t.appleyard@outlook.com.

Submitted for publication May 15, 2022; accepted in revised form October 4, 2022.

SIGNIFICANCE & INNOVATIONS

- Prediction models support earlier intervention in several diseases, but to date, not in osteoarthritis (OA).
- Our systematic review provides a comprehensive and critical synthesis of 31 published multivariable models derived by international research teams to predict the future individual risk of developing OA.
- We found generally good performance and evidence of increasing use of internal and external validation. However, a focus on knee OA, a reliance on a restricted number of cohort data sets, mainly from higher-income countries, and use of data sources that may be challenging to scale up in routine practice, may limit the applicability of many existing prediction models in general populations.
- The emergence of prediction modeling using routine health care data and improvements in analytic methods may help address some, but not all, of these limitations.

remains uncommon. The current systematic review was motivated by a desire to understand how close such a prospect may be in OA and what remaining limitations may need to be addressed.

OA is a chronic, painful condition that poses significant challenges to public health. In recent years, the disability-adjusted life-years associated with OA have risen markedly, estimated by the Global Burden of Disease project to have increased 34% between 1990 and 2015 (2). OA also has significant impacts on health care utilization, including surgical intervention (7,8), and total costs are estimated to represent up to 0.5% of high-income nations' gross domestic product (9). With health care–associated costs of OA predicted to rise (10,11), validated risk prediction models are needed to identify high-risk patients, permitting the communication of risk, stratification of care, and attempts at risk-informed prevention. Furthermore, models to predict disease incidence may provide insight into the classification and diagnosis of early OA, which has increasing interest due to its chronic, progressive nature, alongside a growing focus on the prognosis, rather than solely the diagnosis, of the condition (12,13).

Several studies have developed risk prediction models for OA outcomes, but to date, no systematic synthesis of this evidence has been published. Our systematic review identifies and critically synthesizes published studies deriving and validating multivariable risk prediction models for predicting individualized risk of OA incidence within general populations. The motivating questions for our review were to summarize currently published models, evaluating their applicability to large-scale use in clinical practice.

MATERIALS AND METHODS

Literature search. We conducted preliminary literature searches before finalizing our search strategy and specifying our research protocol following the Preferred reporting Items for Systematic Review and Meta-analysis Protocols guidelines (14) (PROSPERO registration number: 4220446; approved November 2020). We searched PubMed, EMBASE, and Web of Science from inception to December 2021. Our searches used the modified Ingui filter, a generic filter for clinical prediction modified by Geersing et al for greater sensitivity, together with condition-specific terms relating to OA (15). To increase specificity for risk, rather than prognosis/progression, the terms [onset OR inciden*] were also applied (for full search strategy, see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25035). We screened the reference lists of included articles and published abstracts of recent international annual conferences of the Osteoarthritis Research Society International (conference years 2017–2019), the European Alliance of Associations for Rheumatology (2018–2020) and the American College of Rheumatology (2018–2020) to identify models in development that may have subsequently been published.

Eligibility criteria. We included any original study of a longitudinal design (including randomized controlled trials, cohort and [nested] case–control studies) conducted in a general population sample that developed, compared, or validated a multivariable prediction model to predict an individual's risk of future OA incidence, irrespective of the time span for prediction. Articles presenting a clinical prediction score based on a model, as well as those evaluating prediction model impact, were also eligible. Eligible definitions of OA included symptomatic, radiographic, and symptomatic-radiographic. We excluded studies of hospital inpatients and other selective settings, prognostic models of patients with existing symptomatic or radiographic disease, and those using arthroplasty as the sole outcome. Cross-sectional studies, case reports/series and conference abstracts were excluded. Titles and abstracts of studies were required in English; no language restriction was placed on articles eligible for full-text review.

Screening. Search results of the 3 searched databases were exported to the reference software Rayyan (16). TA undertook deduplication. Title screening was undertaken by a single reviewer (TA, MJT, DA, or GP), with a sample of decisions checked by a second reviewer. Authors then worked in pairs (TA and MJT, and DA and GP, blinded to the other's decision) to screen abstracts, with conflicts resolved by a third reviewer (MJT or GP) not involved in the original decision. Upon full-text review, paired authors again worked independently with a third reviewer to resolve conflict; reasons for exclusion were documented.

Data extraction. Data extraction of eligible studies was performed by paired reviewers (TA and MJT, DA and GP), using a shared Microsoft Excel (17) worksheet incorporating items for extraction as outlined by Cochrane (18) and the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies checklist (19). Data extracted included general study information, followed by model-specific data, for instance relating to study design, sample size, outcome definition, and included predictors. Last, we extracted performance metrics (overall fit, discrimination, and calibration). The data extraction template is available in Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.25035. Where studies presented final models for multiple eligible outcomes, information was extracted for each model. In those with more than 1 model per outcome, a final model was identified, based on the authors' own designation or inferred from their description of the model-building process and intended application.

Risk-of-bias assessment. For all included articles, we assessed the risk of bias across 4 predetermined domains (participants, outcome, predictors, and analysis) using the Prediction study Risk of Bias Assessment Tool (20). Risk of bias and applicability were scored as low, high, or of unclear risk, with applicability appraised with respect to large-scale use in diverse general populations and in contexts where imaging may not be routinely available or recommended. Risk-of-bias assessments were conducted by 1 reviewer, checked by a second reviewer.

Narrative synthesis. Given the heterogeneous nature of study designs and model content, we conducted a narrative synthesis of results. Final models were grouped by the outcome joint of interest (index joint: knee, hip, hand, other, any) to reflect prior evidence of joint-specific risk factors and then synthesized by a specific outcome definition (e.g., radiographic, symptomatic, symptomatic-radiographic). Performance measures were summarized, with calibration relating to the agreement between predicted versus observed risk, discrimination assessing whether patients with the outcome (at a given threshold) have higher risk prediction scores (21). The area under the receiver operating characteristic curve (AUC), as the most commonly reported standard metric for discrimination, was displayed (with 95% confidence intervals [95% CIs], where reported) in a forest plot for each model across derivation (i.e., apparent performance), internal validation (assessment of performance typically within a subset of the original data set, for instance by bootstrapping or cross-validation) and external validation (different sample to derivation) phases (22). An AUC of 0.5 suggests that the model demonstrates no discrimination, ≥0.7 and ≥0.8 were deemed good and excellent, respectively, accepting that such thresholds are quite arbitrary. Predictors included in final models were tabulated and color-coded by the mode of assessment. To reduce the volume of information presented, we grouped different or multiple measurements that had been used to capture the same construct. However, a spreadsheet of this tabulation with minimal grouping of predictors was retained as supplementary data. Throughout the synthesis, studies were generally presented in the order of the year of publication to help discern trends over time. Patients and members of the public were not involved in this systematic review.

RESULTS

The search yielded 10,129 articles. After deduplication, followed by title and then abstract screening, 62 articles were taken forward for full-text screening, of which 20 were eligible for inclusion. A further study was added during data extraction through reference searches, and 5 were added upon rerunning of searches in December 2021. As a result, 26 studies were included in the final analysis (Figure 1).

General characteristics of included studies. We included 26 eligible studies reporting 31 final multivariable prediction models for incident OA, published between 2010 and 2022, using study populations from 15 unique data sources in the US (9 studies), The Netherlands (8 studies), UK (4 studies), Sweden (2 studies), Canada, China, and Norway (Table 1). The median prediction horizon was 8 years (range 2–41 years), the median number of participants/joints with the outcome of interest was 121.5 (range 27–12,803), and the median number of predictors included in final models was 6 (range 3–24). Regression analysis was used in 24 models (commonly logistic regression or generalized estimating equations), while 7 involved machine learning approaches (e.g., [deep] neural networks). Internal validation was undertaken for 19 models, 7 were externally validated, and 2 were both internally and externally validated.

