

Clinical pathways before total knee replacement in England: a population-based retrospective cohort study using sequence analysis and multilinked electronic health records

MATHEW, Smitha, PARRY, Emma, PEAT, George <<http://orcid.org/0000-0002-9008-0184>> and YU, Dahai <<http://orcid.org/0000-0002-8449-7725>>

Available from Sheffield Hallam University Research Archive (SHURA) at:
<https://shura.shu.ac.uk/38053/>

This document is the Published Version [VoR]

Citation:

MATHEW, Smitha, PARRY, Emma, PEAT, George and YU, Dahai (2026). Clinical pathways before total knee replacement in England: a population-based retrospective cohort study using sequence analysis and multilinked electronic health records. *The Lancet Primary Care*, 2 (1): 100098. [Article]

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Clinical pathways before total knee replacement in England: a population-based retrospective cohort study using sequence analysis and multilinked electronic health records



Smitha Mathew, Emma Parry, George Peat, Dahai Yu



Summary

Background The UK National Institute for Health and Care Excellence recommends a structured care pathway for osteoarthritis; however, real-world practice often diverges from these guidelines, and the patient journey through primary care before knee replacement is not well described. We aimed to characterise real-world care pathways before total knee replacement and to identify patient characteristics associated with different patterns of care using sequence analysis and linked electronic health records.

Methods In this population-based retrospective cohort study, we identified patients aged 45 years or older who underwent primary total knee replacement between Jan 1, 2007, and Dec 31, 2021, from the Clinical Practice Research Datalink Aurum database, linked to Hospital Episode Statistics data in England, UK. Sequence analysis was applied to classify health-care patterns in the 10 years preceding total knee replacement, based on care strategies such as consultations, analgesic prescriptions, referrals, and knee surgeries. Multinomial logistic regression was used to determine the associations between patients' characteristics and care pathways.

Findings We included data for 129 470 patients in the analysis (mean age 69.7 years [SD 9.4]; 73 212 [56.5%] were female, 56 258 [43.5%] were male). Sequence analysis identified five distinct care pathways: low health-care use (48 922 [37.8%] of 129 470 patients), moderate health-care use (30 947 [23.9%] patients), opioids predominant (25 983 [20.1%] patients), gabapentinoids and antidepressants predominant (5910 [4.6%] patients), and non-steroidal anti-inflammatory drugs predominant (17 708 [13.7%] patients). Patients from the most deprived areas had higher odds of being in the opioids predominant (odds ratio 3.86, 95% CI 3.57–4.18) and gabapentinoids and antidepressants predominant (3.50, 3.07–3.99) clusters, relative to the low health-care use cluster. In addition, female sex, age above 75 years, BMI of 25.0 kg/m² or higher, multiple comorbidities, and being Black or South Asian ethnicity (vs White ethnicity) were significantly associated with the odds of membership in all other clusters relative to the low health-care use cluster.

Interpretation The identification of distinct and non-guideline-concordant pathways highlights persistent gaps in primary care management of patients with osteoarthritis. These findings underscore the need for targeted primary care interventions to support timely, evidence-based, non-pharmacological management and to promote more consistent and equitable care before knee replacement.

Funding National Natural Science Foundation of China, Henan Key Research and Development Special Project, and National Institute for Health and Care Research Applied Research Collaboration West Midlands.

Copyright © 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Osteoarthritis is a leading cause of disability that substantially restricts an individual's ability to carry out daily activities.¹ In 2023, more than 635 million people were estimated to have osteoarthritis worldwide, and the knee is the most commonly affected joint.² An estimated 10 million people in the UK are affected by osteoarthritis, with knee osteoarthritis accounting for nearly half of these cases.³ Osteoarthritis is predominantly managed in primary care and accounts for approximately 10% of all primary care consultations in the UK.⁴

Total knee replacement is a definitive and cost-effective intervention for managing end-stage knee osteoarthritis.⁵

In England and Wales, approximately 134 652 primary total knee replacement procedures were performed in 2024 for osteoarthritis.⁶ Although total knee replacement is a definitive intervention for end-stage disease, there remains a critical need to optimise conservative care during the earlier stages of osteoarthritis. The UK National Institute for Health and Care Excellence (NICE) outlines a structured care pathway for osteoarthritis, beginning with diagnosis and self-management, followed by core non-surgical treatments (including patient education, exercise, and weight management), a broad range of non-pharmacological and pharmacological interventions, and referral for consideration of joint surgery.⁷ Within this pathway, primary care is

Lancet Prim Care 2026;
2: 100098

Published Online January 29,
2026
<https://doi.org/10.1016/j.lanprc.2025.100098>

Henan Key Laboratory of
Chronic Disease Prevention and
Therapy & Intelligent Health
Management, the First
Affiliated Hospital, Zhengzhou
University, Zhengzhou, China
(S Mathew MSc, D Yu PhD);
Primary Care Centre Versus
Arthritis, School of Medicine,
Keele University, Staffordshire,
UK (S Mathew, E Parry PhD,
G Peat PhD, D Yu); Centre for
Applied Health & Social Care
Research, Sheffield Hallam
University, Sheffield, UK (G Peat)

Correspondence to:
Dr Dahai Yu, Henan Key
Laboratory of Chronic Disease
Prevention and Therapy &
Intelligent Health Management,
the First Affiliated Hospital,
Zhengzhou University,
Zhengzhou 450052, China
d.yu@keele.ac.uk

Research in context

Evidence before this study

Total knee replacement is a definitive and cost-effective intervention for managing end-stage knee osteoarthritis. The UK National Institute for Health and Care Excellence recommends a structured care pathway for osteoarthritis, beginning with diagnosis and self-management, followed by core non-surgical treatments such as patient education, physical activity, and weight management, progressing to broader non-pharmacological and pharmacological interventions and referral for joint surgery when appropriate. However, real-world care pathways often differ from recommended guidelines and remain poorly characterised. We searched MEDLINE from database inception to July 31, 2025, using the following search terms: ("care pathways" OR "clinical pathways" OR "care patterns" OR "care trajectories" OR "patterns") AND ("osteoarthritis" OR "knee replacement" OR "knee arthroplasty"). The search results were limited to peer-reviewed publications in English. We found 14 studies that investigated treatment patterns in osteoarthritis, and seven that had specifically examined knee osteoarthritis. However, these studies primarily focused on the prevalence of individual care strategies or short-term episodes of care, rather than providing a comprehensive assessment of longitudinal care pathways. Mapping such pathways allows for a more complete

understanding of patient journeys, revealing patterns and deviations from guideline-recommended care.

Added value of this study

To our knowledge, this study offers the first comprehensive mapping of real-world management trajectories preceding total knee replacement using large-scale UK electronic health record data and an innovative sequence analysis approach. Unlike previous studies that examined individual treatments or short-term care episodes, this research captures entire care pathways spanning 10 years, revealing five distinct care pathway typologies that reflect heterogeneous patterns of osteoarthritis management in routine clinical practice.

Implications of all the available evidence

A clear understanding of health-care pathways helps identify service gaps and highlight areas for improvement. Substantial socioeconomic inequalities were evident, with individuals from the most deprived areas being more likely to receive non-guideline-based management. These findings are crucial for enhancing health-care delivery, reducing disparities in service provision, and informing evidence-based policies that promote more effective and equitable care for people with knee osteoarthritis.

the principal setting in which osteoarthritis is diagnosed, evidence-based non-surgical treatments are initiated, and decisions regarding referral to secondary care are made. However, the real-world care pathways followed by patients in primary care remain unclear and might deviate substantially from recommended guidelines.

Evidence shows that many patients with knee osteoarthritis do not receive guideline-recommended care.^{8–14} A Danish cohort study found that only 23% of patients received recommended treatments before or within 6 months of orthopaedic consultation, while 62% received non-recommended treatments.¹⁴ US and Dutch studies have reported frequent opioid use, corticosteroid injections, imaging, and referrals to secondary care before total knee replacement, and a Swedish study showed underuse and delay of physiotherapy.^{8,12,13} These studies highlight wide variation and suboptimal alignment with guidelines, but they focused mainly on individual treatments or short-term care episodes rather than the overall pattern of care experienced by patients over time. To address this gap, we used large-scale linked electronic health records (EHRs) and sequence analysis to map and understand 10-year care pathways preceding total knee replacement, and we examined how these care pathways varied by patient characteristics to identify disparities and opportunities for improving osteoarthritis care. Characterising patients' journeys through primary and secondary care before total knee replacement is essential for identifying where primary care strategies might enhance the equity, quality, and

timeliness of care in the UK and other primary care-led health systems worldwide.

