Title: Lifetime risk and genetic predisposition to post-traumatic OA of the knee in the UK Biobank

Hollis B (PhD), Chatzigeorgiou C (PhD), Southam L (BSc), Hatzikotoulas, K (PhD), Kluzek S (MD, PhD), Williams A (MD), the Genetics of Osteoarthritis Consortium, Zeggini E (PhD), Jostins-Dean L (PhD)*, Watt FE (MD, PhD)*±

*Joint last authors

±Full name, postal address and e-mail of corresponding author:

Dr Fiona Elizabeth Watt, MD, PhD

f.watt@imperial.ac.uk

- *Centre for Inflammatory Disease, 9th Floor, Commonwealth Building, Department of Immunology and Inflammation, Hammersmith Campus, Imperial College London, London, Du Cane Road, London, W12 0NN. United Kingdom*
- *Centre for Osteoarthritis Pathogenesis Versus Arthritis, Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, United Kingdom*

Dr Benjamin Hollis, PhD

Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, United Kingdom

benjamin.hollis@kennedy.ox.ac.uk

Associate Professor Luke Jostins-Dean, PhD

Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, United Kingdom

luke.jostins@kennedy.ox.ac.uk

Ms Lorraine Southam, BSc

Institute of Translational Genomics, Helmholtz, Munich, Germany

lorraine.southam@helmholtz-muenchen.de

Dr Konstantinos Hatzikotoulas, PhD

Institute of Translational Genomics, Helmholtz, Munich, Germany

hatzikotoulas@helmholtz-muenchen.de

Dr Charikleia Chatzigeorgiou, PhD

Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, United Kingdom

charikleia.chatzigeorgiou@mssm.edu

Associate Professor Stefan Kluzek, MD, PhD

Centre for Sport, Exercise and OA, University of Nottingham, Nottingham, United Kingdom

Stefan.Kluzek@nottingham.ac.uk

Mr Andy Williams, MD

Fortius Clinic, London, United Kingdom

andy.williams@fortiusclinic.com

Professor Eleftheria Zeggini, PhD

Institute of Translational Genomics, Helmholtz, Munich, Germany

eleftheria.zeggini@helmholtz-muenchen.de

Genetics of Osteoarthritis Consortium

(see Notes section for membership)

Running title: Genetics of PTOA in UK Biobank

Abstract

Objective: Acute knee injury is associated with post-traumatic OA (PTOA). Very little is known about the genome-wide associations of PTOA when compared with idiopathic OA (iOA). Our objective was to describe the development of knee OA after knee injury and its genetic associations in UK Biobank (UKB).

Design: Clinically significant structural knee injuries in those <=50 years were identified from electronic health record and self-reported data in 502,409 UKB participants. Time-to-first knee OA code was compared in injured cases and age-/sex-matched non-injured controls using Cox Proportional Hazards models. A time-to-OA genome-wide association study (GWAS) sought evidence for PTOA risk variants 6 months-20 years following injury. Evidence for associations of two iOA polygenic risk scores (PRS) was sought.

Results: Of 4233 knee injury cases, 1896 (44.8%) were female (mean age at injury 34.1 years [SD10.4]). Over a median of 30.2 (IQR19.5-45.4) years, 1096 (25.9%) of injured cases developed knee OA. The overall hazards ratio (HR) for knee OA after injury was 1.81[1.70,1.93],P=$8.9x10^{-74}$. Female sex and increasing age at injury were associated with knee OA following injury (HR1.15[1.02,1.30];1.07[1,07,1.07] respectively). OA risk was highest in the first 5 years after injury (HR3.26[2.67,3.98]), persisting for 40 years. In 3074 knee injury cases included in the time-to-OA GWAS, no variants reached genome-wide significance. iOA PRS was not associated with time-to-OA (HR 0.43[0.02,8.41]).

Conclusions: Increasing age at injury and female sex appear to be associated with future development of PTOA in UKB, the risk of which was greatest in the 5 years after injury. Further international efforts towards a better-powered meta-analysis will definitively elucidate genetic similarities and differences of PTOA and iOA.

Keywords: osteoarthritis; knee; injury; post-traumatic; genetic; risk

Introduction

Whilst osteoarthritis (OA) pathogenesis is incompletely understood, its risk factors are now well-delineated(1). Acute joint injury, such as intra-articular ligament rupture, fracture, or meniscal tear all increase the risk of OA considerably(2-4). OA following a joint injury (so-called 'post-traumatic osteoarthritis', PTOA) is difficult to predict, but reportedly develops in 30-50% of people over 5-10 years(5, 6). Considering the knee, commonly affected by trauma and by OA, such PTOA is thought to constitute approximately 12% of all knee OA(7). The excess risk of knee OA after injury is reportedly 4-15 fold, varying depending on the type of injury(4, 8-10). This is notable as individuals are typically young at the time of knee injury, developing PTOA in young/mid-life, leading to life-long knee symptoms and impaired quality-of-life, clinical management difficulties (considering joint replacement is typically at older ages) and related substantial healthcare/societal costs.

Beyond this impact, many wider questions remain. How to establish an individual's risk of future OA at the time of injury, and how to mitigate this risk by secondary prevention is currently unclear(11). To answer this question, one must consider the risk of developing idiopathic knee OA (iOA, without history of knee injury) and what factors are associated with PTOA development.

A number of injury-related, imaging and clinical factors such as concomitant injuries, effusion-haemarthrosis at the time of injury and re-injury have all been associated with either worse clinical outcomes or future structural PTOA in longitudinal cohorts(12-15). However, little is known about how generalisable these features are, or how factors associated with iOA (such as increasing age, sex and increased body mass index (BMI) or genetic factors) influence the risk of PTOA. The dynamics of excess OA risk posed by a joint injury are also unknown, such as when this becomes indistinguishable from background iOA risk. This requires examination of population level data. Epidemiological knowledge of PTOA is important to its genetic investigation, the development of risk prediction models, as well as defining any period of modifiable risk with a view to preventive interventions. However, determining clinically significant acute knee joint injuries and PTOA cases from electronic healthcare records (EHR) is not straightforward(4, 16).

Genetic variation is known to be an important aetiological factor for OA. There are now very large genome wide association studies (GWAS) which provide high quality evidence for up to 100 single nucleotide polymorphisms (SNPs) contributing to risk of iOA at any site, with some being knee-specific(17, 18). However, whether genetic variation associated with iOA contributes to PTOA risk following a joint injury is not known. One 2013 study considered the known OA loci at that time, ascertaining cases by retrospective report of knee injury, finding some association with existing OA genetic variants(19). No further studies with more specific case ascertainment, in larger numbers or with more recent genetic knowledge have been carried out since to our knowledge. The ability to generate so-called polygenic risk scores (PRS), which look at cumulative genetic risk across many genome-wide loci could be clinically important in predicting PTOA.

A final important question that genetic studies could help to answer is how similar or different iOA and PTOA truly are, by defining which risk genes and pathogenic pathways are shared or distinct (3). . Pre-clinical and clinical studies seeking to identify novel interventions in PTOA which translate to OA in general would benefit(11).

Our objectives were to describe OA development over time after acute knee injury, comparing to those without injury, including the effect of relevant covariates; to establish any maximum time after injury that should be considered a 'window' for PTOA development for genetic studies; to test if there were genome-wide associations with developing OA after joint injury; and to test if established iOA genes were collectively associated with PTOA development. To do this we developed a digital classifier for clinically significant knee joint injuries and PTOA in EHR to identify cases in UK Biobank (UKB), with a view to testing the need/feasibility for scaling up. Our overarching hypothesis was that genetic variation contributes to the risk of PTOA and that these genetic factors may be the same as or different from those for iOA.

Method

Source of study participants and setting

UK Biobank (UKB) is a large-scale biomedical database combining medical and genetic data from over 500,000 participants. The anonymised UKB dataset used here includes imputed genome-wide genetic data, linked Hospital Episode Statistics data (ICD10 diagnosis codes and OPCS-4 procedure codes), primary care diagnosis and procedure codes (Read versions 2 and 3, abbreviated to Read2 and Read3), and self-reported (SR) diagnoses and operations. UKB participants gave consent for use of their data at recruitment. This project received UKB approvals in June 2020 (project number 52507).

In addition, GWAS summary statistic data were provided for 80,697 cases (non-overlapping with UKB) by the GO Consortium to create a knee-specific PRS)(17).