Knee OA. *Incident radiographic knee OA.* Of 23 models predicting incident knee OA, 13 defined the outcome radiographically by plain radiography, typically a Kellgren/Lawrence (K/L) grade of 2 or more, although 2 models selected the more severe threshold of K/L grade of ≥3 (23,24). The median number of participants/joints with the outcome of interest for these 13 models was 95 (range 27–474). The median AUC following internal and external validation was 0.77 (range 0.69–0.82, 6 models) and 0.76 (range 0.60–0.86, 4 models, 6 populations), respectively (Figure 2). All 13 models included predictors obtainable from clinical assessment. Most common predictors, featuring in >4 final models, were age, sex, body mass index (BMI), and previous knee injury, as well as self-reported pain, stiffness, and function scores from the Western Ontario and McMaster Universities Osteoarthritis Index. Eight models solely used predictors available from clinical assessment (25–30), a further 6 models included predictors sourced from plain radiographs at baseline (23,24,31–34),

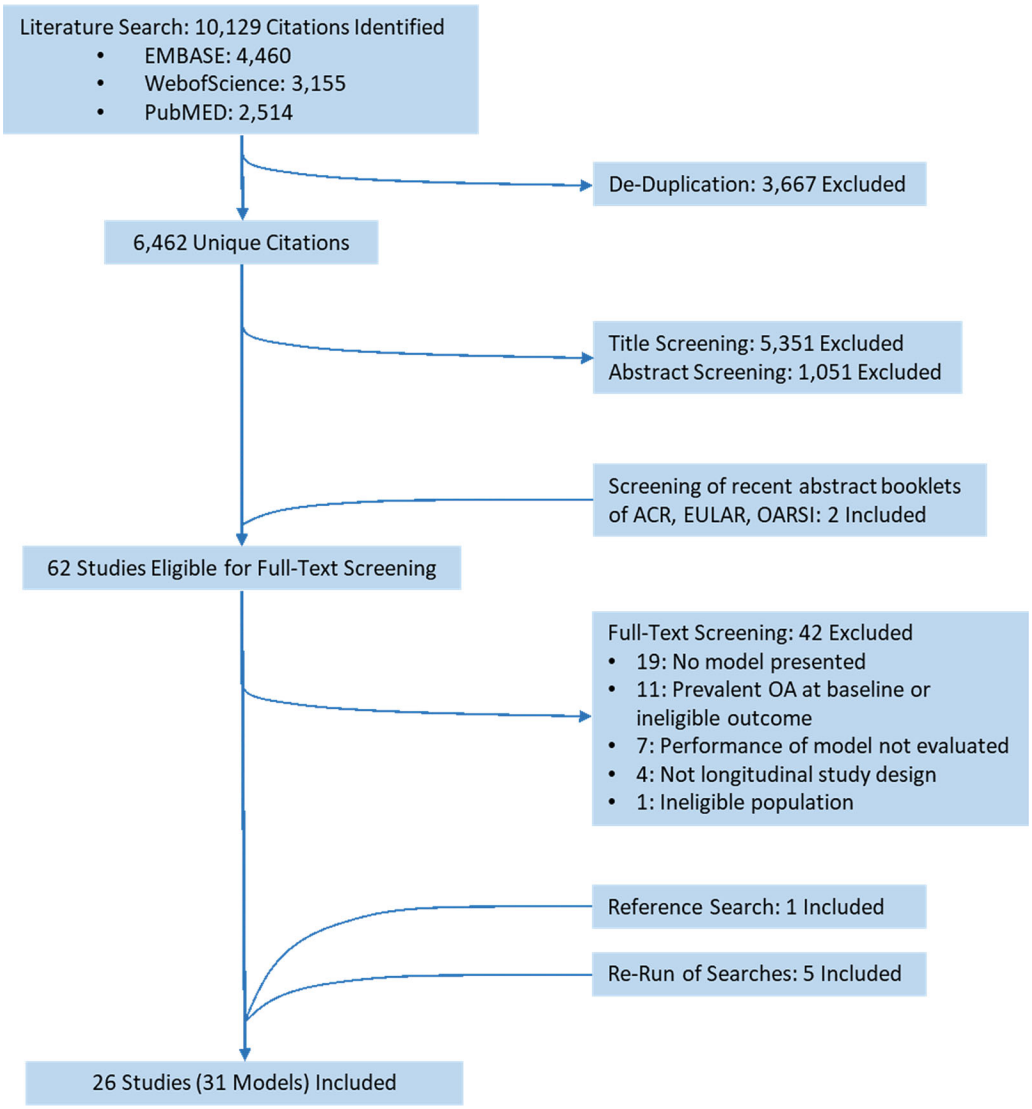

Figure 1. Flow chart of article screening and inclusion. ACR = American College of Rheumatology; EULAR = European Alliance of Associations for Rheumatology; OA = osteoarthritis; OARSI = Osteoarthritis Research Society International.

3 used magnetic resonance imaging (MRI) (34–36), and 5 incorporated serum or urinary biomarkers (included predictors for all models are shown in Figure 3) (24,35,37–39).

Incident symptomatic OA (frequent knee pain). A symptomatic knee OA outcome was used in 5 models, most commonly defined as the onset of frequent knee pain. The median number of participants/joints with the outcome of interest for these models ranged from 51 to 2,103. The median AUC following internal validation was 0.71 (range 0.70–0.78, 5 models). None of these models used predictors beyond those obtainable from clinical assessment or plain radiography. Fernandes et al (27) and Landsmeer et al (28) externally validated their symptomatic models but with contrasting results. External validation in the Osteoarthritis Initiative (OAI) of the model derived from Nottingham data resulted in poor discrimination and calibration (AUC 0.54 [95% CI 0.50–0.58]) (27). Landsmeer et al found relatively little reduction in their model discrimination upon external validation in the Rotterdam Study-III cohort (AUC 0.71 [95% CI 0.62, 0.79]) (28) (Figure 2).

Symptomatic-radiographic knee OA. A definition of knee OA combining both radiographic and symptomatic criteria was used in 4 models. Models by Zhang et al (26) and Landsmeer et al (28) were derived solely using clinical assessment predictors; Chan et al (40) included radiographic predictors, with Lazzarini et al (35) incorporating radiographic, MRI, and biomarker predictors. All models were internally or externally validated. Lazzarini et al (35) and Chan et al (40) internally validated by cross-validation and bootstrapping, respectively. Again, the model derived within the Nottingham cohort did not perform well in the OAI (AUC 0.60 [95% CI 0.58–0.63]), although discrimination remained high in the Genetics of OA and Lifestyle case–control study (AUC 0.79 [95% CI 0.77–0.81]) (26). Landsmeer et al (28)