Sequence analysis is an emerging statistical approach in health-care research and has been widely applied in social science to study life-course trajectories and career patterns.¹⁵ Sequence analysis provides a patient-level, time-sensitive view of care pathways, enabling detailed comparison of individual trajectories and revealing variation and timing of care. When process-mining methods are applied directly to individual-level event logs, they generate very similar patient-level, time-ordered representations of care.¹⁶ In this setting, sequence analysis and process mining are conceptually synonymous, differing mainly in terminology and in specific implementation choices (eg, distance metrics or clustering algorithms) rather than in the structure of the resulting pathways. The robustness and applicability of sequence analysis have been shown across diverse health-care contexts, including chronic disease management and health-care utilisation.^{17–19} Sequence analysis provides a comprehensive understanding of real-world pathways, allowing researchers to identify unmet needs, guide resource allocation, and inform strategies to improve the equity and quality of care for people with osteoarthritis.

Thus, the objectives of our study were to explore the health-care pathways before knee replacement in patients with osteoarthritis using sequence analysis and linked EHRs, and to investigate the association between patient characteristics and the identified care pathways.

Methods

Study design and participants

We did a population-based retrospective cohort study using data from the Clinical Practice Research Datalink (CPRD) Aurum database linked to Hospital Episode Statistics (HES) Admitted Patient Care (APC) and Index of Multiple Deprivation (IMD) data. The CPRD Aurum contains anonymised patient-level primary care data from general practices in England, UK, including demographics, consultations, diagnoses, prescriptions, diagnostic tests, and referrals.²⁰ HES APC provides comprehensive information on all National Health Service (NHS) hospital admissions, while private health care is not captured within CPRD or HES. In the UK, health care is publicly funded through the NHS, which provides care for all residents based on clinical need, including primary and secondary services. The IMD is a composite measure encompassing seven domains of material deprivation: income, employment, education and skills training, health deprivation and disability, barriers to housing and services, local crime, and living environment.²¹ As of the June, 2024 data release, CPRD Aurum comprised records for 47·8 million patients of whom 16·2 million were active from 1596 general practices in England.²²

The study was approved by the independent scientific advisory committee for CPRD research (Protocol #24_003806). As the study used anonymised routinely collected data, the requirement for individual patient consent was waived. No further ethical permissions were required for the analyses of these anonymised patient-level data. We reported the study according to the Reporting of Studies Conducted using Observational Routinely-collected Health Data (RECORD) guidelines (appendix pp 2–8).²³

The study population was extracted by CPRD in July, 2024, and it included all patients with a recorded total knee replacement in HES APC between April 1, 1997, and Dec 31, 2021. Total knee replacements were identified using all relevant Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures (OPCS) codes from the HES APC database (appendix p 9). Eligibility criteria were: (1) age 45 years or older; (2) a recorded primary total knee replacement between Jan 1, 2007, and Dec 31, 2021; (3) continuous registration in the health-care database for a minimum of 10 years before the index year (the year of primary total knee replacement); (4) consent to linkage with HES APC and IMD data; and (5) no recorded total knee replacement within the 10 years preceding the index year. We did not require a recorded diagnosis of osteoarthritis or joint pain as an inclusion criterion because these diagnoses are frequently under-recorded in primary care data, including among patients who subsequently undergo total knee replacement. According to the National Joint Registry, 98% of total knee replacements in the UK are performed for osteoarthritis;⁶ therefore, excluding cases without an explicit osteoarthritis code would risk misclassification due to incomplete

diagnostic coding. The year 2007 was selected as the initial index year because HES APC data have been available in CPRD since 1997, allowing a 10-year look-back period for all potentially eligible participants.

Procedures

For our analysis of care sequences, the study data manager extracted data on care strategies from the CPRD database, including all recorded consultations for knee pain, analgesic prescriptions, referrals for imaging, physiotherapy and specialist care, intra-articular corticosteroid injections, and surgical procedures over the 10-year period, regardless of whether they were specifically related to knee osteoarthritis. Analgesic prescriptions were categorised into oral or topical non-steroidal anti-inflammatory drugs (NSAIDs; eg, ibuprofen, naproxen, diclofenac), non-NSAIDs (eg, paracetamol, capsaicin), weak opioids (eg, codeine 8 mg and paracetamol), moderate opioids (eg, dihydrocodeine 20 mg), strong opioids (eg, tramadol 50 mg), very strong opioids (eg, oxycodone), antidepressants (ie, amitriptyline and duloxetine), and gabapentinoids (ie, gabapentin and pregabalin). Knee surgeries included meniscus surgery, ligament reconstruction, osteotomy, and synovial surgery. All prescribed analgesics were identified using appropriate product codes from the Dictionary of Medicines and Devices, based on lists developed in previous CPRD studies conducted by the University of Keele's Medical Records Research Group (code lists available online). Primary care consultations for knee pain and referrals for imaging, physiotherapy, orthopaedic specialist care, and intra-articular corticosteroid injections were defined using Systematised Nomenclature of Medicine Clinical Terms (SNOMED CT) codes (appendix pp 10–14). Surgical procedures were identified using OPCS codes (appendix pp 14–15).

We also extracted patients' age, sex, BMI, comorbidities, smoking status, drinking status, ethnicity, and geographical region. Sex was self-reported by patients at the time of general practice registration and recorded in CPRD demographic records. Ethnicity was obtained from linked HES data, where it is self-reported by patients at hospital admission and recorded using standard CPRD categories. The list of comorbidities was based on Charlson's 17 pre-defined clinical conditions,²⁴ identified through the presence of relevant diagnostic codes in patients' health records over the 10-year period. All covariates were measured at the index year, except for BMI, smoking status, and drinking status, which were defined using the most recent value recorded within the 10 years before the index year to account for a higher likelihood of missing data. Neighbourhood deprivation was assessed using the 2019 English IMD rank. We implemented several measures to minimise potential sources of bias, including data completeness checks, use of validated code lists, and restriction to patients with continuous registration and linked data availability.