Eligibility criteria for knee injury

Inclusion criteria for those with knee injury were: 1) all UKB participants with ongoing consent for data analysis; and 2) evidence of a knee injury of known timing, with all of: a) a clinically significant injury to an internal structure of the joint (plausibly sufficient to increase the risk of OA *a priori*); b) to the location of the knee joint; c) due to acute trauma; and d) occurring between the ages 18-50 (Supplementary Table 1). Exclusion criteria were applied to remove individuals with confounding conditions, such as pre-existing OA or chronic knee instability(Supplementary Methods).

Definition of an acute knee injury was by a combination of codes available in one or more associated Hospital or Primary care EHR episode code sets and/or SR codes, which individually or jointly fulfilled inclusion criteria 2a-c within a three month 'injury window' (which began at first evidence of knee injury, using the earliest record)(18, 20). Code lists were established through manual curation, including review by an expert panel, and were further checked using cross-mapping tools (Supplementary Methods). Sensitivity and specificity of ICD-10 code lists were validated using coding data from existing cohorts of joint injury (to assess sensitivity), degenerative joint injury (to assess

specificity) and knee OA (to assess outcome sensitivity) (see Supplementary Methods, Supplementary Table 2).

Definition of knee OA outcome in time-to-event analyses

Knee OA was defined as diagnostic or procedure codes compatible with definite, probable or possible knee OA and/or total knee replacement/revision, consistent with knee OA (Supplementary methods). The first knee OA code was used. This needed to be six months or more after the knee injury.

Study populations

Sizes of study populations following application of the inclusion criteria, then the exclusion criteria are reported in Figure 1 (additional exclusion criteria for genetic studies, how summary statistics described is below, in Supplementary Methods). Different sources of identification of eligible individuals were recorded (Supplementary Figure 1). Based on injury and OA codes, we defined those with: knee injury ('knee injury'); no knee injury and no knee OA ('non injured controls'). For some analyses, we also defined knee OA, excluding cases of prior knee injury ('iOA').

Matched cohort time-to-event analysis

The relationship between knee injury and knee OA was assessed in a matched cohort time-to-event survival analysis. Knee injury individuals were matched on sex and date of birth (to within 3 months) with non-injured controls in a 1:10 ratio. The same exclusion criteria as applied to the knee injury group were applied to controls (those who died prior to the date of injury of their matched case were removed). Controls were assigned a 'pseudo-date' of injury, corresponding to the date of injury of their matched case. Data were censored at death or at latest UKB/EHR record (31/03/2021 at time of analysis - the most recently released UKB dataset).

The effect of joint injury on time-to-OA outcome or censor was assessed using a Cox Proportional Hazards (PH) model, including age at injury, sex and injury status as covariates aswell as the interactions of age and sex with injury. The primary analysis was conducted using those with knee injury ascertained from all data sources combined. Pre-defined sensitivity analyses then examined

cases ascertained by SR or EHR sources only. A further Cox regression was performed to establish how the relative risk of PTOA varied with time after joint injury, by estimating the hazard ratio for joint injury across pre-defined strata of time since injury of 0-4, 5-9, 10-19, 20-29, 30-39 and 40+ years (specifically, we fitted a model with an interaction between joint injury and time strata as a time-varying categorical variable). As it was previously unknown how long excess risk of PTOA remained after a knee injury, we used these data to select an upper time threshold after injury to apply in our subsequent analyses (above which OA would be unlikely to be related to the injury).

Genome-wide association studies

Genome-wide association studies (GWAS) were employed to assess genetic associations for traits related to knee injury and subsequent PTOA. For all studies, standard quality control (QC) measures were applied, as described in Supplementary Methods.

The primary GWAS was a time-to-event analysis (where the event is first evidence of OA), using a Cox PH model implemented in SPACox(21), with age at injury, sex, genotyping chip and first 10 genetic principal components (PCs) received from UKB included as covariates. Data were censored at death, latest UKB/EHR record or after 20 years (as the upper limit of our 'observation period' established in initial analysis, described above and in results). Two further GWAS to support this work were carried out, a case-control analysis of iOA vs. uninjured controls, and a case-control analysis of knee injury vs. uninjured controls, both described in Supplementary Methods.

For all genome-wide analyses, standard thresholds of significance were predefined: genome-wide significance was $P < 5x10^{-8}$ and a suggestive level of significance $P < 1\times10^{-6}$. Suggestively significant associations were queried using the Open Targets Genetics Portal to identify candidate genes and other phenotypes associated with variants(22).

iOA Polygenic Risk Scores

PRS for iOA were generated using two datasets: 1) the summary statistics for a knee OA GWAS within the GO consortium dataset, excluding UKB(17), and 2) the lead variants for our UKB iOA case-

control GWAS described above (see Supplementary Methods). Associations between iOA PRSs and knee injury were carried out using Cox PH models.

Codes with missing dates were excluded from the analysis. All other data were included.

Power calculations

We performed two sets of power analyses, the first carried out at the study onset to guide the analysis, and the second carried out after results were gathered, to guide future studies. Both are outlined in Supplementary Methods.

Robustness analyses

Post-hoc, we checked if any differences noted in non-matched baseline characteristics between the two cohort groups affected our findings, by adding them as additional covariates in the matched cohort time-to-event analysis.

Results

Characteristics of knee injury cases in UKB

Following application of eligibility criteria to a total of 502,465 UKB participants, 4,233 individuals were identified as knee injury cases from the various sources (Figure 1, Supplementary Figure 1). There were approximately similar numbers of males and females experiencing knee injury (1,896, 44.8% female). More males were ascertained by EHR codes (1,493, 64.7%) but a higher proportion of females by SR (Table 1). Mean age at injury was 34.1 years (SD 10.4). This was higher in EHR-identified cases (39.6 years[SD 8.4]) than in self-reported cases (28.4 years[SD 9.0]), as would be expected from differences in sampling (EHR data is only reliably present within the last 25 years).

The median follow-up from injury to end of observation was 30.2 years for all cases (IQR 19.5-45.4), and was again longer for self-reports (34.0 years;IQR 23.3-42.3) than EHR cases (21.7 years;IQR 15.2-29.2). Overall, a total of 1,096 (25.9%) injured individuals developed knee OA within the observation period, with a median time from injury to OA of 21.7 years (IQR 10.5-34.7). The proportion of knee injury cases developing OA was higher in EHR cases (24.3%) than SR cases (12.9%). The median time-to-OA was longer in SR cases (29.0 years, IQR 19.3-37.4) than in EHR-only (13.4 years, IQR 7.0-21.7).

The relationship between knee injury and knee OA

A Cox PH model was used to assess the relationship between knee injury and knee OA in the matched cohort study (Figure 2, Table 2, Supplementary Table 3). As expected, knee injury increased the risk of future knee OA (HR 1.81, 95%CI 1.75-1.87). Female sex and increasing age at injury increased the risk for subsequent OA (HR 1.41; 95%CI 1.35,1.48 and 1.07 per year; 95%CI 1.07,1.07 respectively). Of note, there was evidence for an interaction between sex and injury, with a far weaker, though still significant, effect of female sex in those developing OA after an injury than in those without injury (HR 1.15; 95%CI 1.02,1.30, HR for injury-by-sex interaction, 0.77; 95%CI 0.67,0.87, Supplementary Table 4). All these associations held in sensitivity analyses by source of identification. For those with knee injury ascertained by EHR-only, the effect of knee injury on OA was greater than in all source data (HR 3.06, 95%CI 2.97,3.16) and present but weaker in SR-only

cases (HR 1.18; 95%CI 1.05,1.30). Females were at greater risk from OA in SR-only source data (HR 1.68; 95%CI 1.54,1.83) than in EHR-only cases (HR 1.32; 95%CI 1.22,1.43). Knee injured individuals were noted to have similar physical activity and BMI to the matched control cohort, but tended to be more affluent (by two socioeconomic measures) and more white (Supplementary Table 3). Inclusion of these additional factors in the model as a post hoc robustness analysis did not substantially affect our findings (Supplementary Table 5). This analysis did, however, reveal another interesting difference between PTOA and iOA: within the uninjured matched cohort, lower household income and higher deprivation were associated with a higher rates of progression to iOA ($P < 2.2 \times 10^{-16}$ and 4.6×10^{-4} respectively), but among injured individuals the effect of socioeconomic status on progression to PTOA was much smaller and non-significant (P = 0.0603 and 0.318 respectively, interaction P value = 0.0016).