Table 1. Descriptive characteristics of included studies*

General information				Information relating to final models						
Author, year (ref.)	Cohort data sets (country)	Modeling technique	Relevant models reported	Outcome definition	Source of predictors	Max. predict. horizon, years†	Predictors, no.	Participants (with outcomes), no.‡	Int. Val.	Ext. Val.
Knee OA										
Peat et al, 2010 (25)	CAS-K§ (UK)	Regression	2	ROA	SR, PE	3	10	122 (42)	N	N
Zhang et al, 2011 (26)	Nott.§, OAI, GOAL (UK, US)	Regression	2	1) ROA, 2) SROA	1) SR, 2) SR	1) 12, 2) 12	1) 6, 2) 6	1) 179 (99), 2) 179 (99)	1) N, 2) N	1) Y, 2) Y
Kinds et al, 2012 (31)	CHECK§ (NL)	Regression	12	ROA	SR, PE, XR	5	4	985j (189j)	Y	N
Kerkhof et al, 2014 (32)	RS-I§, RS-II, Ching. (NL, UK)	Regression	6	ROA	SR, PE, XR	9.4	5	2,628 (474)	N	Y
Riddle et al, 2016 (23)	MOST†, OAI (US)	Regression	2	ROA	SR, PE, XR	5	4	2,824j (262j)	N	Y
Fernandes et al, 2017 (27)	Nott.§, OAI (UK, US)	Regression	1	FKP	SR	12	7	1,203 (?)	Y	Y
Janvier et al, 2017 (33)	OAI§ (US)	Regression	18	ROA	SR, PE, XR	2	4	344j (79j)	Y	N
Lazzarini et al, 2017 (35)	PROOF§ (NL)	M-L	3	1) SROA, 2) FKP, 3) ROA	1) SR, PE, XR, MRI, SB; 2) SR, PE, XR; 3) SR, PE, MRI, SB	1) 2.5, 2) 2.5, 3) 2.5	1) 8, 2) 7, 3) 5	1) 354 (39), 2) 351 (51), 3) 321 (27)	1) Y, 2) Y, 3) Y	1) N, 2) N, 3) N
Sharma et al, 2017 (34)	OAI§ (US)	Regression	25	ROA	SR, PE, XR, MRI	8	12	841j (53j)	N	N
Driban et al, 2018 (24)	OAI§ (US)	Regression	1	ROA	SR, PE, XR, SB	4	4	162 (54)	Y	N
Camacho-Encina et al, 2019 (37)	OAI§, OAI (US)	Regression	2	ROA	SR, PE, SB	8	7	327 (146)	N	Y
Landsmeer et al, 2019 (28)	PROOF§, RS-III (NL)	Regression	2	1) FKP, 2) SROA	1) SR, PE; 2) SR,PE	1) 6.5, 2) 6.5	1) 7, 2) 7	1) 237 (75), 2) 235 (70)	1) N, 2) N	1) Y, 2) Y
Magnusson et al, 2019 (29)	SWE Con.§ (SWE)	Regression	2	RDOA	SR, PE, EHR	41	3	40,118 (2,052)	Y	N
Garriga et al, 2020 (38)	Ching.§ (UK)	Regression	2	ROA	SR, PE, UB	4	3	649 (95)	Y	N
Chan et al, 2021 (40)	OAI§ (US)	M-L	5	SROA	SR, PE, XR	2	24	2,640 (1,254)	Y	N
Joseph et al, 2022 (36)	OAI§ (US)	M-L	3	ROA	SR, PE, MRI	8	10	710 (124)	Y	N
Lourido et al, 2021 (39)	OAI§ (US)	Regression	5	ROA	SR, PE, SB	8	6	200 (86)	Y	N
Wang et al, 2021 (30)	CHARLS§ (China)	Regression	1	FKP	SR, PE	4	10	8,193 (815)	Y	N
Guan et al, 2022 (48)	OAI§ (US)	M-L	4	FKP	SR, PE, XR	2	6	4200j (2103j)	Y	N
Hip OA										
Castano-Betancourt et al, 2013 (41)	RS-I§ (NL)	Regression	14	ROA/THR	SR, PE, XR	11	9	688 (119)	N	N
Saberi Hosnijeh et al, 2018 (42)	RS-I§, RS-II, CHECK (NL)	Regression	5	ROA/THR	SR, PE, XR, SB	10	13	4,575j (258j)	Y	Y
Hirvasniemi et al, 2019 (43)	CHECK§ (NL)	M-L	10	ROA/THR	SR, PE, XR	10	5	987j (43j)	Y	N
Gielis et al, 2020 (44)	CHECK§ (NL)	Regression	5	ROA/THR	SR, PE, XR	8	4	1,002 (185)	Y	N
Hand OA										
Magnusson et al, 2018 (45)	SWE Con.§ (SWE)	Regression	1	RDOA	SR, PE	41	6	40,118 (212)	Y	N
Johnsen et al, 2020 (46)	HUNT2§ (NOR)	Regression	5	RDOA	1) SR, 2) PE	1) 23, 2) 23	1) 5, 2) 5	1) M: 17,153 (206), 2) F: 18,682 (732)	1) Y, 2) Y	1) N, 2) N
OA (any joint)										
Black et al, 2020 (47)	CPCSSN§ (CAN)	Regression	1	RDOA	EHR	5	5	383,117 (12,803)	Y	N

* CAN = Canada; CAS-K = Clinical Assessment Study – Knee; CHARLS = China Health and Retirement Longitudinal Study; CHECK = Cohort Hip and Cohort Knee study; Ching. = Chingford Study; CPCSSN = Canadian Primary Care Sentinel Surveillance Network; EHR = electronic health record; Ext. = external; F = female; FKP = frequent knee pain; GOAL = Genetics of Osteoarthritis and Lifestyle study; HUNT2 = Helseundersøkelsen i Nord-Trøndelag study; Int. = internal; M = male; Max. = maximum; M-L = machine learning; MOST = Multicenter Osteoarthritis Study; MRI = magnetic resonance imaging; NL = Netherlands; NOR = Norway; Nott. = Nottingham Cohort; OA = osteoarthritis; OAI = Osteoarthritis Initiative; PE = physical examination; Predict. = prediction; PROOF = PRevention of knee Osteoarthritis in Overweight Females Study; RDOA = recorded diagnosis of OA; ref. = reference; ROA = radiographic OA; RS = Rotterdam Study; SB = serum biomarker; SR = self-report; SROA = symptomatic-radiographic OA; SWE = Sweden; SWE Con. = Swedish Conscription Cohort; THR = total hip replacement; UB = urine biomarker; Val. = validation; XR = X-ray (plain radiography).
† Certain studies had follow-up at interim intervals.
‡ j = joint level of interest, i.e., 985j (189j) for Kinds et al (31) represents 985 knees included, of which 189 knees developed the outcome.
§ Denotes derivation data set; other samples are those in which external validation was undertaken.