For the code lists see <https://www.keele.ac.uk/hcdr/>

See Online for appendix

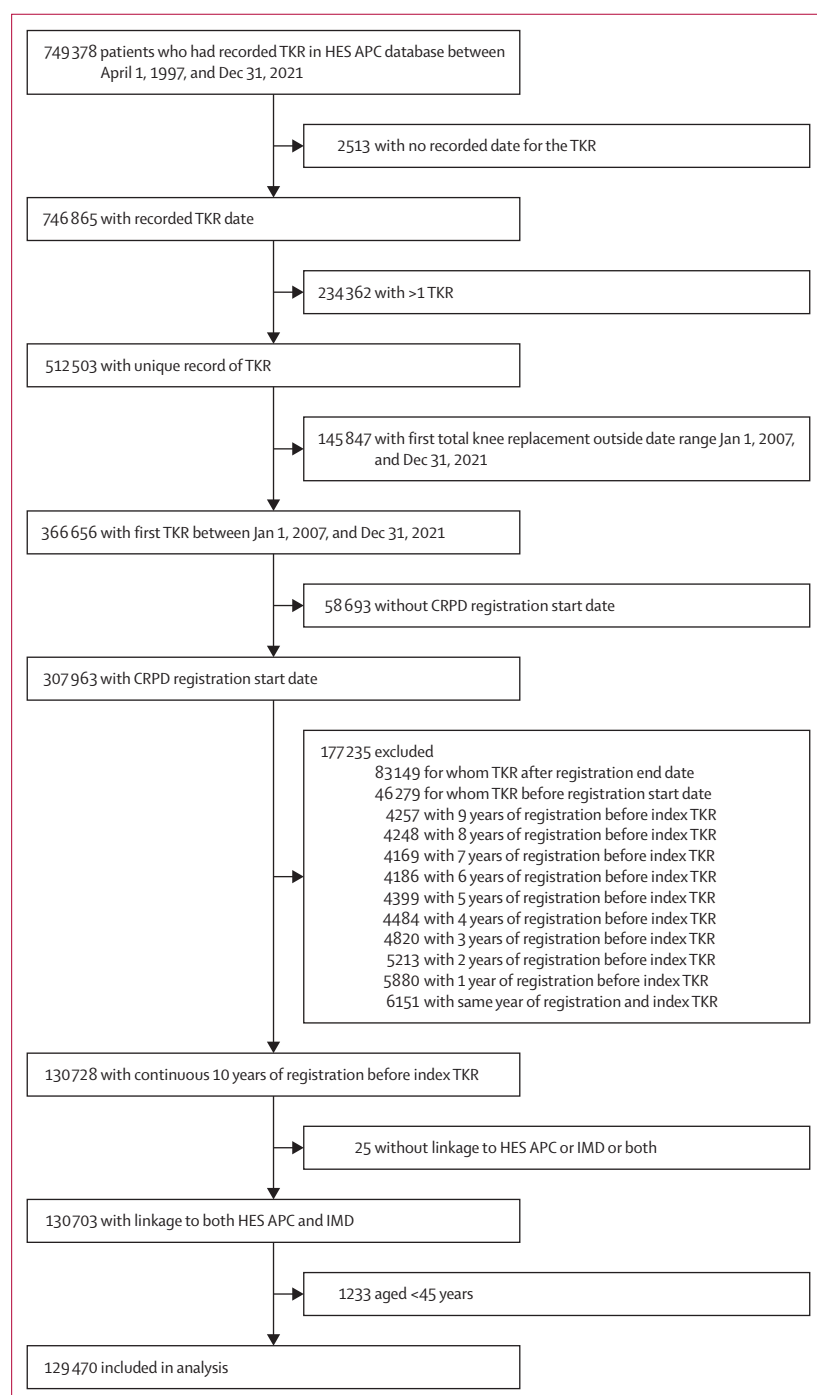


Figure 1: CPRD flowchart

APC=Admitted Patient Care. CPRD=Clinical Practice Research Datalink. HES=Hospital Episode Statistics. IMD=Index of Multiple Deprivation. TKR=total knee replacement.

Statistical analysis

We used descriptive statistics to summarise the key characteristics of the study population, with mean (SD) for continuous variables and counts and percentages for categorical variables. We examined the annual prevalence of

	Patients (N=129 470)
Sex	
Male	56 258 (43·5%)
Female	73 212 (56·5%)
Mean age, years	69·7 (9·4)
Age group	
45–54 years	8580 (6·6%)
55–64 years	29 349 (22·7%)
65–74 years	49 024 (37·9%)
75–84 years	36 373 (28·1%)
≥85 years	6144 (4·7%)
Mean BMI, kg/m ²	29·71 (6·0)
BMI group	
Underweight (<18·5 kg/m ²)	1384 (1·1%)
Normal (18·5–24·9 kg/m ²)	24 344 (18·8%)
Overweight (25·0–29·9 kg/m ²)	43 200 (33·4%)
Obese (≥30·0 kg/m ²)	52 320 (40·4%)
Missing	8222 (6·4%)
Comorbidities	
0	54 055 (41·8%)
1	41 490 (32·0%)
2	20 983 (16·2%)
≥3	12 942 (10·0%)
Smoking status	
Never smoker	67 994 (52·5%)
Ex-smoker	39 648 (30·6%)
Current smoker	18 722 (14·5%)
Missing	3106 (2·4%)
Drinking status	
Never drinker	10 244 (7·9%)
Ever drinker	14 943 (11·5%)
Missing	104 283 (80·5%)
Ethnicity	
White	120 659 (93·2%)
Black	1461 (1·1%)
South Asian	3723 (2·9%)
Other or unknown	3627 (2·8%)
Geographical region	
North East	6226 (4·8%)
North West	26 374 (20·4%)
Yorkshire & The Humber	5682 (4·4%)
East Midlands	3164 (2·4%)
West Midlands	24 457 (18·9%)
East of England	5910 (4·6%)
London	13 452 (10·4%)
South East	26 300 (20·3%)
South West	17 905 (13·8%)
Index of Multiple Deprivation decile	
1 (least deprived)	15 780 (12·2%)
2	14 721 (11·4%)
3	15 058 (11·6%)
4	15 415 (11·9%)
5	13 012 (10·1%)
6	13 205 (10·2%)

(Table 1 continues on next page)

	Patients (N=129 470)
(Continued from previous page)	
7	12 415 (9.6%)
8	10 606 (8.2%)
9	9958 (7.7%)
10 (most deprived)	8741 (6.8%)
Missing	559 (0.4%)
Index year	
2007	8455 (6.5%)
2008	9003 (7.0%)
2009	8683 (6.7%)
2010	9037 (7.0%)
2011	9012 (7.0%)
2012	9019 (7.0%)
2013	9151 (7.1%)
2014	10 119 (7.8%)
2015	10 282 (7.9%)
2016	10 553 (8.2%)
2017	10 434 (8.1%)
2018	9736 (7.5%)
2019	10 244 (7.9%)
2020	4943 (3.8%)
2021	799 (0.6%)

Data are n (%) or mean (SD).

Table 1: Baseline characteristics of patients

different care strategies over the 10 years preceding the index total knee replacement, stratified by age, sex, and other covariates. We used sequence analysis to identify clusters of patients who followed similar care pathways before total knee replacement. A full methodological description of the sequence analysis is in the appendix (pp 16–18). Briefly, the initial step in sequence analysis is to define the categorical states and the time interval used to construct the care sequence for each patient. A state represents a specific health-care event or care strategy (eg, a knee pain consultation, physiotherapy referral, or analgesic prescription) recorded within a given time interval. We used 6-month intervals over the 10-year observation period, resulting in a sequence of 20 care events for each patient. Dissimilarities between sequences were calculated using optimal matching with transition rate-based substitution costs and a cost of one for an insertion or deletion. Patients were grouped according to similar care patterns using the CLARA (Clustering Large Applications) algorithm. We determined the optimal number of clusters using a combination of cluster quality indices and clinical interpretability. We used state distribution plots to summarise the sequences. These plots show, for each timepoint, the proportion of patients in each care state, providing a visual overview of how care patterns changed over time.

Associations between cluster membership and patient characteristics were examined using the χ^2 test. In post-hoc

analyses, standardised Pearson residuals were calculated to identify cells contributing most to the overall χ^2 statistic; absolute residual values of 2 or higher indicate that the observed count differs significantly from the expected count under the null hypothesis. We fitted a single multinomial logistic regression model in which all patient characteristics were entered simultaneously, thereby mutually adjusting each factor for the others. As this was an exploratory analysis, we did not identify confounders for specific exposure–outcome associations; instead, estimates reflect associations from one mutually adjusted model. Associations were reported as odds ratios (ORs) with 95% CIs. Missing data for BMI (n=8222), IMD (n=559), and smoking status (n=3106) were handled using multiple imputation by chained equations under the assumption that data were missing at random.²⁵ The imputation model included all variables from the multinomial regression model as well as the cluster membership. Variable-specific imputation models were applied: linear regression for BMI, ordered logistic regression for IMD, and multinomial logistic regression for smoking status. Height and weight were included as auxiliary variables to support BMI imputation. A total of 100 imputations were generated, and estimates were combined using Rubin's rules to obtain overall estimates and SEs.

We did subgroup analyses stratifying the population by sex, age group (≤ 65 years and > 65 years), and calendar period (2007–11, 2012–16, and 2017–21). We did sensitivity analyses to assess the robustness of the findings using alternative dissimilarity measures, including the longest common subsequences, hamming distance, and dynamic hamming distance, and by applying a 12-month time interval. In addition, we did a sensitivity analysis restricted to the period 2007–19 to examine care patterns before the COVID-19 pandemic.

All statistical analyses were done using Stata (version 18.0). Statistical significance was set at $p < 0.05$. Sequence analysis and visualisation were performed using the TraMineR and WeightedCluster packages in R version 4.2.1.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or the writing of the report.

Results

We identified 129 470 eligible individuals who underwent primary total knee replacement during the study period (figure 1). The characteristics of study participants are shown in table 1. Mean age was 69.7 years (SD 9.4), 73 212 (56.5%) of 129 470 participants were female and 56 258 (43.5%) were male, 52 320 (40.4%) were classified as obese, 120 659 (93.2%) were White, and 8741 (6.8%) were from the most deprived areas.

The prevalence of care strategies over the 10 years preceding the index total knee replacement stratified by age, sex, and other covariates is presented in the appendix (pp 19–42).