Defining a time threshold for considering PTOA development after injury

A second Cox PH model was used to examine the time-dependent risk for PTOA following knee injury. Within the first 5 years, the risk of future OA among injured individuals compared with the control group was at its highest (HR 3.26; 95%CI 2.67,3.98) (Table 3). The risk (considering all data sources) then decreased gradually over time, though it remained elevated after 40+ years following knee injury (HR 1.39; 95%CI 1.18,1.63). There was no clear 'dividing line' (even after 40 years, a small number of OA cases could be due to knee injury decades earlier based on the excess cases in the knee injured group at this time, and were therefore, in some sense, 'PTOA'). However after 20 years, the overall HR dropped below 2. We also considered that at a median age of 54 (after 20 years' follow up), there were likely to be increasingly proportionately more iOA (i.e. OA that would have occurred even without joint injury) than PTOA cases (i.e. excess cases attributed to joint injury): prevalence rates for OA of at least 10% are expected in this age group, then rising sharply with age(23). A threshold of 20 years after injury was therefore selected as being reasonable clinically and statistically to examine the genetic associations of excess OA risk for this study. Of note, in sensitivity analyses by source of case ascertainment, the trend was consistent in EHR-only cases, though with

relatively higher hazard ratios for each time stratrum (HR 6.53, 95%CI 5.14,8.29 at <5 years) and this excess risk was present but gradually less detectable for 10 - 40 years from injury but not after (40+ years following injury, HR 1.49, 95%CI 0.93,2.39). For SR data, an effect was seen between 5 and 10 years following injury (HR 1.64, 95%CI 1.05,2.58), with no evidence of a difference in risk in other strata.

Time-to-event GWAS for OA in knee injured participants

Following genetic QC exclusions, 3,074 participants with knee injury were included in the GWAS for time-to-OA (Figure 1). OA outcomes were censored at 20 years after injury as evidenced above, giving 377 PTOA cases (12.3%). No variants reached genome-wide significance. One locus met suggestive significance, a variant on chromosome 11 (rs2412204, P=1.17 x 10^{-7}, MAF=0.13, Figure 3, Supplementary Table 6). This variant is reported in the Open Targets database as associated with "Acquired spondylolisthesis" in UKB(P=0.0019)(22, 24). A post-hoc sensitivity analysis using only cases identified by EHR codes did not reveal any additional genome-wide significant hits, though the sample size was low (Supplementary Figure 2).

Power calculations using our UKB data indicated that a sample size between 6,000 and 9,500 injury cases would be required to detect variants in a full time-to-event GWAS with moderate hazard ratios (HR 1.2-1.25; assumed MAF=0.3), with 80% power at genome-wide significance (P<5 x 10^{-8} , Figure 4A). To compare PTOA and iOA associations in a GWAS, odds ratios of >1.2 and MAF of 0.3 could be detected at genome-wide significance with a sample size of 2,750 PTOA cases (assuming an excess of iOA cases, Figure 4B – note that these same power calculations apply to a GWAS of knee injured cases with excess numbers of uninjured controls). Because this study was underpowered to detect PTOA variants in a full GWAS, whether the genetic component of PTOA was driven by the known genetic associations of iOA was investigated by applying PRS for iOA in Cox PH models among injured participants. A PRS was constructed from 54 lead variants for knee-specific OA in GO (excluding UKB participants; Supplementary Methods & Supplementary Table 7). This showed no association with time-to-OA when added as an additional covariate in our analyses (HR = 0.43; 95%CI 0.02,8.41)

(Supplementary Table 8). These findings held in sensitivity analyses by source of case identification (EHR-only and SR-only). A second PRS, constructed from the 64 lead variants in a GWAS for iOA among UKB participants, suggested possible evidence for association with time-to-OA (HR 18.87; 95%CI 1.71,208.70) (Supplementary Methods, Supplementary Figure 3 and Supplementary Tables 9, 10). However, the confidence intervals were very large and these findings were not consistent in sensitivity analyses in EHR-only or SR-only data sources (Supplementary Table 10). Of note, rs2412204, the SNP suggestively associated with time-to-OA after knee injury, was not associated with knee iOA in either of these datasets (GO, P=0.059; UKB, P=0.93). Furthermore, on post-hoc testing, none of the constituent SNPs in either PRS showed significant associations with PTOA (i.e. P>0.05 following adjustment for multiple testing, data not shown).

GWAS for knee injury

In order to place PTOA genetic associations in context, examination of genome-wide associations with knee injury alone is needed (especially in studies where PTOA cases are compared with non-injured or population controls). A case-control GWAS for knee injury was performed in up to 33,300 individuals (3,074 knee injury cases and 30,226 non injured controls). No SNPs were associated with experiencing a knee injury at a genome-wide significance level ($P<5\times10^{-8}$) (Supplementary Figure 4). Four SNPs showed evidence of association at a suggestive level of significance ($P<1\times10^{-6}$) (Supplementary Table 11). rs2412204 was not one of these and showed no association with joint injury (P=0.81).

Discussion

We identified cases of clinically significant knee joint injury in a large dataset and longitudinally evaluated knee PTOA and its associated genetic risk. We developed and tested an approach identifying joint injury and PTOA from digital records with high specificity and moderate sensitivity, though likely with some differences based on the data source. The substantial size of UKB (half a million participants) means we add to what is known about the epidemiology of PTOA in a general population, as well as its genetic risk.

Knee injury occurred at a mean age of 34 years, in broadly similar numbers of men and women. As expected, knee inury substantially increased the risk of future knee OA. By defining rates of PTOA over predefined time periods and comparing these in non-injured, excess risk of PTOA could reasonably be demonstrated for at least 20 years after the injury (approximately twice the rate of iOA even at the end of this period). This is perhaps a longer risk period than anticipated and has relevance for our understanding of PTOA. The greatest risk was in the first 5 years after injury (three-fold greater than the background rate of iOA and 6-fold if only EHR data considered). This is the first time to our knowledge that longitudinal dynamics of OA risk after injury have been studied in a large population in this way(4, 5). Considering all available follow-up, knee OA was detected in 26% of those with knee injury after a median follow-up time of 21 years. This was lower than expected: frequencies of 30-70% over shorter time periods are reported, depending on cohort and injury type(5, 6).

Robust evidence is presented for increasing age at the time of knee injury being a substantial risk factor for future PTOA: every additional year of age increased PTOA risk by 7%. We replicate findings from a meta-analysis which recently reported age at time of knee injury being a risk factor for PTOA(25). Though female sex appeared to be an independent risk factor for PTOA after injury exposure, this effect was considerably more modest than for iOA as a whole, or compared with the cumulative effect of age. The strength of association also varied between data sources. This finding should be replicated in independent datasets: caution should be exercised, in view of the potential

for the effect of sex (and sex hormones) on OA differing over the lifecourse(26, 27). Others' findings differ, reporting either no sex effect or a possible increased risk in males(4, 9, 25). Here, relatively greater contamination with iOA in females at later follow-up times could have influenced our observations. Overall, these findings support the importance of age and sex in PTOA risk, but imply that their relative contribution may be different from iOA. Furthermore, though the occurrence of injuries appeared to be affected by socioeconomic factors, subsequent PTOA outcomes were found to be largely uninfluenced by such factors, highlighting a potential further difference between PTOA and iOA.

We also assessed genome-wide significant associations with PTOA in a time-to-event analysis, testing whether genetic variants could influence PTOA development after injury, and found no robustly significant signal, potentially due to power constraints. If genetic variation predisposes to PTOA, this could be mediated by increased knee injury risk in the first place, or the evolution of PTOA following the injury. For this reason, we carried out a separate GWAS of knee injury cases and non injured controls. Again, no SNPs reached genome-wide significance, in what was a sufficiently powered analysis to detect modest effects. Though candidate gene studies have been carried out, there had been no published GWAS of experiencing a knee injury(28).

Another possibility was that known iOA knee genes could account for at least some of the genetic risk of PTOA. In the only previous study by Valdes, some association was found for then known OA hits and reported past history of knee injury(19). However, the understanding of which variants should constitute a knee OA PRS has expanded recently(17). Our analyses showed limited evidence for known OA risk variants having a role in PTOA. However power was limited and larger-scale studies of PTOA are warranted.