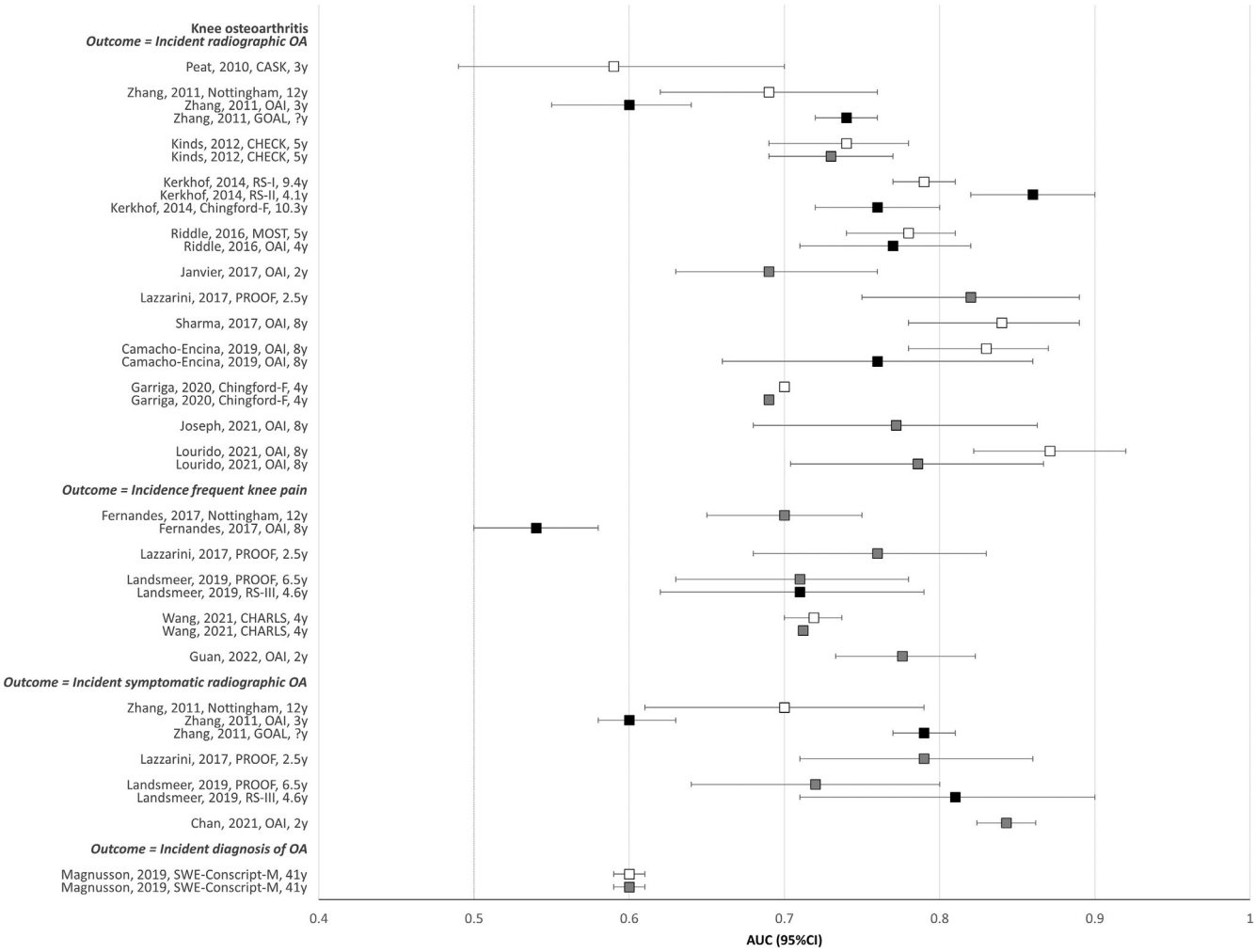

Figure 2. Forest plot of the area under the curve (AUC) with 95% confidence intervals (95% CIs) or multivariable prediction models for incident knee osteoarthritis (OA), stratified by outcome, ordered by year of publication. The first column lists author, year of publication, data set, and prediction horizon. The white marker shows model development, the grey marker shows internal validation, and the black marker shows external validation. CAS-K = Clinical Assessment Study–Knee; CHARLS = China Health and Retirement Longitudinal Study; CHECK = Cohort Hip and Cohort Knee study; Chingford-F = Chingford female; GOAL = Genetics of Osteoarthritis and Lifestyle study; MOST = Multicenter Osteoarthritis Study; OAI = Osteoarthritis Initiative; PROOF = PRevention of knee Osteoarthritis in Overweight Females Study; RS-I = Rotterdam Study 1; RS-II = Rotterdam Study 2; RS-III = Rotterdam Study 3; SWE-Conscript-M = Swedish Military Conscription Dataset.

also validated their combined-criteria model; the AUC in the Rotterdam Study-III was 0.81 (95% CI 0.71–0.90) (Figure 2).

Recorded diagnosis of knee OA. A 2019 study by Magnusson et al was the only identified study predicting knee OA using electronic health record (EHR) data (29). They identified male patients undergoing military conscription in Sweden in 1969 (n = 40,118), and using the National Patient Register, determined who subsequently developed knee OA (n = 2,052). Knee OA was defined as the entry of a relevant International Classification of Diseases, Ninth Revision (ICD-9) or ICD-10 code within the patient's EHR between 1987 and 2010. The AUC of the model following internal validation was 0.60 (95% CI 0.59–0.61) (Figure 2).

Hip OA. Of 4 models derived to predict hip OA (41–44), all originated in The Netherlands and used the same definition: a composite outcome of K/L grade of ≥2 or total hip replacement (THR). The median number of participants/joints was 994.5. The 2 earliest models were derived in the Rotterdam Study-I cohort (41,42), the latter 2 used the Cohort Hip and Cohort Knee (CHECK) study (43,44). All models featured age, sex, and BMI, together with radiographic parameters. A baseline K/L grade (0/1) was used in 3 models (41–43); the remainder used the presence of joint space narrowing and osteophytes (upon which the K/L grade is calculated) (44). The most recently published models incorporated trabecular bone texture (43) and patented automated hip shape via a machine learning algorithm (44). Discrimination of the latter model in particular was high (the AUC from internal validation was 0.86 [95% CI 0.83–0.90]) with near-perfect calibration (44). External validation was undertaken only by Saberi Hosnijeh et al, finding a reduction in