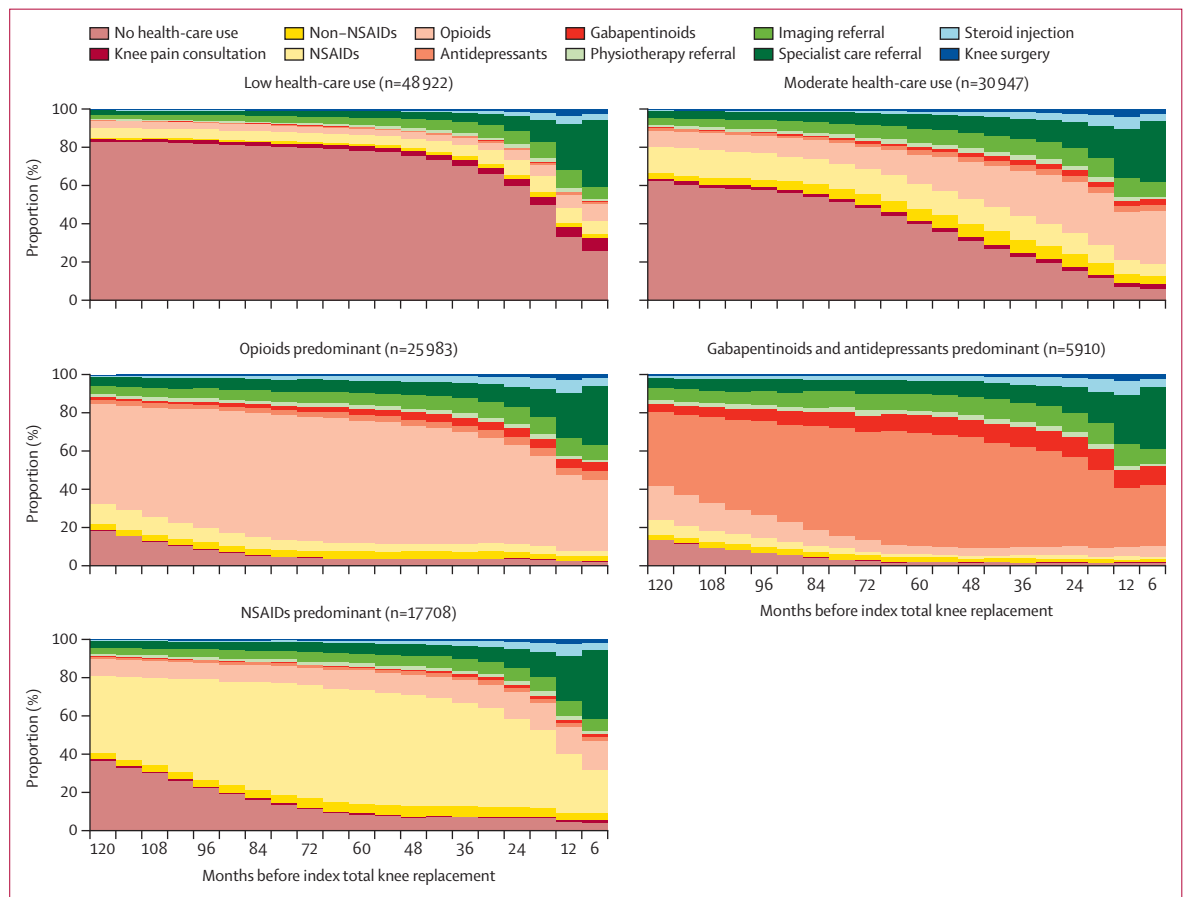


Figure 2: State distribution plot of care pattern typologies

The x-axis shows months before the index total knee replacement, and the y-axis shows the proportion of patients in each care state. Different colours indicate different care strategies. NSAIDs=non-steroidal anti-inflammatory drugs. For computational feasibility, we combined all types of surgery (meniscus surgery, ligament surgery, osteotomy, and synovial surgery) into the category of knee surgery (appendix p 16).

Consultations and analgesic prescriptions generally increased gradually over time, with a notable rise in the 2 years before surgery. Physiotherapy referrals generally remained consistently low (<10%), whereas referrals for imaging, injections, and specialists care generally increased substantially closer to surgery.

Sequence analysis identified five distinct clusters based on similar care patterns (figure 2). The optimal number of clusters was chosen using cluster quality indices (appendix p 43) and clinical interpretability. Cluster 1 included 48 922 (37·8%) of 129 470 individuals and was characterised by consistently low health-care use between 10 and 3 years before surgery, followed by a gradual increase in the 2 years preceding total knee replacement. This group was labelled as low health-care use. Cluster 2 included 30 947 (23·9%) individuals and who showed a moderate level of health-care use that increased progressively over time; this group was labelled as moderate health-care use. Cluster 3 included 25 983 (20·1%) and exhibited a predominant use of opioids, it was hence labelled as opioids predominant. Cluster 4 included 5910 (4·6%)

individuals and was characterised by sustained use of gabapentinoids and antidepressants and was labelled as gabapentinoids and antidepressants predominant. Finally, cluster 5 included 17 708 (13·7%) individuals and was characterised by predominant NSAIDs use; this group was therefore labelled as NSAIDs predominant.

Patients' characteristics across clusters of similar care patterns are presented in table 2. Table 3 summarises the results of the multinomial logistic regression analysis conducted to identify potential predictors of cluster membership. We calculated ORs to estimate the likelihood of patients belonging to each cluster, using the low health-care use group as the reference cluster. Patients from the most deprived areas had significantly higher odds of belonging to the opioids predominant cluster (OR 3·86, 95% CI 3·57–4·18) and gabapentinoids and antidepressants predominant cluster (3·50, 3·07–3·99) than did patients from the least deprived areas, relative to the low health-care use cluster. In addition, female sex, age above 75 years, BMI of 25·0 kg/m² or higher, multiple comorbidities, and being Black or South Asian ethnicity (vs White ethnicity) were