There are some limitations to our work. Only 790 people met our criteria for PTOA in UK Biobank; following genetic exclusions and gating on 20 years, this dropped to 377. This is between 12% and 17% of all patients with a knee injury and less than 1% of all knee OA in UKB. Frequencies of both injury and PTOA were therefore much lower than expected(7). Our eligibility criteria were stringent,

with an upper age of 50 years at injury and further censoring at 20 years of follow-up actively reducing case numbers. This intentionally conservative approach guarded against the inclusion of individuals with iOA. There are known limitations of EHR case ascertainment, particularly its sensitivity for early OA. Diagnostic OA coding depends on interaction with clinical services and may be applied late (if at all) in the clinical pathway. This is sometimes around the time of referral for joint replacement(29, 30) (though only 72/1096 (6.6%) of our OA outcomes relied on knee arthroplasty codes). There is also potential for collider/ascertainment biases, i.e. someone with a knee injury is more likely to have orthopedic follow-up, be investigated and diagnosed with OA than an uninjured person. This was why a time-to-event analysis in those with knee injuries was our preferred methodology(18). The UKB is a non-random sampling of the UK population(31); furthermore, the age at recruitment for UKB was 40-70 years, but knee injuries tend to occur earlier in life. Though UKB provides long follow-up intervals supporting detection of incident PTOA, we rely on combinations of hospital records and/or SR of knee injuries from some time ago(18, 20). These record types do not generally record laterality, either of the injury or of the OA, so occurrence of some cases of OA in the contralateral knee cannot be excluded and could have reduced power. Lastly, BMI is a well-established risk factor for iOA(1), but its effects on PTOA risk are not well-established; BMI within 2 years of injury was described as a baseline characteristic but not analysed, as it had high missingness in this dataset.

It was of note that there were differences in relative rates of OA depending on source of data, with higher rates of progression to PTOA among EHR data, suggesting that medical records may be more effective at identifying PTOA cases than recalled history of injury. Over time, it appeared that SR data became more comparable proportionately to the EHR data. This could be because recollection of significant injury becomes more precise over a longer period of time, or just that SR data are generally less sensitive to case detection. The phenotypic approach used should be planned carefully in further studies, considering appropriate sensitivity analyses where there are different methods to ascertain injuries.

Whilst not a limitation as such, it should be noted that the study was not designed to assess the relative risks of different injury subtypes. There is already much work on this (4, 9, 10, 16). Given that acute mechanical insult to the joint is associated with a ubiquitous molecular response to injury (3, 32), and to increase power, it was always our intention to combine all clinically significant knee injuries, noting there would be heterogeneity in their relative risk for PTOA.

Large international efforts in cohorts who can accurately identify cases of knee injury and/or PTOA, such as in the GO consortium, could likely achieve the necessary sample size for adequate power to detect genome-wide significant signals. This UKB study provides a methodological roadmap justifying analytical approaches for larger coordinated efforts.

In summary, we present a first study testing the effects of genetic variation alongside other important predictors on the risk of PTOA following a significant knee injury in a large UK population. Our findings justify fully powered studies and international meta-analyses to definitively answer remaining questions. Studies of genetic architecture of the two conditions would conclusively tell us how similar or different these diseases really are, including whether PTOA is a distinct endotype of OA. This would improve our chances of identifying effective targets for preventing or slowing OA. Identifying individuals at moderate to high risk of OA would allow us to focus preventative efforts.

Notes

Genetics of Osteoarthritis Consortium:

Cindy G. Boer,[1] Konstantinos Hatzikotoulas,[2] Lorraine Southam,[2] Lilja Stefánsdóttir,[3] Yanfei Zhang,[4] Rodrigo Coutinho de Almeida,[5] Tian T. Wu,[6] Jie Zheng,[7] April Hartley,[7,8] Maris Teder-Laving,[9] Anne Heidi Skogholt,[10] Chikashi Terao,[11] Eleni Zengini,[12] George Alexiadis,[13] Andrei Barysenka,[2] Gyda Bjornsdottir,[3] Maiken E. Gabrielsen,[10] Arthur Gilly,[2] Thorvaldur Ingvarsson,[14,15] Marianne B. Johnsen,[10,16,17] Helgi Jonsson,[18,14] Margreet Kloppenburg,[19] Almut Luetge,[10] Sigrun H. Lund,[3] Reedik Mägi,[9] Massimo Mangino,[20] Rob R.G.H.H. Nelissen,[21] Manu Shivakumar,[22] Julia Steinberg,[2,23] Hiroshi Takuwa,[24,25] Laurent F. Thomas,[10,26,27,28] Margo Tuerlings,[5] arcOGEN Consortium, HUNT All-In Pain, ARGO Consortium, Regeneron Genetics Center, George C. Babis,[29] Jason Pui Yin Cheung,[30] Jae Hee Kang,[31] Peter Kraft,[32] Steven A. Lietman,[33] Dino Samartzis,[30,34] P. Eline Slagboom,[5] Kari Stefansson,[3,14] Unnur Thorsteinsdottir,[3,14] Jonathan H. Tobias,[8,7] André G. Uitterlinden,[1] Bendik Winsvold,[10,35,36] John-Anker Zwart,[10,35] George Davey Smith,[7,37] Pak Chung Sham,[38] Gudmar Thorleifsson,[3] Tom R. Gaunt,[7] Andrew P. Morris,[39] Ana M. Valdes,[40] Aspasia Tsezou,[41] Kathryn S.E. Cheah,[42] Shiro Ikegawa,[24] Kristian Hveem,[10,43] Tõnu Esko,[9] J. Mark Wilkinson,[44] Ingrid Meulenbelt,[5] Ming Ta Michael Lee,[4,45] Joyce B.J. van Meurs,[1] Unnur Styrkársdóttir,[3] and Eleftheria Zeggini[2,46]

[1]Department of Internal Medicine, Erasmus MC, Medical Center, 3015CN Rotterdam, the Netherlands
[2]Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany
[3]deCODE Genetics/Amgen Inc., 102 Reykjavik, Iceland
[4]Genomic Medicine Institute, Geisinger Health System, Danville, PA 17822, USA
[5]Department of Biomedical Data Sciences, Section Molecular Epidemiology, Postzone S05-P Leiden University Medical Center, 2333ZC Leiden, the Netherlands
[6]Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong, China

7MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK
8Musculoskeletal Research Unit, Translation Health Sciences, Bristol Medical School, University of Bristol, Southmead Hospital, Bristol BS10 5NB, UK
9Estonian Genome Center, Institute of Genomics, University of Tartu, 51010 Tartu, Estonia
10K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway
11Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa 230-0045, Japan
124th Psychiatric Department, Dromokaiteio Psychiatric Hospital, 12461 Athens, Greece
131st Department of Orthopaedics, KAT General Hospital, 14561 Athens, Greece
14Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland
15Department of Orthopedic Surgery, Akureyri Hospital, 600 Akureyri, Iceland
16Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, 0316 Oslo, Norway
17Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, 0424 Oslo, Norway
18Department of Medicine, Landspitali The National University Hospital of Iceland, 108 Reykjavik, Iceland
19Departments of Rheumatology and Clinical Epidemiology, Leiden University Medical Center, 9600, 23OORC Leiden, the Netherlands
20Department of Twin Research and Genetic Epidemiology, Kings College London, London SE1 7EH, UK
21Department of Orthopaedics, Leiden University Medical Center, 9600, 23OORC Leiden, the Netherlands
22Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
23Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW, Sydney, NSW 1340, Australia
24Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo 108-8639, Japan
25Department of Orthopedic Surgery, Shimane University, Shimane 693-8501, Japan
26Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, 7491 Trondheim, Norway
27BioCore-Bioinformatics Core Facility, Norwegian University of Science and Technology, 7491 Trondheim, Norway
28Clinic of Laboratory Medicine, St. Olavs Hospital, Trondheim University Hospital, 7030 Trondheim, Norway
292nd Department of Orthopaedics, National and Kapodistrian University of Athens, Medical School, Nea Ionia General Hospital Konstantopouleio, 14233 Athens, Greece
30Department of Orthopaedics and Traumatology, The University of Hong Kong, Pokfulam, Hong Kong, China
31Department of Medicine, Brigham and Women's Hospital, 181 Longwood Ave, Boston, MA 02115, USA
32Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA
33Musculoskeletal Institute, Geisinger Health System, Danville, PA 17822, USA
34Department of Orthopaedic Surgery, Rush University Medical Center, Chicago, IL 60612, USA
35Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital and University of Oslo, 0450 Oslo, Norway
36Department of Neurology, Oslo University Hospital, 0424 Oslo, Norway
37Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK
38Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong, China
39Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, University of Manchester, Manchester M13 9LJ, UK
40Faculty of Medicine and Health Sciences, School of Medicine, University of Nottingham, Nottingham, Nottinghamshire NG5 1PB, UK
41Laboratory of Cytogenetics and Molecular Genetics, Faculty of Medicine, University of Thessaly, Larissa 411 10, Greece
42School of Biomedical Sciences, The University of Hong Kong, Pokfulam, Hong Kong, China
43HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, 7600 Levanger, Norway
44Department of Oncology and Metabolism and Healthy Lifespan Institute, University of Sheffield, Sheffield S10 2RX, UK
45Institute of Biomedical Sciences, Academia Sinica, 115 Taipei, Taiwan
46Technical University of Munich and Klinikum Rechts der Isar, TUM School of Medicine, 81675 Munich, Germany.