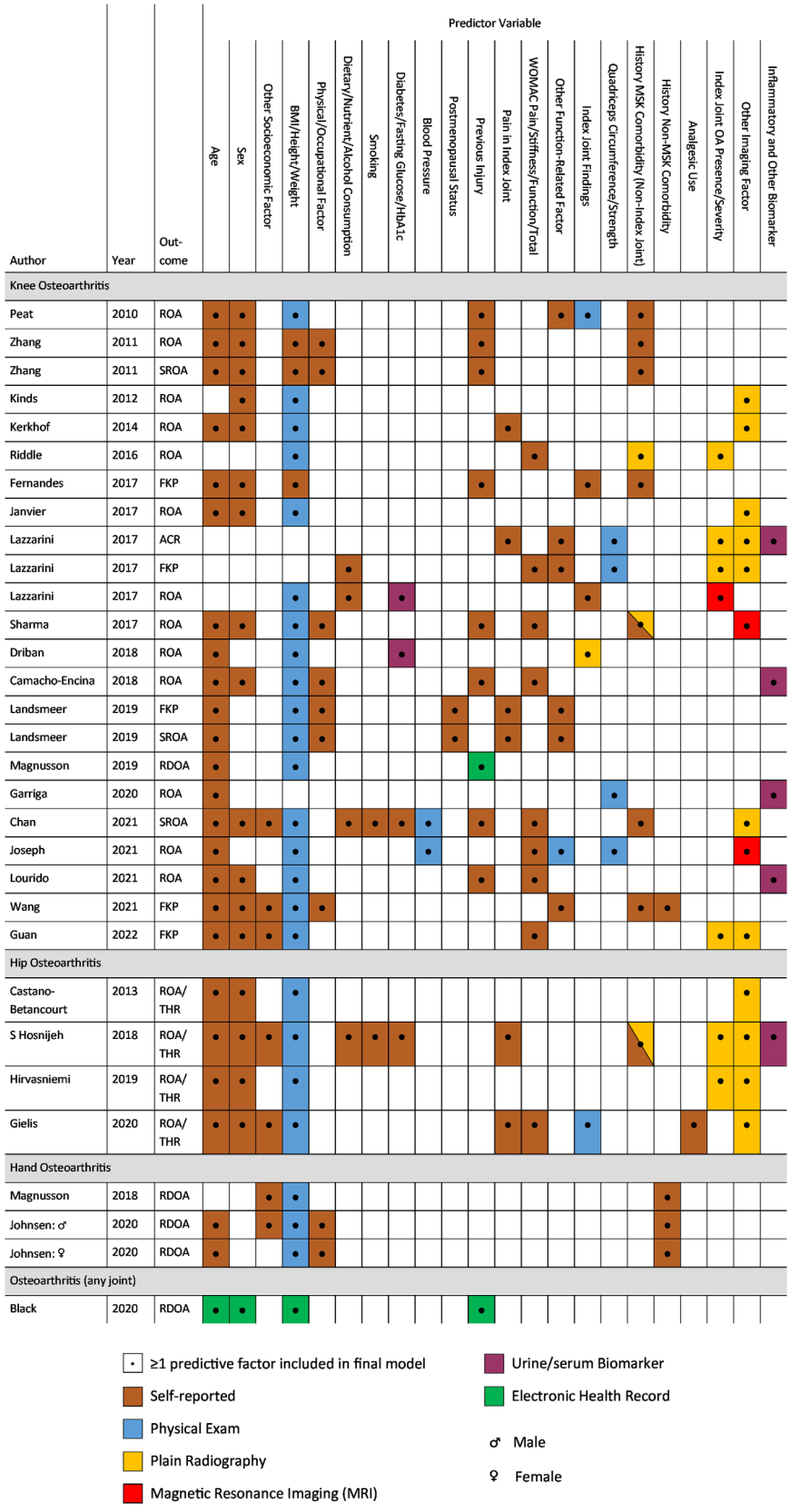

Figure 3. Predictor variables and domains represented in final models, by mode of assessment. BMI = body mass index; FKP = frequent knee pain; $HbA1_c$ = glycosylated hemoglobin; MSK = musculoskeletal; OA = osteoarthritis; RDOA = recorded diagnosis of OA; ROA = radiographic OA; SROA = symptomatic radiographic OA; THR = total hip replacement; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.

performance in both the Rotterdam Study-II (AUC 0.75 [95% CI 0.72–0.79]) and more notably in the CHECK cohort data set (AUC 0.71 [95% CI 0.66–0.75]) (42) (Figure 4).

Hand OA. We identified 2 Scandinavian studies (deriving 3 models) predicting the incidence of hand OA (45,46). Magnusson et al (45) used the aforementioned military conscription cohort and linked to the National Patient Register to identify participants who developed hand OA (using ICD-10 codes) between 1998 and 2010. A small proportion (n = 212, 0.5%) of their study population (n = 40,118) developed the outcome. Their final model included education level, BMI, and sleep problems; the AUC following internal validation was 0.62 (95% CI 0.58–0.64) (45). Building upon this, Johnsen et al (46) used the Norwegian Nord-Trøndelag Health Study cohort to externally validate the original model of Magnusson et al, as well as a simplified algorithm (46). Model performance was comparable, with AUCs of 0.60 (95% CI 0.56–0.64) and 0.62 (95% CI 0.57–0.65), respectively (Figure 4). Johnsen et al (46) then developed their own prediction models for hand OA in male subjects and female subjects, separately, using diagnostic ICD-9 and ICD-10 codes within the Norwegian National Patient Register. Of note, no improvement in performance was observed with the addition of a genetic risk score in models for either patient sex, or with reproductive and hormonal factors in female participants (46). While both studies underwent internal validation, neither was externally validated.

OA (any joint). Black et al, using the Canadian Primary Sentinel Surveillance Network, was the only study that sought to predict the incidence of OA irrespective of joint (47). They identified 383,117 eligible patients, of whom 12,803 received a billing or problem-list code for OA within 5 years of cohort entry. Their model consisted of 5 predictors routinely collected within EHR data: age, sex, BMI, prior leg injury, and osteoporosis diagnosis. Both discrimination (AUC 0.84 [95% CI 0.83–0.85]) and calibration were good following 10-fold cross-validation (Figure 4).

Calibration summary. Of 12 models internally validating their model derived for knee OA outcomes, 5 included calibration

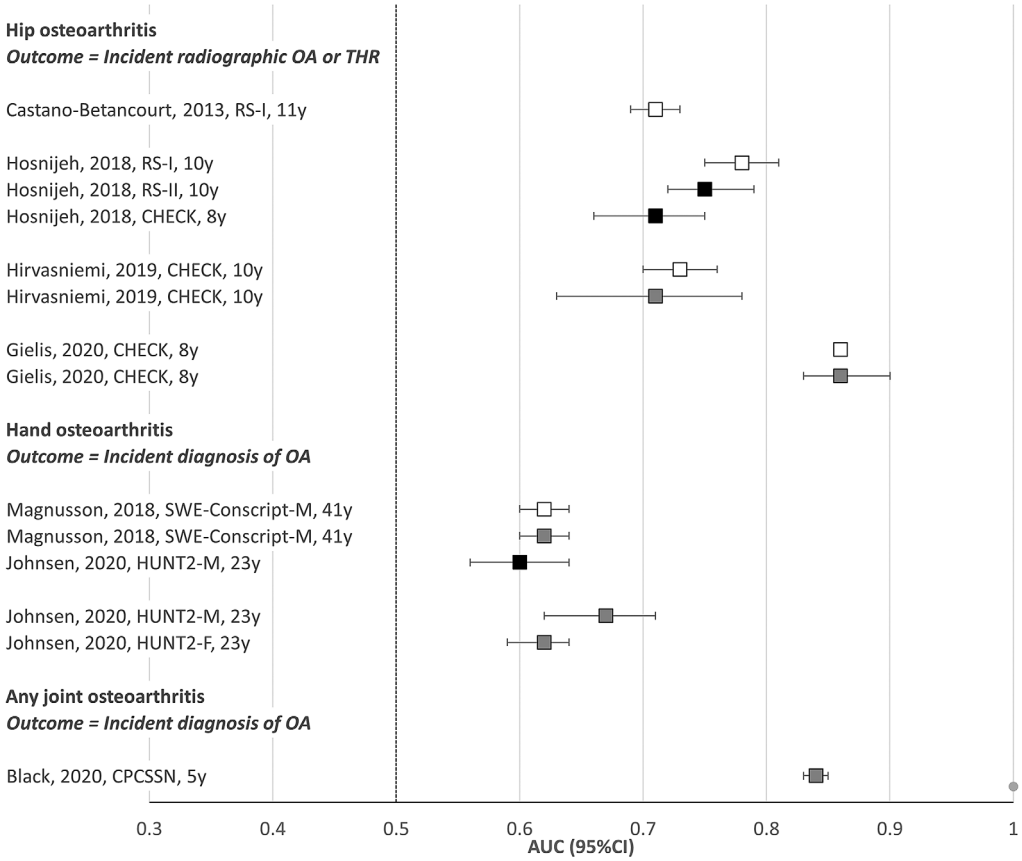

Figure 4. Forest plot of the area under the curve (AUC) with 95% confidence intervals (95% CIs) for multivariable prediction models for incident hip, hand, and any joint osteoarthritis (OA), ordered by year of publication. The first column lists author, year of publication, data set, and prediction horizon. The white marker shows model development, the grey marker shows internal validation, and the black marker shows external validation. CHECK = Cohort Hip and Cohort Knee study; CPCSSN = Canadian Primary Care Sentinel Surveillance Network; HUNT2-F = Helseundersøkelsen i Nord-Trøndelag study–female; HUNT2-M = Helseundersøkelsen i Nord-Trøndelag study–male; RS-I = Rotterdam Study 1; RS-II = Rotterdam Study 2; SWE-Conscript-M = Swedish Military Conscription Dataset; THR = total hip replacement.

assessment. Earlier models (27,31) appraised calibration using the Hosmer-Lemeshow statistic, and 4 studies presented and appraised calibration plots (27,29,30,38). Five of the 6 models for knee OA undergoing external validation assessed calibration using Hosmer-Lemeshow; only Fernandes et al (27) presented findings visually.