	Low health-care use (n=48 922)		Moderate health-care use (n=30 947)		Opioids predominant (n=25 983)		Gabapentinoids and antidepressants predominant (n=5910)		NSAIDs predominant (n=17 708)		p value*
	N (%)	Residual	N (%)	Residual	N (%)	Residual	N (%)	Residual	N (%)	Residual	
Sex											
Male	25 333 (51.8%)	47.1	12 690 (41.0%)	-10.0	8812 (33.9%)	-34.7	1307 (22.1%)	-33.9	8116 (45.8%)	6.9	<0.0001
Female	23 589 (48.2%)	-47.1	18 257 (59.0%)	10.0	17 171 (66.1%)	34.7	4603 (77.9%)	33.9	9592 (54.2%)	-6.9	..
Age group											
45-54 years	3970 (8.1%)	16.8	2412 (7.8%)	9.5	1212 (4.7%)	-14.2	367 (6.2%)	1.3	619 (3.5%)	-18.0	<0.0001
55-64 years	12 561 (25.7%)	20.1	7543 (24.4%)	8.2	4657 (17.9%)	-20.4	1358 (23.0%)	0.6	3230 (18.2%)	-15.2	..
65-74 years	18 242 (37.3%)	-3.3	11 137 (36.0%)	7.8	10 052 (38.7%)	3.1	2297 (38.9%)	1.6	7296 (41.2%)	9.9	..
75-84 years	12 280 (25.1%)	-18.7	8347 (27.0%)	5.0	8514 (32.8%)	18.8	1610 (27.2%)	1.5	5622 (31.8%)	11.7	..
≥85 years	1869 (3.8%)	-12.2	1508 (4.9%)	1.2	1548 (6.0%)	10.3	278 (4.7%)	0.2	941 (5.3%)	3.8	..
BMI group											
Underweight (<18.5 kg/m ²)	444 (1.0%)	-4.6	300 (1.0%)	-2.1	353 (1.5%)	5.5	76 (1.4%)	1.8	211 (1.3%)	1.7	<0.0001
Normal (18.5-24.9 kg/m ²)	9865 (21.4%)	8.8	5431 (18.6%)	-7.2	4502 (18.9%)	-5.2	988 (18.1%)	-3.7	3558 (21.5%)	4.8	..
Overweight (25.0-29.9 kg/m ²)	17 760 (38.5%)	16.1	9984 (34.2%)	-5.8	7671 (32.2%)	-12.5	1726 (31.6%)	-6.3	6059 (36.5%)	2.6	..
Obese (≥30.0 kg/m ²)	18 109 (39.2%)	-21.7	13 459 (46.1%)	11.8	11 323 (47.5%)	15.1	2670 (48.9%)	8.8	6759 (40.8%)	6.7	..
Comorbidities											
0	25 147 (51.4%)	54.9	12 398 (40.1%)	-6.9	7673 (29.5%)	-44.7	1721 (29.1%)	-20.2	7116 (40.2%)	-4.6	<0.0001
1	14 655 (30.0%)	-12.6	10 130 (32.7%)	3.0	8875 (34.2%)	8.2	1913 (32.4%)	0.5	5917 (33.4%)	4.2	..
2	6133 (12.5%)	-27.9	5173 (16.7%)	2.8	5379 (20.7%)	22.0	1245 (21.1%)	10.4	3053 (17.2%)	4.0	..
≥3	2987 (6.1%)	-36.4	3246 (10.5%)	3.3	4056 (15.6%)	33.8	1031 (17.5%)	19.5	1622 (9.2%)	-4.0	..
Smoking status											
Never smoker	25 856 (55.1%)	7.2	16 101 (52.9%)	-3.8	13 160 (51.2%)	-9.4	3045 (52.2%)	-2.5	9819 (56.3%)	7.8	<0.0001
Ex-smoker	14 354 (30.6%)	-4.6	9766 (32.1%)	3.0	8265 (32.2%)	3.0	1852 (31.8%)	0.7	5407 (31.0%)	-1.1	..
Current smoker	6712 (14.3%)	-4.0	4599 (15.1%)	1.5	4277 (16.6%)	9.2	935 (16.0%)	2.7	2209 (12.7%)	-8.6	..
Drinking status											
Never drinker	2696 (33.2%)	-16.7	2534 (40.4%)	-0.4	2898 (48.9%)	14.8	752 (50.3%)	7.9	1352 (40.3%)	-0.4	<0.0001
Ever drinker	5436 (66.9%)	16.8	3739 (59.6%)	0.4	3033 (51.1%)	-14.8	743 (49.7%)	-7.9	2004 (59.7%)	0.4	..
Ethnicity											
White	46 225 (94.5%)	14.4	28 741 (92.9%)	-2.6	23 962 (92.2%)	-7.0	5507 (93.2%)	-0.1	16 224 (91.6%)	-9.0	<0.0001
Black	317 (0.7%)	-12.8	442 (1.4%)	5.7	428 (1.7%)	8.9	71 (1.2%)	0.5	203 (1.2%)	0.2	..
South Asian	696 (1.4%)	-24.8	1014 (3.3%)	4.8	1023 (3.9%)	11.5	217 (3.7%)	3.8	773 (4.4%)	12.8	..
Other or unknown	1684 (3.4%)	10.9	750 (2.4%)	-4.6	570 (2.2%)	-6.6	115 (2.0%)	-4.1	508 (2.9%)	0.6	..
Geographical region											
North East	2007 (4.1%)	-9.3	1388 (4.5%)	-3.1	1706 (6.6%)	14.8	389 (6.6%)	6.5	736 (4.2%)	-4.4	<0.0001
North West	9125 (18.7%)	-12.0	6325 (20.4%)	0.3	6099 (23.5%)	13.9	1209 (20.5%)	0.2	3616 (20.4%)	0.2	..
Yorkshire & The Humber	1980 (4.1%)	-4.7	1329 (4.3%)	-0.9	1150 (4.4%)	0.3	334 (5.7%)	4.9	889 (5.0%)	4.4	..
East Midlands	1221 (2.5%)	0.9	694 (2.2%)	-2.6	655 (2.5%)	0.9	160 (2.7%)	1.3	434 (2.5%)	0.1	..
West Midlands	9111 (18.6%)	-1.9	5895 (19.1%)	0.8	5243 (20.2%)	5.9	982 (16.6%)	-4.6	3226 (18.2%)	-2.5	..
East of England	2405 (4.9%)	4.7	1375 (4.4%)	-1.2	1014 (3.9%)	-5.7	252 (4.3%)	-1.1	864 (4.9%)	2.2	..
London	4987 (10.2%)	-1.8	3365 (10.9%)	3.2	2656 (10.2%)	-1.0	525 (8.9%)	-3.9	1919 (10.8%)	2.1	..
South East	11 156 (22.8%)	17.4	6219 (20.1%)	-1.1	4242 (16.3%)	-17.9	1212 (20.5%)	0.4	3471 (19.6%)	-2.5	..
South West	6930 (14.2%)	2.7	4357 (14.1%)	1.5	3218 (12.4%)	-7.5	847 (14.3%)	1.2	2553 (14.4%)	2.4	..
Index of Multiple Deprivation											
1 (least deprived)	7336 (15.1%)	24.7	3519 (11.4%)	5.3	2081 (8.0%)	-23.2	558 (9.5%)	-6.7	2286 (12.9%)	3.0	<0.0001
2	6302 (13.0%)	13.9	3483 (11.3%)	1.0	2234 (8.6%)	-16.0	583 (9.9%)	-3.8	2119 (12.0%)	2.5	..
3	6275 (13.0%)	11.0	3501 (11.3%)	2.2	2558 (9.9%)	-10.3	612 (10.4%)	-3.2	2112 (12.0%)	1.2	..
4	6236 (12.9%)	7.8	3750 (12.1%)	1.1	2641 (10.2%)	-9.9	644 (10.9%)	-2.5	2144 (12.1%)	0.7	..
5	5002 (10.3%)	2.1	3122 (10.4%)	0.1	2423 (9.3%)	-4.6	577 (9.8%)	-0.8	1888 (10.7%)	2.8	..
6	4802 (9.9%)	-3.1	3221 (10.4%)	1.2	2697 (10.4%)	0.9	634 (10.7%)	1.3	1851 (10.5%)	1.0	..
7	4249 (8.8%)	-8.1	3102 (10.0%)	2.8	2840 (10.9%)	8.0	583 (9.9%)	0.7	1641 (9.3%)	1.7	..
8	3368 (7.0%)	-12.9	2635 (8.5%)	2.2	2710 (10.4%)	14.5	498 (8.4%)	0.6	1395 (7.9%)	1.7	..
9	2803 (5.8%)	-20.2	2449 (7.9%)	1.5	2915 (11.2%)	23.7	572 (9.7%)	5.8	1219 (6.9%)	4.5	..
10 (most deprived)	2080 (4.3%)	-27.6	2134 (6.9%)	1.0	2859 (11.0%)	30.4	642 (10.8%)	12.8	1026 (5.8%)	5.6	..

(Table 2 continues on next page)

	Low health-care use (n=48 922)		Moderate health-care use (n=30 947)		Opioids predominant (n=25 983)		Gabapentinoids and antidepressants predominant (n=5910)		NSAIDs predominant (n=17 708)		p value*
	N (%)	Residual	N (%)	Residual	N (%)	Residual	N (%)	Residual	N (%)	Residual	
(Continued from previous page)											
Index year											
2007	2903 (5.9%)	-6.8	1911 (6.2%)	2.9	2002 (7.7%)	8.6	226 (3.8%)	-8.6	1413 (8.0%)	8.4	<0.0001
2008	3210 (6.6%)	-4.3	1970 (6.4%)	4.7	2011 (7.7%)	5.6	282 (4.8%)	-6.8	1530 (8.6%)	9.5	..
2009	3088 (6.3%)	-4.4	1951 (6.3%)	3.2	1946 (7.5%)	5.6	261 (4.4%)	-7.2	1437 (8.1%)	8.1	..
2010	3235 (6.6%)	-4.0	2054 (6.6%)	2.7	1989 (7.7%)	4.8	316 (5.4%)	-5.0	1443 (8.2%)	6.6	..
2011	3173 (6.5%)	-5.2	2199 (7.1%)	1.2	1950 (7.5%)	3.9	367 (6.2%)	-2.3	1323 (7.5%)	2.9	..
2012	3234 (6.6%)	-3.9	2259 (7.3%)	2.6	1900 (7.3%)	2.5	336 (5.7%)	-4.0	1290 (7.3%)	1.8	..
2013	3197 (6.5%)	-5.8	2209 (7.1%)	0.6	1993 (7.7%)	4.2	430 (7.3%)	0.6	1322 (7.5%)	2.2	..
2014	3792 (7.8%)	-0.7	2497 (8.1%)	1.9	2005 (7.7%)	0.7	465 (7.9%)	0.2	1360 (7.7%)	0.7	..
2015	3933 (8.0%)	1.0	2559 (8.3%)	2.4	1944 (7.5%)	3.1	497 (8.4%)	1.4	1349 (7.6%)	1.7	..
2016	4226 (8.6%)	5.0	2511 (8.1%)	0.3	1994 (7.7%)	3.1	554 (9.4%)	3.5	1268 (7.2%)	5.2	..
2017	4178 (8.5%)	5.0	2576 (8.3%)	2.0	1925 (7.4%)	4.3	572 (9.7%)	4.7	1183 (6.7%)	7.3	..
2018	3927 (8.0%)	5.4	2398 (7.8%)	1.8	1722 (6.6%)	6.1	583 (9.9%)	7.0	1106 (6.3%)	6.9	..
2019	4230 (8.7%)	7.6	2518 (8.1%)	1.7	1702 (6.6%)	9.1	678 (11.5%)	10.4	1116 (6.3%)	8.5	..
2020	2225 (4.6%)	10.7	1156 (3.7%)	0.9	784 (3.0%)	7.5	283 (4.8%)	4.0	495 (2.8%)	7.6	..
2021	371 (0.8%)	5.1	179 (0.6%)	1.00	116 (0.5%)	3.9	60 (1.0%)	4.0	73 (0.4%)	3.8	..