Acknowledgements

This research has been conducted using the UK Biobank (UKB) Resource under Application Number 52507. We thank UKB and its participants. This work uses data provided by patients and collected by the NHS as part of their care and support.

FW is supported by a UKRI Future Leaders Fellowship (MR/S016538/1 and MR/S016538/2) and has also received support from Kennedy Trust for Rheumatology Research and the NIHR Oxford Biomedical Research Centre (BRC). This work was also supported in part by the Centre for OA Pathogenesis Versus Arthritis (grant numbers 20205 and 21621) of which FW is a co-applicant. FW is

a member of the Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis (grant number 21595).

The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

LJD is supported by Wellcome trust fellowship grant 208750/Z/17/Z.

Tissue samples and/or data were obtained from the Oxford Musculoskeletal Biobank and were collected with informed donor consent in full compliance with national and institutional ethical requirements, the UK Human Tissue Act, and the Declaration of Helsinki (HTA Licence 12217 and Oxford REC C 09/H0606/11). We thank the Oxford Knee Surgery Team including Andrew Price, William Jackson and Nicholas Bottomley and our centre tissue coordinators Louise Hill and Katherine Groves who coordinated this study. We thank the participants of the KICK, MenTOR and OMB studies. We acknowledge Professor Andrew Price and the other MenTOR and OxKIC investigators and Professor Tonia Vincent and other KICK investigators in the use of these cohort data for validation purposes and the support of the Versus Arthritis Centre for Osteoarthritis Pathogenesis in these various studies.

We thank Oxford University Hospitals Foundation NHS Trust (OUHFNT) for providing linked data for the purposes of research and their expert hospital coders for their advice on this project: Celeste Ward, Chris Middlemass, Mary O'Donnell (all OUHFNT).

Lastly we acknowledge the more recently formed GO-PTOA working group's members for their discussion of this work (https://www.genetics-osteoarthritis.com/active-working-groups/gwas-meta-analysis-of-post-traumatic-osteoarthritis-of-the-knee/index.html).

Author Contributions

All authors contributed to the conception and design of the study, or acquisition of data, or analysis and interpretation of data; all authors contributed to drafting and/or revision and all approved the final version to be submitted. Specific contributions are outlined below:

Design, concept: FW,LJD, EZ

Codes lists and analysis: BH, CC, LJD, LS, KH

Clinical expertise, design of criteria: FW, SK, AW, LJD

Writing manuscript: BH, FW, LJD

Reviewing and revising manuscript: all

FW and LJD take joint responsibility for the integrity of the work as a whole.

Role of funding source

This work was funded by a UKRI Future Leaders Fellowship (MR/S016538/1 and MR/S016538/2) to FW and has also received support from Kennedy Trust for Rheumatology Research and LJD and BH received support from Wellcome trust grant 208750/Z/17/Z.

The funder(s) of the study were not involved in the study design; collection, analysis, and interpretation of data; writing of the report; or decision to submit the paper for publication.

Conflict of Interest

FW has received personal consulting fees in 2020 from Pfizer for an unrelated compound. She is an associate editor for Osteoarthritis and Cartilage. Her institution received funding from Versus Arthritis for her role as lead of a Versus Arthritis Research Advisory Group.

No other authors declare a conflict of interest.

Studies including humans

Participants in UK Biobank (UKB) gave written informed consent to take part and could also withdraw their consent and data at any time. UKB has approval from the North West Multi-centre Research Ethics Committee (MREC), Haydock, UK as a Research Tissue Bank (RTB). This project and the related use of anonymised participant data was approved in June 2020 by UK Biobank (see Supplementary methods for futher details and also for ethical approvals of studies used in validation study).

Data Statement

The codelists for case ascertainment for this study have been made available. Links to these are given in Supplementary materials. UKB data can be accessed upon application (see

https://www.ukbiobank.ac.uk). The GO Consortium make their summary data available via the MSK Knowledge Portal (see https://www.genetics-osteoarthritis.com/data/index.html). Summary statistics for relevant analyses in UKB presented here will be uploaded to repositories including GWAS Catalog at the point of publication https://www.ebi.ac.uk/gwas/.

Figure legends

Figure 1. Flow diagram of study populations and related analyses in UK Biobank

Study population selection is shown on the left and pre-defined analyses on the right.

Inclusion and exclusion criteria were applied to select groups of individuals for all analyses. In addition, for genetic studies, genetic QC exclusion criteria were applied ahead of genetic studies.

Pre-defined comparisons in analyses are indicated by dotted lines.

Abbreviations: QC, quality control; GWAS, genome wide association study; iOA, idiopathic OA

Figure 2. Time-to-osteoarthritis analysis for knee injured vs. non-injured participants in UK Biobank

Individuals with knee injury (1) were matched 1:10 with non-injured controls, (0), by age and sex. Time, years (yrs), is time from injury . A 'pseudo age at injury' (index) for each non-injured control corresponded to the matched injured indiviudal. The upper panel shows a Kaplan Meier curve of OA-free survival probability as a function of time since injury. The lower panel shows the number at risk (i.e. number of cases that had neither developed OA nor been censored) at different time periods in our dataset.

Figure 3. Manhattan plot for GWAS of time-to-osteoarthritis in knee injured participants in UK Biobank

Manhattan plot is shown where the red line represents threshold for genome-wide significance ($P < 5 \times 10^{-8}$). The blue line represents suggestive significance level ($P < 1 \times 10^{-6}$).

Figure 4. Power calculations to detect genome wide significant associations for post-traumatic OA

Power plots including selected effect sizes, with assumed MAF of 0.3 are shown for genome wide association studies for **A**, time to event (PTOA) after joint injury, and **B**, comparing groups with PTOA and iOA (or knee injured individuals and healthy controls) in a case-control analysis.

HR, hazards ratio; OR, odds ratio; PTOA, post-traumatic OA; iOA, idiopathic OA

References

1. Glyn-Jones S, Palmer AJ, Agricola R, Price AJ, Vincent TL, Weinans H, et al. Osteoarthritis. Lancet. 2015;386(9991):376-87.

2. Showery JE, Kusnezov NA, Dunn JC, Bader JO, Belmont PJ, Jr., Waterman BR. The Rising Incidence of Degenerative and Posttraumatic Osteoarthritis of the Knee in the United States Military. J Arthroplasty. 2016;31(10):2108-14.

3. Watt FE. Posttraumatic osteoarthritis: what have we learned to advance osteoarthritis? Curr Opin Rheumatol. 2021;33(1):74-83.

4. Snoeker B, Turkiewicz A, Magnusson K, Frobell R, Yu D, Peat G, et al. Risk of knee osteoarthritis after different types of knee injuries in young adults: a population-based cohort study. Br J Sports Med. 2020;54(12):725-30.

5. Lohmander LS, Englund PM, Dahl LL, Roos EM. The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. Am J Sports Med. 2007;35(10):1756-69.

6. Frobell RB, Roos HP, Roos EM, Roemer FW, Ranstam J, Lohmander LS. Treatment for acute anterior cruciate ligament tear: five year outcome of randomised trial. BMJ. 2013;346:f232.

7. Brown TD, Johnston RC, Saltzman CL, Marsh JL, Buckwalter JA. Posttraumatic osteoarthritis: a first estimate of incidence, prevalence, and burden of disease. J Orthop Trauma. 2006;20(10):739-44.

8. Khan T, Alvand A, Prieto-Alhambra D, Culliford DJ, Judge A, Jackson WF, et al. ACL and meniscal injuries increase the risk of primary total knee replacement for osteoarthritis: a matched case-control study using the Clinical Practice Research Datalink (CPRD). Br J Sports Med. 2019;53(15):965-8.