All models for hip (except the earliest [41]), hand, and any-joint OA were internally validated. All but 1 model presented calibration plots visually, reporting reasonable or good agreement between expected and observed outcomes. Models by Gielis et al (44) and Black et al (47) demonstrated excellent calibration.

Risk-of-bias summary. The most common sources of potential bias included the extensive use of univariable analysis to select predictors for inclusion in final models, and suboptimal handling of missing data, competing risks, and cohort attrition (domain 4). The inclusion of THR (a measure of both incidence and progression) in composite outcome definitions of incident hip OA was also flagged (Figure 5).

We judged model applicability in terms of the ability to be implemented at scale in diverse general (adult) populations. Within those terms, applicability was typically judged to be poor. Poor applicability was most often due to the need for predictors obtained by imaging or biologic samples that may not be routinely available, recommended, or affordable for such application (domain 2). Ethnicity and other social stratifiers were seldom considered or included in final models, and we were often unsure whether data sets used to derive and validate models had drawn from a sufficiently diverse population to be applied at scale in general populations (domain 1).

DISCUSSION

We sought to systematically identify and critically evaluate existing multivariable risk prediction models for OA incidence and to consider their potential application at scale in diverse populations to advance individual risk-informed preventive action. Our review identified 26 studies deriving 31 multivariable risk prediction models. A total of 16 models published since 2018 suggests a growing field, attracting machine learning approaches and novel biomarkers, but one that remains centered around a relatively small number of mature cohort data sets of knee (and to a lesser extent hip) OA incidence in high-income countries.

Importantly, our review identifies a general lack of inclusion of social stratifiers beyond age, sex, and occupation-associated risk. Of note, until a study in 2021 by Chan et al (and subsequently Guan et al in 2022) (40,48), no stratifiers relating to ethnicity and markers of deprivation existed, factors that are associated with disparities in both incidence and prevalence of OA (49,50). The lack of such predictors, as well as income, education, and geographic location, may also contribute to a lack of applicability, and usability in, wider populations.

With some exceptions, notably when predicting a future recorded diagnosis of OA across very long prediction horizons, model discrimination after validation ranged from an AUC of 0.70 to 0.85. This range of performance relates to heterogeneous models with prediction horizons from 2 to 12 years and predictors whose collection and processing varies in cost and complexity. In several models undergoing internal or external validation, calibration was either not reported or relied on the Hosmer-Lemeshow test statistic, which is recognized as problematic and no longer recommended (51). Better approaches, however, including the visual display of calibration plots (29,38,42–47) and reported intercept and slope (38,43,47) were used in several more recent studies. Poor calibration of some models was attributed to the inherent unpredictability of incident OA over very long prediction horizons (45), and also to challenges in identifying suitably comparable cohort data sets for external validation (27,42). We identified no recent examples of externally validated models supported by moderate or strong evidence of good calibration. Acknowledging that the development of a single prediction model for OA may be challenging, particularly across different target populations, within different health care systems, or across long prediction horizons, from adolescence to disease onset, is important. Predicting individual risk and preventive intervention remain achievable but may require several models in different settings and contexts.

Amid our critical appraisals of risk of bias and areas for methodologic improvement were several positives: the use of internal or external validation was common, greatly facilitated by data sharing, the foresight to design overlapping data points across different cohort studies, the shift toward more careful evaluation of model calibration, and a common practice of including certain core predictors (age, sex, BMI, previous injury). We would also encourage others to emulate, where possible, the initial approach of Johnsen et al to testing and adapting a previously published model rather than assuming the need to derive another new model (46).

We found little evidence of patient and public involvement and engagement in included studies. This may contribute to a lack of clarity on potential applications and routes to implementation, and studies may be strengthened with clear rationale statements alongside integration of patient and public involvement and engagement, for instance by following the Guidance for Reporting Involvement of Patients and the Public (52). Our own review can be criticized on this point, limited by being an unfunded project, with no means for required remuneration. Lack of patient and public involvement is an area for future strengthening.

Our prospectively registered review used a replicable search strategy without language restriction in the search phase across 3 electronic databases, which was rerun prior to submission and was supplemented by searches of reference lists and conference abstracts. Pairs of reviewers working independently and using recommended checklists and risk-of-bias tools performed study selection, data extraction, and risk-of-bias assessments. Our

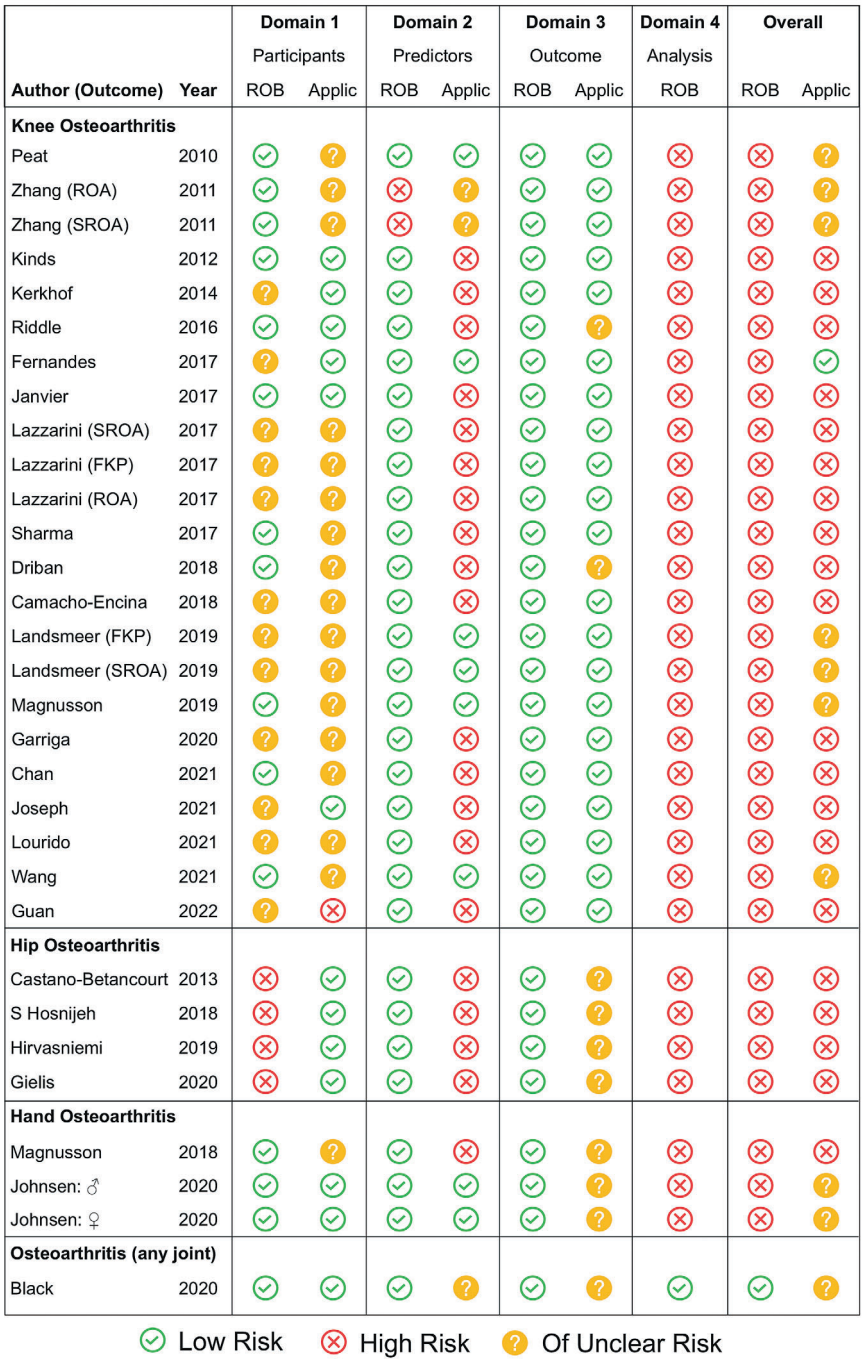

Figure 5. Risk-of-bias assessment using the Prediction study Risk of Bias Assessment Tool. ROB = risk of bias; Applic = applicability; FKP = frequent knee pain; ROA = radiographic osteoarthritis; SROA = symptomatic-radiographic osteoarthritis; ♂ = male; ♀ = female.