Data are n (%) and adjusted residuals. An absolute residual of ≥ 2 indicates the observed count differs significantly from the expected count under the null hypothesis ($p < 0.05$). Positive residuals denote over-representation, negative residuals under-representation. NSAIDs=non-steroidal anti-inflammatory drugs. *p value for χ^2 test for association between patients' characteristics and identified care pathways.

Table 2: Distribution of care pattern typology across patients' characteristics with χ^2 test and post-hoc residuals (N=129 470)

significantly associated with the odds of membership in all other clusters relative to the low health-care use cluster. Multinomial regression results for smoking status, index of multiple deprivation, geographical region, and year of surgery differed by cluster and are shown in table 3.

Subgroup analysis by sex showed that female participants had care patterns consistent with those observed in the overall cohort, while the clustering among male participants resulted in five distinct clusters with two different care patterns (appendix pp 44–45). Notably, the gabapentinoids and antidepressants predominant pattern observed among female participants was not evident in the male participants, indicating potential sex differences in pharmacological management before total knee replacement.

Subgroup analyses by age group showed clusters broadly consistent with those of the primary analysis (appendix pp 46–47). Similarly, stratification by the calendar periods yielded consistent clustering results (appendix pp 48–50); however, there was a notable increase in the use of gabapentinoids over time.

Sensitivity analyses using alternative dissimilarity measures (longest common subsequences, hamming distance, and dynamic hamming distance) showed clustering patterns consistent with those observed in the primary analysis (appendix pp 51–53). Although the overall structure of the clusters remained stable, slight variations were noted in the distribution of individuals across clusters. In addition, sensitivity analysis using a 12-month time interval produced similar results, further supporting the robustness and stability of the identified care patterns over time

(appendix p 54). Finally, the sensitivity analysis restricted to the pre-COVID-19 period (2007–19) showed care patterns that were consistent with those observed in the primary analysis (appendix p 55).

Discussion

In this study, we analysed health-care pathways among 129 470 patients who underwent primary total knee replacement predominantly for osteoarthritis between 2007 and 2021 in England, UK, focusing on routinely collected care strategies aligned with NICE guidelines. Using sequence analysis, we identified five distinct typologies of care pathways over the 10 years preceding total knee replacement: low health-care use, moderate health-care use, opioids predominant, gabapentinoids and antidepressants predominant, and NSAIDs predominant. These findings highlight substantial variation in care pathways, with cluster membership significantly associated with female sex, age above 75 years, BMI ≥ 25.0 kg/m², multiple comorbidities, non-White ethnicity, and neighbourhood deprivation.

To our knowledge, this is the first study to investigate long-term care pathways preceding total knee replacement using sequence analysis on large-scale routinely collected EHRs. A previous Danish study identified 1143 unique treatment pathways among 2574 patients, but its focus was limited to the uptake of guideline-adherent non-surgical treatments before and 6 months after orthopaedic consultation.¹⁴ A previous US study assessed health-care utilisation over the 2 years preceding total knee replacement, focusing on the frequency of pharmacological treatments,

	Moderate health-care use	Opioids predominant	Gabapentinoids and antidepressants predominant	NSAIDs predominant
Sex				
Male	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Female	1.60 (1.55–1.65)	2.24 (2.16–2.31)	4.18 (3.91–4.47)	1.26 (1.22–1.31)
Age group				
45–54 years	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
55–64 years	0.99 (0.93–1.05)	1.23 (1.14–1.33)	1.18 (1.04–1.33)	1.63 (1.48–1.79)
65–74 years	1.02 (0.96–1.08)	1.91 (1.78–2.06)	1.33 (1.18–1.50)	2.58 (2.36–2.83)
75–84 years	1.13 (1.06–1.20)	2.37 (2.20–2.55)	1.29 (1.14–1.47)	2.92 (2.66–3.20)
≥85 years	1.37 (1.26–1.50)	2.92 (2.64–3.23)	1.41 (1.18–1.68)	3.29 (2.92–3.71)
BMI group				
Underweight (<18.5 kg/m ²)	1.00 (0.87–1.16)	1.19 (1.03–1.37)	1.28 (1.00–1.65)	1.05 (0.89–1.24)
Normal (18.5–24.9 kg/m ²)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Overweight (25.0–29.9 kg/m ²)	1.11 (1.06–1.16)	1.11 (1.06–1.16)	1.09 (1.00–1.18)	1.06 (1.01–1.11)
Obese (≥30.0 kg/m ²)	1.41 (1.35–1.47)	1.60 (1.53–1.68)	1.42 (1.31–1.53)	1.27 (1.21–1.33)
Comorbidities				
0	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
1	1.37 (1.33–1.42)	1.85 (1.78–1.92)	1.81 (1.69–1.94)	1.34 (1.29–1.40)
2	1.66 (1.59–1.73)	2.58 (2.46–2.70)	2.83 (2.62–3.07)	1.58 (1.50–1.66)
≥3	2.14 (2.02–2.26)	3.89 (3.67–4.12)	5.03 (4.60–5.50)	1.68 (1.57–1.80)
Smoking status				
Never smoker	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Ex-smoker	1.18 (1.14–1.22)	1.26 (1.22–1.31)	1.29 (1.21–1.38)	1.05 (1.01–1.09)
Current smoker	1.14 (1.09–1.19)	1.34 (1.27–1.40)	1.28 (1.18–1.39)	0.95 (0.90–1.01)
Ethnicity				
White	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Black	1.87 (1.61–2.17)	1.95 (1.67–2.29)	1.36 (1.04–1.78)	1.83 (1.53–2.20)
South Asian	2.31 (2.08–2.56)	2.84 (2.56–3.15)	2.39 (2.02–2.81)	3.43 (3.08–3.83)
Other or unknown	0.75 (0.68–0.82)	0.74 (0.67–0.82)	0.57 (0.47–0.69)	0.98 (0.88–1.08)
Geographical region				
North East	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
North West	1.03 (0.95–1.11)	0.82 (0.76–0.89)	0.69 (0.60–0.78)	1.06 (0.97–1.17)
Yorkshire & The Humber	1.04 (0.95–1.15)	0.80 (0.72–0.89)	0.98 (0.83–1.16)	1.23 (1.09–1.38)
East Midlands	0.87 (0.77–0.98)	0.74 (0.66–0.84)	0.76 (0.62–0.93)	0.96 (0.83–1.10)
West Midlands	0.96 (0.89–1.04)	0.73 (0.67–0.79)	0.59 (0.51–0.67)	0.92 (0.83–1.01)
East of England	0.92 (0.84–1.02)	0.64 (0.58–0.71)	0.67 (0.56–0.79)	0.97 (0.86–1.09)
London	0.87 (0.80–0.94)	0.53 (0.49–0.58)	0.47 (0.40–0.54)	0.83 (0.75–0.92)
South East	0.90 (0.84–0.98)	0.58 (0.54–0.63)	0.70 (0.61–0.79)	0.85 (0.77–0.94)
South West	0.95 (0.87–1.02)	0.60 (0.55–0.65)	0.68 (0.59–0.78)	0.98 (0.88–1.07)
Index of Multiple Deprivation				
1 (least deprived)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
2	1.13 (1.06–1.20)	1.19 (1.11–1.27)	1.21 (1.07–1.37)	1.06 (0.99–1.14)
3	1.14 (1.07–1.21)	1.37 (1.28–1.47)	1.30 (1.15–1.47)	1.07 (1.00–1.14)
4	1.22 (1.15–1.29)	1.42 (1.33–1.53)	1.36 (1.21–1.54)	1.09 (1.01–1.16)
5	1.25 (1.17–1.33)	1.61 (1.50–1.73)	1.51 (1.33–1.71)	1.18 (1.10–1.27)
6	1.32 (1.24–1.41)	1.83 (1.71–1.97)	1.72 (1.52–1.94)	1.20 (1.11–1.29)
7	1.43 (1.34–1.52)	2.16 (2.01–2.32)	1.76 (1.55–1.99)	1.20 (1.11–1.29)
8	1.50 (1.40–1.60)	2.51 (2.34–2.71)	1.87 (1.64–2.14)	1.28 (1.18–1.39)
9	1.64 (1.53–1.76)	3.16 (2.93–3.41)	2.50 (2.20–2.85)	1.35 (1.24–1.47)
10 (most deprived)	1.87 (1.73–2.02)	3.86 (3.57–4.18)	3.50 (3.07–3.99)	1.48 (1.35–1.63)
Index year				
2007	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
2008	0.92 (0.85–1.00)	0.89 (0.82–0.96)	1.10 (0.91–1.32)	0.97 (0.89–1.06)