9. Muthuri SG, McWilliams DF, Doherty M, Zhang W. History of knee injuries and knee osteoarthritis: a meta-analysis of observational studies. Osteoarthritis Cartilage. 2011;19(11):1286-93.

10. Poulsen E, Goncalves GH, Bricca A, Roos EM, Thorlund JB, Juhl CB. Knee osteoarthritis risk is increased 4-6 fold after knee injury - a systematic review and meta-analysis. Br J Sports Med. 2019;53(23):1454-63.

11. Mason D, Englund M, Watt FE. Prevention of posttraumatic osteoarthritis at the time of injury: Where are we now, and where are we going? J Orthop Res. 2021;39(6):1152-63.

12. Van Ginckel A, Verdonk P, Witvrouw E. Cartilage adaptation after anterior cruciate ligament injury and reconstruction: implications for clinical management and research? A systematic review of longitudinal MRI studies. Osteoarthritis Cartilage. 2013;21(8):1009-24.

13. van Meer BL, Oei EH, Meuffels DE, van Arkel ER, Verhaar JA, Bierma-Zeinstra SM, et al. Degenerative Changes in the Knee 2 Years After Anterior Cruciate Ligament Rupture and Related Risk Factors: A Prospective Observational Follow-up Study. Am J Sports Med. 2016;44(6):1524-33.

14. Potter HG, Jain SK, Ma Y, Black BR, Fung S, Lyman S. Cartilage injury after acute, isolated anterior cruciate ligament tear: immediate and longitudinal effect with clinical/MRI follow-up. Am J Sports Med. 2012;40(2):276-85.

15. Garriga C, Goff M, Paterson E, Hrusecka R, Hamid B, Alderson J, et al. Clinical and molecular associations with outcomes at 2 years after acute knee injury: a longitudinal study in the Knee Injury Cohort at the Kennedy (KICK). Lancet Rheumatol. 2021;3(9):e648-e58.

16. Khan T, Alvand A, Prieto-Alhambra D, Culliford DJ, Judge A, Jackson WF, et al. ACL and meniscal injuries increase the risk of primary total knee replacement for osteoarthritis: a matched case-control study using the Clinical Practice Research Datalink (CPRD). Br J Sports Med. 2018.

17. Boer CG, Hatzikotoulas K, Southam L, Stefansdottir L, Zhang Y, Coutinho de Almeida R, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell. 2021;184(24):6003-5.

18. Zengini E, Hatzikotoulas K, Tachmazidou I, Steinberg J, Hartwig FP, Southam L, et al. Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. Nat Genet. 2018;50(4):549-58.

19. Valdes AM, Doherty SA, Muir KR, Wheeler M, Maciewicz RA, Zhang W, et al. The genetic contribution to severe post-traumatic osteoarthritis. Ann Rheum Dis. 2013;72(10):1687-90.

20. DeBoever C, Tanigawa Y, Aguirre M, McInnes G, Lavertu A, Rivas MA. Assessing Digital Phenotyping to Enhance Genetic Studies of Human Diseases. Am J Hum Genet. 2020;106(5):611-22.

21. Bi W, Fritsche LG, Mukherjee B, Kim S, Lee S. A Fast and Accurate Method for Genome-Wide Time-to-Event Data Analysis and Its Application to UK Biobank. Am J Hum Genet. 2020;107(2):222-33.

22. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nat Genet. 2021;53(11):1527-33.

23. Swain S, Sarmanova A, Mallen C, Kuo CF, Coupland C, Doherty M, 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(6):792-801.

24. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018;50(9):1335-41.

25. Whittaker JL, Losciale JM, Juhl CB, Thorlund JB, Lundberg M, Truong LK, et al. Risk factors for knee osteoarthritis after traumatic knee injury: a systematic review and meta-analysis of randomised controlled trials and cohort studies for the OPTIKNEE Consensus. Br J Sports Med. 2022;56(24):1406-21.

26. Watt FE. Hand osteoarthritis, menopause and menopausal hormone therapy. Maturitas. 2016;83:13-8.

27. Gulati M, Dursun E, Vincent V, Watt FE. The influence of sex hormones on musculoskeletal pain and osteoarthritis. Lancet Rheumatol. 2023;5(4):E225-E38.

28. Maffulli N, Margiotti K, Longo UG, Loppini M, Fazio VM, Denaro V. The genetics of sports injuries and athletic performance. Muscles Ligaments Tendons J. 2013;3(3):173-89.

29. Yu D, Jordan KP, Peat G. Underrecording of osteoarthritis in United Kingdom primary care electronic health record data. Clin Epidemiol. 2018;10:1195-201.

30. Ferguson RJ, Prieto-Alhambra D, Walker C, Yu D, Valderas JM, Judge A, et al. Validation of hip osteoarthritis diagnosis recording in the UK Clinical Practice Research Datalink. Pharmacoepidemiol Drug Saf. 2019;28(2):187-93.

31. Swanson JM. The UK Biobank and selection bias. Lancet. 2012;380(9837):110.

32. Watt FE, Ismail HM, Didangelos A, Peirce M, Vincent TL, Wait R, et al. Src and fibroblast growth factor 2 independently regulate signaling and gene expression induced by experimental injury to intact articular cartilage. Arthritis Rheum. 2013;65(2):397-407.

Table 1. Sources and characteristics of knee injured individuals from UK Biobank

* From injury to end of observation period/censoring
Hospital codes are comprised of ICD-10 and OPCS-4 codes. OPCS-4 stands for OPCS Classification of Interventions and Procedures, version 4, a procedural classification which codes operations, procedures and interventions performed during inpatient stays, day case surgery and some outpatient treatments in NHS hospitals in the UK.
Primary care codes are comprised of Read Codes, a coded thesaurus of clinical terms used by the NHS. There are two versions, v2 and v3, which have been used by the NHS since 1985.
EHR codes are a combination of both Hospital and Primary care codes.
Self-reports is self-reported data on knee injury, given as part of UK Biobank data collection.

ICD-10, International Classification of Diseases, Tenth revision; OPCS, Office of Population Censuses and Surveys; EHR, electronic healthcare records; SD, standard deviation; IQR, inter-quartile range; NHS, UK National Health Service

Source	Cases, N	Sex, Female N (%)	Mean age at injury (SD), yrs	OA cases, N (%)	Median time to OA (IQR), yrs	Median follow-up time* (IQR), yrs
Hospital codes	738	265 (35.9)	42.6 (5.5)	149 (20.2)	9.9 (6.0, 13.7)	17.3 (12.4, 21.3)
Primary Care Codes	1,937	687 (35.5)	38.9 (9.0)	320 (16.5)	16.2 (7.6, 25.0)	20.9 (12.8, 28.4)
All EHR Codes	2,306	813 (35.3)	39.6 (8.4)	562 (24.3)	13.4 (7.0, 21.7)	21.7 (15.2, 29.2)
Self-reports	2,240	1,229 (54.9)	28.4 (9.0)	288 (12.9)	29.0 (19.3, 37.4)	34.0 (23.3, 42.3)
All sources	4,233	1,896 (44.8)	34.1 (10.4)	1,096 (25.9)	21.7 (10.5, 34.7)	30.2 (19.5, 45.4)

Table 2. Effect of knee injury, age and sex on time to knee osteoarthritis

Cox Proportional Hazard model for effects of knee injury and other covariates on time to knee OA. Individuals with injury were matched with non-injured controls (1:10) by age and sex (see Figure 1). A 'pseudo age at injury' (index) was assigned for each non-injured control corresponded to the time of injury of the matched individual's injury.