review has several limitations. First, risk-of-bias tools specifically for prediction modeling studies using machine learning techniques were under development at the time of our review (53). Second, several studies derived multiple models for the same outcome. The designation of a final model relied on 2 reviewers' independent judgement based first on the authors' description but was not necessarily the best performing or most applicable model reported. Third, while knowing all of the candidate predictors considered in model development would be of interest, this information was often lacking or partially reported. We opted not to try to synthesize this information. We also refrained from attempting to calculate events per variable for each model because of lack of available information on candidate predictors and because EPV is no longer recommended as a guide to sample size (54,55). Fourth, we did not undertake meta-analysis due to study heterogeneity, nor meta-regression due to an insufficient

number of models. Finally, models were developed by their authors for many reasons. Models judged by us to have low applicability to large-scale use in diverse general populations may be highly applicable for other purposes, such as enriched recruitment to clinical trials or within selected clinical settings.

We are unaware of any other previously published review of multivariable risk prediction models for OA incidence with which to compare our findings, although we note a published protocol of a review in development relating to prognostic models for knee OA (56). Reviews in other fields have found similar concerns over methodologic quality and applicability of multivariable prediction models for disease incidence (57–60). An excess of model development and a lack of rigorous external validation by independent research teams is a recurrent theme. In the more established field of cardiovascular risk prediction, use of registry and EHR data sets appears more common, and efforts are underway to adapt models for application in low- and middle-income countries (4). Such attempts may signal directions for future development of individual risk prediction in OA. Challenges on the validity and completeness of coding and the availability of information on important predictors within routine EHR data are well recognized. The approach of Black et al (47), however, suggests that mitigating some of these may be possible, to produce prediction models with good performance and a prospect of implementation within existing national health systems. However, whether routine EHR data can support accurate risk prediction models is unclear, specifically for hip OA and hand OA. Aspects of hip morphology appear to add important predictive value but will have limited availability in routine records for general populations. Furthermore, the consistent use of composite outcomes including THR may limit both applicability and accuracy in predicting incident disease, an implication for hip OA model development that may be highlighted following more extensive external validation. Substantial under- and misdiagnosis of hand OA poses a different challenge.

Johnsen et al (46), in their separate prediction models of incident hand OA in both males and females, did not find a significant improvement in model performance with the addition of a genetic risk score. Genetic association within OA is a growing field, with ongoing identification of associated variants (61). While the predictive value of these novel variants remains unknown, we believe a model feasible for widespread implementation in clinical practice should use routinely available predictors. Furthermore, previous literature suggests that the accurate prediction of outcomes requires associations of the strength rarely observed in studies (62). Consequently, the addition of predictors such as genetic risk scores to core predictors such as age may not significantly change model performance, and may explain the apparent null result in the models of Johnsen et al (46).

Our review excluded several studies that we feel deserve specific mention. We excluded studies that relied solely on joint replacement as the outcome because of the risk of conflating predictors of incidence and progression. However, the separation of incidence from progression in OA can be contested. Approaches to modeling changes in symptom and disease severity, classifying cohort enrollment, or censorship as a spectrum rather than binary events, such as by Halilaj et al (63) and Widera et al (64), may still contain relevant information. In addition, the linked studies of Losina et al (65) and Michl et al (66) provide evidence that is highly relevant to introducing individual risk models for OA in 1 scalable format: patient self-evaluation using an online OA risk calculator. Of note, this calculator used relatively simple-to-report predictors based on the earlier Nottingham risk prediction model, derived by Zhang et al (26). Beyond these studies, there is a lack of evaluation of the impact of risk prediction models for OA used in clinical practice.

In summary, we identified 31 multivariable prediction models for OA from 26 published studies. While interest is growing among researchers, applicability to clinical practice in diverse general populations is often lacking, and only a paucity of evaluation exists of the impact of implementation. We suggest that models may benefit from clearly stating their rationale, and by integrating patient and public involvement and engagement and predictors such as ethnicity. Furthermore, the progression toward viable risk prediction models for OA that are applicable across a number of settings would be aided with a focus on routinely available predictors and with wider external validation of models in varied populations. Last, growing interest in machine learning techniques, as well as the classification of OA disease as a progression rather than dichotomous incidence, warrants updated research guidelines to better appraise such innovative approaches.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Appleyard had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Appleyard, Peat.
Acquisition of data. Appleyard, Thomas, Antcliff, Peat.
Analysis and interpretation of data. Appleyard, Thomas, Antcliff, Peat.

REFERENCES

1. Chen L. Overview of clinical prediction models. Ann Transl Med 2020; 8:71.
2. Vos T, Allen C, Arora M, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388:1545–602.
3. United Nations. Transforming our world: the 2030 agenda for sustainable development. 2015. URL: https://sdgs.un.org/2030agenda.