(Table 3 continues on next page)

	Moderate health-care use	Opioids predominant	Gabapentinoids and antidepressants predominant	NSAIDs predominant
(Continued from previous page)				
2009	0.94 (0.87–1.03)	0.88 (0.81–0.96)	1.05 (0.87–1.26)	0.93 (0.85–1.02)
2010	0.94 (0.87–1.02)	0.86 (0.79–0.93)	1.20 (1.00–1.44)	0.89 (0.81–0.97)
2011	1.02 (0.94–1.10)	0.84 (0.77–0.91)	1.40 (1.17–1.66)	0.82 (0.75–0.90)
2012	1.01 (0.93–1.10)	0.78 (0.72–0.85)	1.22 (1.02–1.45)	0.77 (0.71–0.85)
2013	0.99 (0.91–1.07)	0.82 (0.75–0.89)	1.56 (1.31–1.85)	0.80 (0.73–0.88)
2014	0.93 (0.86–1.01)	0.68 (0.63–0.74)	1.39 (1.17–1.64)	0.68 (0.62–0.74)
2015	0.92 (0.85–1.00)	0.63 (0.58–0.68)	1.44 (1.22–1.70)	0.65 (0.59–0.71)
2016	0.84 (0.78–0.91)	0.60 (0.55–0.65)	1.50 (1.27–1.76)	0.56 (0.52–0.62)
2017	0.87 (0.80–0.94)	0.58 (0.54–0.63)	1.54 (1.31–1.82)	0.53 (0.49–0.58)
2018	0.87 (0.80–0.93)	0.56 (0.51–0.61)	1.70 (1.44–2.00)	0.53 (0.48–0.58)
2019	0.83 (0.77–0.90)	0.50 (0.46–0.54)	1.81 (1.54–2.12)	0.49 (0.45–0.54)
2020	0.74 (0.67–0.81)	0.45 (0.41–0.50)	1.46 (1.21–1.76)	0.42 (0.37–0.47)
2021	0.73 (0.61–0.89)	0.45 (0.36–0.56)	2.12 (1.55–2.90)	0.38 (0.29–0.49)

Data are odds ratios (95% CI). NSAIDs=non-steroidal anti-inflammatory drugs.

Table 3: Multinomial logistic regression model for association between patients' characteristics and membership to different clusters, relative to the low health-care use cluster

but did not differentiate between patient-level trajectories or capture broader patterns of care.⁸ A previous Dutch study examined management strategies during the 3 years following osteoarthritis diagnosis, describing the most frequently recorded care modalities but not longitudinal care patterns.¹² Similarly, a Swedish study compared treatment strategies over a 10-year period between patients with and without total knee replacement, but it did not explore detailed care sequences or pathways over time.¹³ In contrast, our study applies a data-driven clustering approach to longitudinal EHR data, offering novel insights into real-world health-care utilisation in the 10 years preceding total knee replacement by capturing key care strategies.

Our findings indicate that 37.8% (48 922 of 129 470) of patients were classified in the low health-care use cluster and progressed to total knee replacement with relatively few previous interventions, potentially missing key components of guideline-recommended care, such as physiotherapy or stepwise analgesic management. In contrast, 23.9% (30 947 of 129 470) of patients were classified in the moderate health-care use cluster and followed a more prolonged care pathway, engaging with several non-surgical interventions before undergoing surgery. The remaining patient groups were primarily managed with pharmacological treatments, including NSAIDs, opioids, gabapentinoids, and antidepressants, many of which were initiated early and continued over long periods. These patterns suggest that a substantial proportion of patients might be receiving care that diverges from current clinical guidelines.

Furthermore, in multinomial logistic regression analysis we found that female patients were more likely to receive analgesics than male patients, indicating potential sex differences in pain management and prescribing patterns. These findings align with earlier evidence showing that older patients, especially women, often present with greater

mobility restriction and more advanced osteoarthritis before knee replacement.²⁶ Older individuals also had higher odds of belonging to groups with long-term use of NSAIDs and opioids than did younger individuals, possibly reflecting a high burden of comorbidities and more frequent interactions with health-care services. Patients from South Asian and Black backgrounds were more likely to be in analgesic-predominant clusters compared with patients of White ethnicity, suggesting disparities in care and greater reliance on pharmacological rather than guideline-recommended treatments.²⁷ Regional variation was evident, with patients in Yorkshire & the Humber more likely to be in the NSAIDs predominant group than patients in the North East region. Compared with the North East region, all other regions had lower odds of belonging to the clusters characterised by extensive use of opioids, gabapentinoids, and antidepressants, reflecting variation in prescribing behaviours, health-care access, and local clinical practices.²⁸

Patients from the most deprived areas were more likely to be in the opioids predominant and gabapentinoids and antidepressants predominant clusters than patients in least deprived areas, consistent with previous evidence describing socioeconomic differences in health-care utilisation.¹⁸ NHS care is free at the point of delivery, but means-tested prescription charges might partly explain socioeconomic differences in medication use. NSAIDs and opioids use declined over time, reflecting evolving guidelines that shifted from paracetamol and NSAIDs as first-line treatments towards greater emphasis on non-pharmacological management and concerns about long-term opioid safety. Conversely, use of gabapentinoids and antidepressants increased from 2010 onwards, suggesting rising reliance on these medications for chronic pain management.^{29,30}

Referrals for imaging, physiotherapy, specialist care, corticosteroid injections, and knee surgeries were broadly similar across all clusters. Following analgesic prescriptions, referrals for imaging and specialist care represented the next most common care strategies. Physiotherapy is a core recommended treatment for the management of osteoarthritis; however, this study found evidence of underutilisation, suggesting that many patients did not receive physiotherapy referrals in primary care. It is important to note that physiotherapy codes in EHRs reflect referral occurrence rather than treatment continuity and therefore might not capture ongoing or repeated physiotherapy sessions. Moreover, self-referrals and private physiotherapy are not routinely recorded in primary care systems, which might contribute to underestimation of true physiotherapy engagement.

Our findings highlight substantial inequalities in pre-total knee replacement care by socioeconomic status, ethnicity, sex, and geographical region, with higher odds of opioids, gabapentinoids, and antidepressants use in the most deprived areas suggesting greater reliance on pharmacological rather than guideline-recommended non-surgical management. These disparities raise concerns about the equity and quality of osteoarthritis care delivered in primary care. Policy efforts should prioritise improving access to core non-surgical treatments, particularly in deprived or underserved areas, and national initiatives to standardise best practice, reduce prescribing variation, and early-stage conservative care. Educational initiatives could enhance clinicians' confidence in delivering evidence-based non-pharmacological care and in guiding patients through appropriate treatment options. Future research in primary care settings should evaluate targeted interventions to improve adherence to guideline-recommended non-surgical management, assess their impact on pain, function, and time to surgery, and explore how best to address inequalities identified in our study.

The strengths of this study include the use of a large, population-based cohort with a 10-year retrospective follow-up using routinely collected EHRs, providing a robust picture of real-world care pathways before total knee replacement. Moreover, the application of sequence analysis offers a novel, data-driven method to identify and characterise care trajectories, capturing both the order and timing of interventions.

This study also has several limitations. First, the requirement for a minimum of 10 years of registration before the index total knee replacement might have introduced selection bias. However, the distribution of patients with shorter registration histories was similar to that of those included (appendix pp 56–57), suggesting that any resulting bias is likely to be minimal. Second, treatment use was measured as a binary indicator (yes or no) within 6-month intervals, which does not capture differences in frequency, dose, or duration. Recording only one care event per interval inevitably results in a loss of detail, masking

multiple treatments and the true order of transitions within the same period. Third, we did not examine care from diagnosis to total knee replacement, as osteoarthritis diagnoses are inconsistently coded in EHRs. Instead, we used a fixed 10-year pre-surgery period for consistent follow-up. Consequently, cluster differences might partly reflect variation in osteoarthritis duration before index year. Fourth, health-care use not directly attributable to knee osteoarthritis might introduce some misclassification. However, this approach provides a more comprehensive view of patients' overall health-care engagement and multimorbidity, which are integral to understanding real-world patterns of care preceding total knee replacement. Fifth, analgesic prescribing might be underestimated because some pain medications are available over the counter and therefore not recorded in primary care data. However, as most patients in this cohort were older than 60 years and thus eligible for free prescriptions in the UK, only a small proportion of analgesic use would have occurred outside general practice prescribing records. Sixth, for BMI, smoking status, and drinking status, we used the most recent value recorded within the 10 years preceding the index year, with mean times of 2.6 years (SD 3.8), 1.4 years (1.7), and 3.6 years (2.8), respectively. The lack of timeliness in these measurements is a limitation, as these values might not accurately reflect patients' characteristics close to surgery. Seventh, CPRD does not capture activity in private primary care; therefore, our findings might not fully represent care pathways for patients treated in the private sector. Finally, the very large sample size might have increased statistical power and the likelihood of detecting significant interactions, including effects of small magnitude that might have limited clinical relevance; hence caution should be used when interpreting the data.