HR, hazard ratio; CI, confidence interval; EHR, electronic healthcare record; SR, self-reported data; OA, osteoarthritis

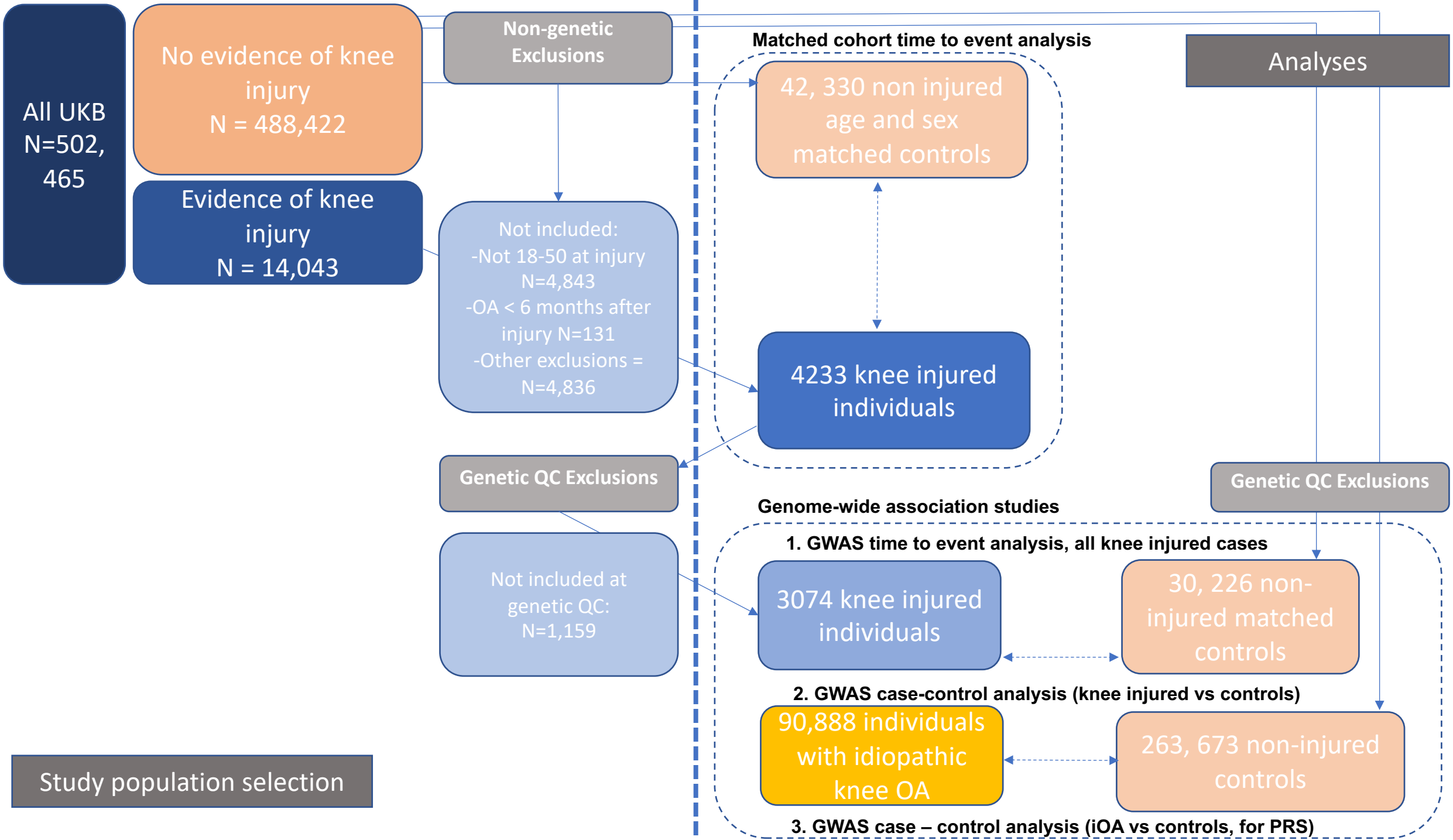
Covariate	All data sources (n = 46,563; OA events = 7,626)		EHR only (n = 25,366; OA events = 2,682)		SR only (n = 24,640; OA events = 2,316)	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
Sex (female)	1.41 (1.35, 1.48)	6.3×10^{-51}	1.32 (1.22, 1.43)	3.4×10^{-12}	1.68 (1.54, 1.83)	3.3×10^{-32}
Knee injury	1.81 (1.70, 1.93)	8.9×10^{-74}	3.06 (2.79, 3.36)	6.2×10^{-123}	1.18 (1.04, 1.33)	0.01
Age at injury	1.07 (1.07, 1.07)	$<1.0 \times 10^{-299}$	1.08 (1.08, 1.09)	1.3×10^{-177}	1.09 (1.08, 1.10)	1.8×10^{-179}

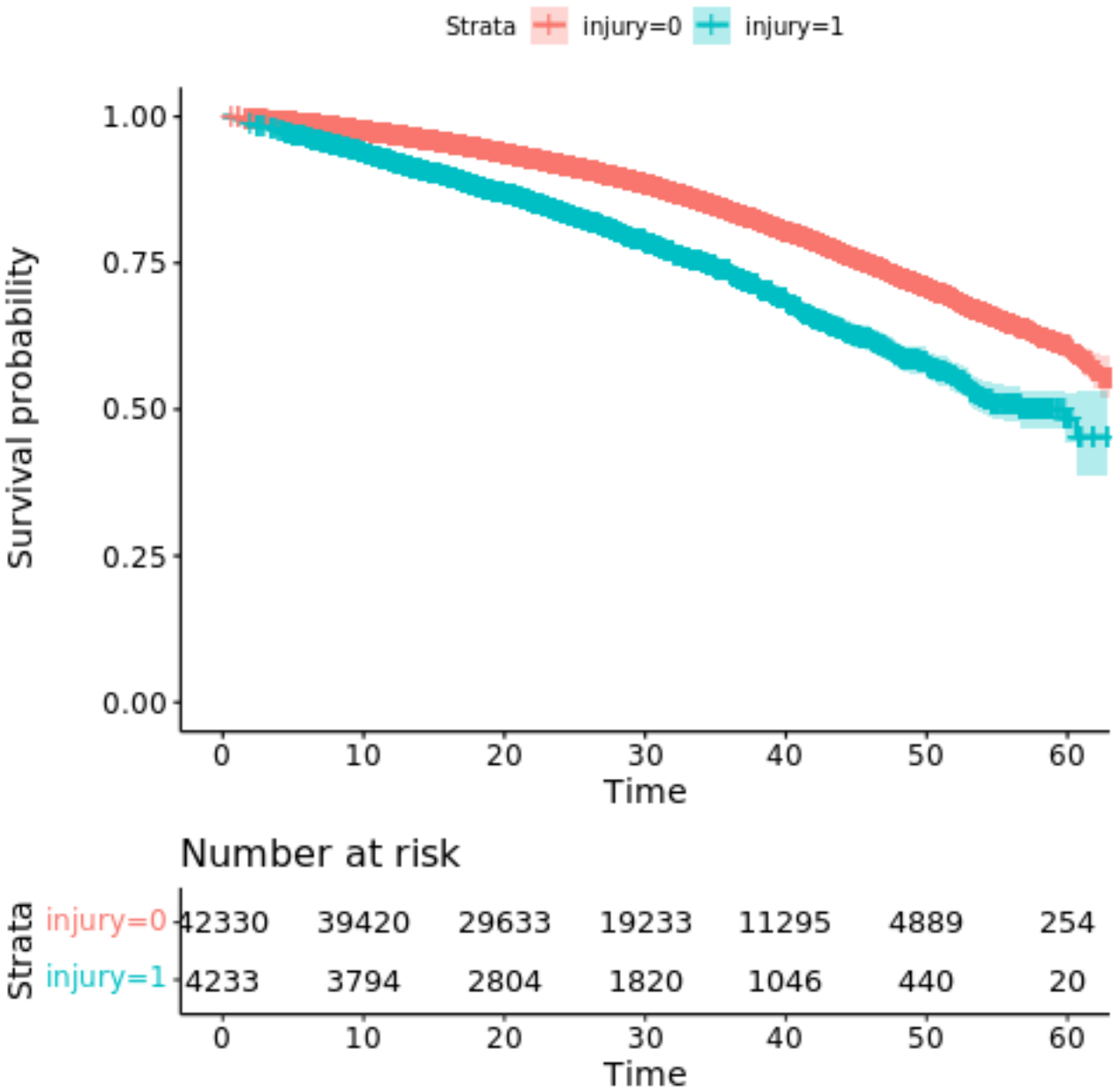
Table 3. Defining a time threshold for cases of post-traumatic osteoarthritis of the knee