4. Kaptoge S, Pennells L, De Bacquer D, et al. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. Lancet Glob Health 2019;7:e1332–45.
5. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007;335:136.
6. Sahle BW, Owen AJ, Chin KL, et al. Risk prediction models for incident heart failure: a systematic review of methodology and model performance. J Card Fail 2017;23:680–7.
7. Pabinger C, Lothaller H, Geissler A. Utilization rates of knee-arthroplasty in OECD countries. Osteoarthritis Cartilage 2015;23:1664–73.
8. Xie F, Kovic B, Jin X, et al. Economic and humanistic burden of osteoarthritis: a systematic review of large sample studies. Pharmacoeconomics 2016;34:1087–100.
9. Puig-Junoy J, Ruiz Zamora A. Socio-economic costs of osteoarthritis: a systematic review of cost-of-illness studies. Semin Arthritis Rheum 2015;44:531–41.
10. Culliford D, Maskell J, Judge A, et al. Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink. Osteoarthritis Cartilage 2015;23:594–600.
11. Ackerman IN, Bohensky MA, Zomer E, et al. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. BMC Musculoskelet Disord 2019;20:90.
12. Emery CA, Whittaker JL, Mahmoudian A, et al. Establishing outcome measures in early knee osteoarthritis. Nat Rev Rheumatol 2019;15:438–48.
13. Croft P, Altman DG, Deeks JJ, et al. The science of clinical practice: disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice. BMC Med 2015;13:20.
14. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;350:g7647.
15. Geersing GJ, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One 2012;7:e32844.
16. Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan: a web and mobile app for systematic reviews. Syst Rev 2016;5:210.
17. Microsoft Corporation. Microsoft Excel. 2021.
18. Cochrane Methods Prognosis. Tools. 2021. URL: https://methods.cochrane.org/prognosis/tools.
19. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11:e1001744.
20. Wolff RF, Moons KG, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51–8.
21. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128–38.
22. Hosmer D, Lemeshow S. Applied logistic regression. 2nd ed. Hoboken (NJ): Wiley; 2000. p. 143–202.
23. Riddle DL, Stratford PW, Perera RA. The incident tibiofemoral osteoarthritis with rapid progression phenotype: development and validation of a prognostic prediction rule. Osteoarthritis Cartilage 2016;24:2100–7.
24. Driban JB, McAlindon TE, Amin M, et al. Risk factors can classify individuals who develop accelerated knee osteoarthritis: data from the Osteoarthritis Initiative. J Orthop Res 2018;36:876–80.
25. Peat G, Thomas E, Duncan R, et al. Is a 'false-positive' clinical diagnosis of knee osteoarthritis just the early diagnosis of pre-radiographic disease? Arthritis Care Research (Hoboken) 2010;62:1502–6.
26. Zhang W, McWilliams DF, Ingham SL, et al. Nottingham knee osteoarthritis risk prediction models. Ann Rheum Dis 2011;70:1599–604.
27. Fernandes GS, Bhattacharya A, McWilliams DF, et al. Risk prediction model for knee pain in the Nottingham community: a Bayesian modelling approach. Arthritis Res Ther 2017;19:59.
28. Landsmeer ML, Runhaar J, van Middlekoop M, et al. Predicting knee pain and knee osteoarthritis among overweight women. J Am Board Fam Med 2019;32:575–84.
29. Magnusson K, Turkiewicz A, Timpka S, et al. A prediction model for the 40-year risk of knee osteoarthritis in adolescent men. Arthritis Care Res (Hoboken) 2019;71:558–62.
30. Wang L, Lu H, Chen H, et al. Development of a model for predicting the 4-year risk of symptomatic knee osteoarthritis in China: a longitudinal cohort study. Arthritis Res Ther 2021;23:65.
31. Kinds MB, Marijnissen AC, Vincken KL, et al. Evaluation of separate quantitative radiographic features adds to the prediction of incident radiographic osteoarthritis in individuals with recent onset of knee pain: 5-year follow-up in the CHECK cohort. Osteoarthritis Cartilage 2012;20:548–56.
32. Kerkhof HJ, Bierma-Zeinstra SM, Arden NK, et al. Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors. Ann Rheum Dis 2014;73:2116–21.
33. Janvier T, Jennane R, Toumi H, et al. Subchondral tibial bone texture predicts the incidence of radiographic knee osteoarthritis: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2017;25:2047–54.
34. Sharma L, Hochberg M, Nevitt M, et al. Knee tissue lesions and prediction of incident knee osteoarthritis over 7 years in a cohort of persons at higher risk. Osteoarthritis Cartilage 2017;25:1068–75.
35. Lazzarini N, Runhaar J, Bay-Jensen AC, et al. A machine learning approach for the identification of new biomarkers for knee osteoarthritis development in overweight and obese women. Osteoarthritis Cartilage 2017;25:2014–21.
36. Joseph GB, McCulloch CE, Nevitt MC, et al. Machine learning to predict incident radiographic knee osteoarthritis over 8 years using combined MR imaging features, demographics, and clinical factors: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2022;30:270–9.
37. Camacho-Encina M, Balboa-Barreiro V, Rego-Perez I, et al. Discovery of an autoantibody signature for the early diagnosis of knee osteoarthritis: data from the Osteoarthritis Initiative. Ann Rheum Dis 2019;78:1699–705.
38. Garriga C, Sanchez-Santos MT, Judge A, et al. Predicting incident radiographic knee osteoarthritis in middle-aged women within four years: the importance of knee-level prognostic factors. Arthritis Care Res (Hoboken) 2020;72:88–97.
39. Lourido L, Balboa-Barreiro V, Ruiz-Romero C, et al. A clinical model including protein biomarkers predicts radiographic knee osteoarthritis: a prospective study using data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2021;29:1147–54.
40. Chan LC, Li HH, Chan PK, et al. A machine learning-based approach to decipher multi-etiology of knee osteoarthritis onset and deterioration. Osteoarthritis Cartilage Open 2021;3:100135.
41. Castano-Betancourt M, Van Meurs JB, Bierma-Zeinstra S, et al. The contribution of hip geometry to the prediction of hip osteoarthritis. Osteoarthritis Cartilage 2013;21:1530–6.
42. Saberi Hosnijeh F, Kavousi M, Boer CG, et al. Development of a prediction model for future risk of radiographic hip osteoarthritis. Osteoarthritis Cartilage 2018;26:540–6.
43. Hirvasniemi J, Gielis WP, Arbabi S, et al. Bone texture analysis for prediction of incident radiographic hip osteoarthritis using machine learning: data from the Cohort Hip and Cohort Knee (CHECK) study. Osteoarthritis Cartilage 2019;27:906–14.

44. Gielis W, Weinans H, Welsing PM, et al. An automated workflow based on hip shape improves personalized risk prediction for hip osteoarthritis in the CHECK study. Osteoarthritis Cartilage 2020;28: 62–70.
45. Magnusson K, Turkiewicz A, Timpka S, et al. Prediction of midlife hand osteoarthritis in young men. Osteoarthritis Cartilage 2018;26: 1027–32.
46. Johnsen MB, Magnusson K, Borte S, et al. Development and validation of a prediction model for incident hand osteoarthritis in the HUNT study. Osteoarthritis Cartilage 2020;28:932–40.
47. Black J, Terry A, Lizotte D. Development and evaluation of an osteoarthritis risk model for integration into primary care health information technology. Int J Med Inform 2020;141:104160.
48. Guan B, Lui F, Mizaian AH, et al. Deep learning approach to predict pain progression in knee osteoarthritis. Skeletal Radiol 2022;51: 363–73.
49. Callahan LF, Cleveland RJ, Allen KD, et al. Racial/ethnic, socioeconomic, and geographic disparities in the epidemiology of knee and hip osteoarthritis. Rheum Dis Clin North Am 2021;47:1–20.
50. 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;23:1323–9.
51. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:230.
52. Staniszewska S, Brett J, Simera I, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. Res Involv Engagem 2017;3:13.
53. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008.
54. Van Smeden M, de Groot JA, Moons KG, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Med Res Methodol 2016;16:163.
55. Riley RD, Ensor J, Snell KI, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441.
56. Zhong J, Si L, Zhang G, et al. Prognostic models for knee osteoarthritis: a protocol for systematic review, critical appraisal, and meta-analysis. Syst Rev 2021;10:149.
57. Kaiser I, Pfahlberg AB, Uter W, et al. Risk prediction models for melanoma: a systematic review on the heterogeneity in model development and validation. Int J Environ Res Public Health 2020;17:7919.
58. Himmelreich JC, Veelers L, Lucassen WA, et al. Prediction models for atrial fibrillation applicable in the community: a systematic review and meta-analysis. Europace 2020;22:684–94.
59. Damen JA, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016;353:i2416.
60. Xu W, Huang J, Yu Q, et al. A systematic review of the status and methodological considerations for estimating risk of first ever stroke in the general population. Neurol Sci 2021;42:2235–47.
61. Boer CG, Hatzikotoulas K, Southam L, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell 2021; 184:4784–818.
62. Pepe MS, Janes H, Longton G, et al. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 2004;159:882–90.
63. Halilaj E, Le Y, Hicks JL, et al. Modeling and predicting osteoarthritis progression: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2018;26:1643–50.
64. Widera, Welsing PM, Ladel C, et al. Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data. Sci Rep 2020;10:8427.
65. Losina E, Klara K, Michl G, et al. Development and feasibility of a personalized, interactive risk calculator for knee osteoarthritis. BMC Musculoskelet Disord 2015;16:312.
66. Michl GL, Katz JN, Losina E. Risk and risk perception of knee osteoarthritis in the US: a population-based study. Osteoarthritis Cartilage 2016;24:593–6.