Our research approach, using nationally representative primary and secondary care data and a transferable sequence analysis framework, enhances the applicability of our findings while acknowledging that variations in coding practices and service availability might influence relevance across diverse primary care settings.

Future research should investigate how these distinct care pathways affect patient outcomes, such as pain and function, and assess their cost-effectiveness to guide more equitable and efficient primary care management of osteoarthritis. Building on these directions, this study provides novel evidence by identifying five long-term care pathways preceding total knee replacement using sequence analysis of routinely collected EHRs.

Contributors

SM was responsible for analysis and drafting of the manuscript. EP, GP, and DY provided their input and guidance at each step of the study. EP, GP, and DY were responsible for revising the manuscript critically and for important intellectual content. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. SM and DY directly accessed and verified the underlying data reported in the manuscript. All authors read and approved the final manuscript.

Declaration of interests

We declare no competing interests.

Data sharing

Access to anonymised patient data from Clinical Practice Research Datalink (CPRD) is subject to a data sharing agreement containing detailed terms and conditions of use following protocol approval from the Medicines and Healthcare products Regulatory Agency Independent Scientific Advisory Committee. This study-specific analysable dataset is therefore not publicly available. Details about Independent Scientific Advisory Committee applications and data costs are available on the CPRD website (<https://www.cprd.com/>).

Acknowledgments

We thank James Bailey for his assistance in retrieving data from the Clinical Practice Research Datalink (CPRD) database. SM was supported by PhD studentship from the Faculty of Medicine and Health Sciences at Keele University. DY was partly funded by the National Natural Science Foundation of China (grant number 82574190), Henan Provincial Department of Science and Technology under the Henan Key Research and Development Special Project (251111313800), and matched funding awarded to the National Institute for Health and Care Research Applied Research Collaboration West Midlands.

References

- Lamova SN, Müller-Ladner U. Editorial: Osteoarthritis—current insights in pathogenesis, diagnosis and treatment. *Curr Rheumatol Rev* 2018; **14**: 91–97.
- Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2023 (GBD 2023) results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME). 2023. <https://vizhub.healthdata.org/gbd-results/> (accessed Jan 7, 2026).
- Versus Arthritis. The state of musculoskeletal health 2024. 2024. <https://versusarthritis.org/media/embkyapu/the-state-of-musculoskeletal-health-2024.pdf> (accessed Aug 2, 2024).
- Swain S, Sarmanova A, Mallen C, et al. Trends in incidence and prevalence of osteoarthritis in the United Kingdom: findings from the Clinical Practice Research Datalink (CPRD). *Osteoarthritis Cartilage* 2020; **28**: 792–801.
- Price AJ, Alvand A, Troelsen A, et al. Knee replacement. *Lancet* 2018; **392**: 1672–82.
- National Joint Registry. 22 NJR Annual Report. 2025. <https://reports.njrcentre.org.uk/> (accessed Nov 19, 2025).
- National Institute for Health and Care Excellence. Osteoarthritis in over 16s: diagnosis and management. NICE guideline [NG226]. 2022; published online Oct 19. <https://www.nice.org.uk/guidance/ng226> (accessed Jan 7, 2026).
- Berger A, Bozic K, Stacey B, Edelsberg J, Sadosky A, Oster G. Patterns of pharmacotherapy and health care utilization and costs prior to total hip or total knee replacement in patients with osteoarthritis. *Arthritis Rheum* 2011; **63**: 2268–75.
- Taqi A, Gran S, Knaggs RD. Patterns of analgesic utilisation among people with knee osteoarthritis: a cohort study using UK primary care data. *J Pharm Policy Pract* 2025; **18**: 2455067.
- Park H-R, Cho S-K, Im SG, et al. Treatment patterns of knee osteoarthritis patients in Korea. *Korean J Intern Med* 2019; **34**: 1145–53.
- Dysart S, Utkina K, Stong L, et al. Insights from real-world analysis of treatment patterns in patients with newly diagnosed knee osteoarthritis. *Am Health Drug Benefits* 2021; **14**: 56–62.
- Arslan IG, van Berkel AC, Damen J, et al. Patterns of knee osteoarthritis management in general practice: a retrospective cohort study using electronic health records. *BMC Primary Care* 2024; **25**: 2.
- Dell'Isola A, Hellberg C, Turkiewicz A, et al. Use of non-surgical treatments on the journey to knee replacement in patients with knee osteoarthritis: a 10-year population-based case-control study. *RMD Open* 2023; **9**: e003422.
- Bruhn SM, Skou ST, Harris LK, et al. Usage of guideline-adherent core treatments for knee osteoarthritis before and after consulting an orthopaedic surgeon: a prospective cohort study. *Osteoarthritis Cartilage* 2023; **5**: 100411.
- Liao TF, Bolano D, Brzinsky-Fay C, et al. Sequence analysis: Its past, present, and future. *Soc Sci Res* 2022; **107**: 102772.
- Hayward CJ, Batty JA, Westhead DR, et al. Disease trajectories following myocardial infarction: insights from process mining of 145 million hospitalisation episodes. *EBioMedicine* 2023; **96**: 104792.
- Mathew S, Peat G, Parry E, Sokhal BS, Yu D. Applying sequence analysis to uncover 'real-world' clinical pathways from routinely collected data: a systematic review. *J Clin Epidemiol* 2024; **166**: 111226.
- Mathew S, Peat G, Parry E, et al. Sequence analysis to phenotype healthcare patterns in adults with musculoskeletal conditions using primary care electronic health records. *Arthritis Care Res (Hoboken)* 2025; **7**: 906–15.
- Henri S, Herrera R, Vanasse A, Forget A, Blais L. Trajectories of care in patients with chronic obstructive pulmonary disease: A sequence analysis. *Can J Respir Crit Care, Sleep Med* 2022; **6**: 237–47.
- CPRD, Clinical Practice Research Datalink. <https://www.cprd.com/> (accessed Jan 27, 2025).
- Ministry of Housing, Communities & Local Government. The English Indices of Deprivation 2019 (IoD2019). 2019. https://assets.publishing.service.gov.uk/media/5d8e26f6ed915d5570c6cc55/IoD2019_Statistical_Release.pdf (accessed Jan 7, 2026).
- CPRD, Release Notes: CPRD Aurum June 2024. 2024. <https://www.cprd.com/sites/default/files/2024-06/2024-06%20CPRD%20Aurum%20Release%20Notes.pdf> (accessed Feb 4, 2025).
- Benchimol EI, Smeeth L, Guttman A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med* 2015; **12**: e1001885.
- Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; **40**: 373–83.
- Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* 2011; **20**: 40–49.
- Jämsen E, Jäntti P, Puolakka T, Eskelinen A. Primary knee replacement for primary osteoarthritis in the aged: gender differences in epidemiology and preoperative clinical state. *Aging Clin Exp Res* 2012; **24**: 691–98.
- Hayanga B, Stafford M, Bécarea L. Ethnic inequalities in healthcare use and care quality among people with multiple long-term health conditions living in the United Kingdom: A systematic review and narrative synthesis. *Int J Environ Res Public Health* 2021; **18**: 12599.
- Mordecai L, Reynolds C, Donaldson LJ, de C Williams AC. Patterns of regional variation of opioid prescribing in primary care in England: a retrospective observational study. *Br J Gen Pract* 2018; **68**: e225–33.
- Appleyard T, Ashworth J, Bedson J, Yu D, Peat G. Trends in gabapentinoid prescribing in patients with osteoarthritis: a United Kingdom national cohort study in primary care. *Osteoarthritis Cartilage* 2019; **27**: 1437–44.
- Ferreira GE, McLachlan AJ, Lin C-WC, et al. Efficacy and safety of antidepressants for the treatment of back pain and osteoarthritis: systematic review and meta-analysis. *BMJ* 2021; **372**: m4825.