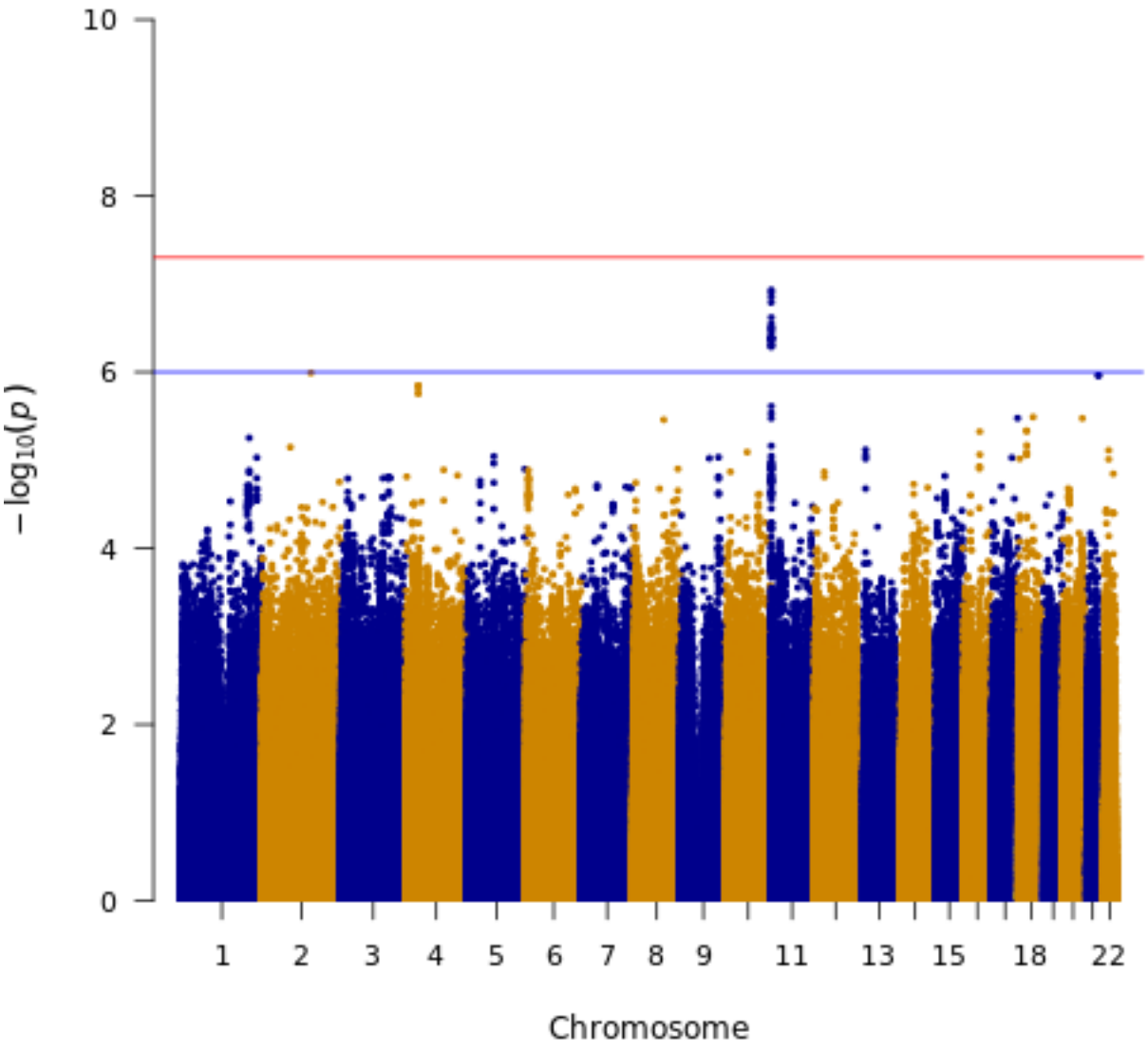
Covariates including time-varying categorical strata for time from injury were included in a Cox proportional hazard model. Results are show for all sources, for EHR and SR.

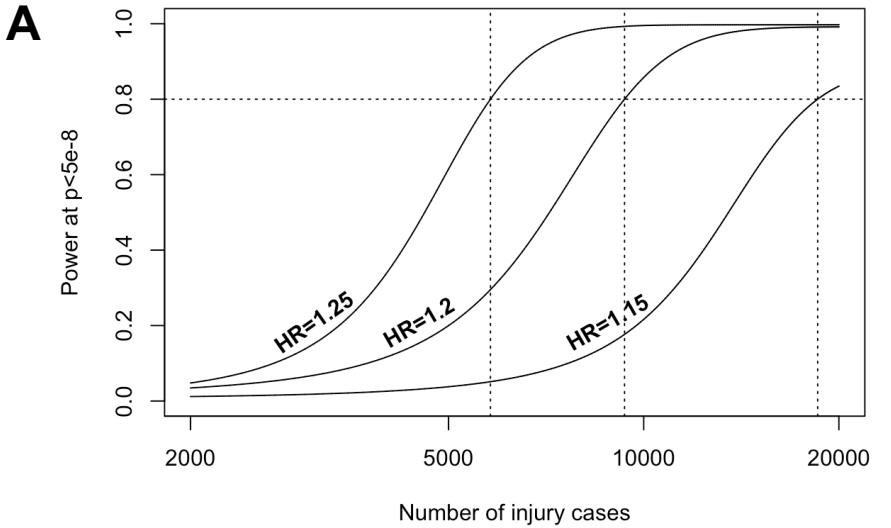
HR, hazard ratio; CI, confidence interval; EHR, electronic healthcare record data; SR, self-reported data; OA, osteoarthritis

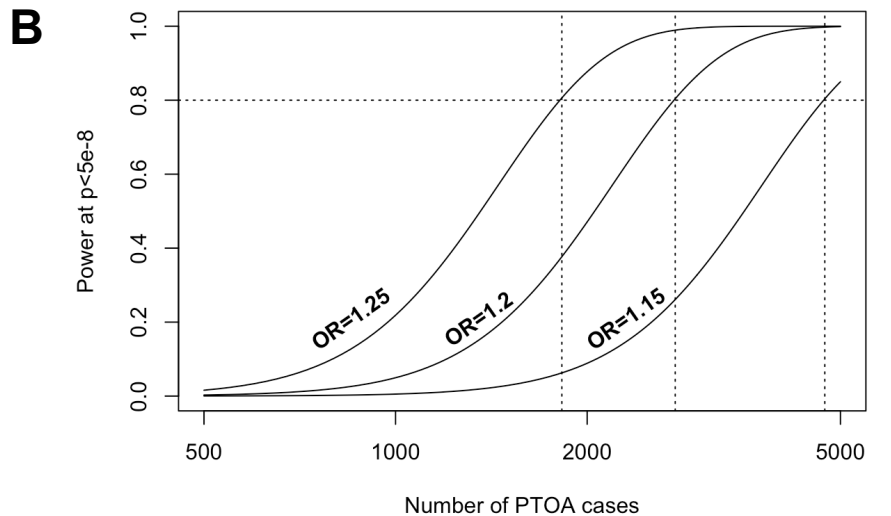
Covariate	All data sources (n = 46,563; OA events = 7,626)		EHR only (n = 25,366; OA events = 2,682)		SR only (n = 24,640; OA events = 2,316)	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
Sex (female)	1.41 (1.35, 1.48)	6.5×10^{-51}	1.32 (1.22, 1.43)	2.7×10^{-12}	1.68 (1.54, 1.83)	3.4×10^{-32}
Age at injury	1.07 (1.07, 1.07)	$<1.0 \times 10^{-299}$	1.08 (1.08, 1.09)	2.0×10^{-177}	1.09 (1.08, 1.10)	1.5×10^{-179}
Injury (< 5 yrs)	3.26 (2.67, 3.98)	5.0×10^{-31}	6.53 (5.14, 8.29)	1.7×10^{-53}	1.50 (0.85, 2.63)	0.16
Injury (≥5 yrs, <10 yrs)	2.27 (1.88, 2.75)	2.2×10^{-17}	3.39 (2.71, 4.25)	3.1×10^{-26}	1.64 (1.05, 2.58)	0.03
Injury (≥10 yrs, <20 yrs)	2.01 (1.76, 2.29)	6.3×10^{-25}	2.97 (2.54, 3.49)	5.1×10^{-41}	0.88 (0.63, 1.21)	0.43
Injury (≥20 yrs, <30 yrs)	1.76 (1.53, 2.03)	7.1×10^{-15}	2.58 (2.09, 3.19)	1.0×10^{-18}	1.24 (0.98, 1.57)	0.07
Injury (≥30 yrs, <40 yrs)	1.43 (1.23, 1.67)	3.1×10^{-6}	2.01 (1.42, 2.83)	6.9×10^{-5}	1.15 (0.92, 1.44)	0.23
Injury (≥40 yrs)	1.39 (1.18, 1.63)	6.2×10^{-5}	1.49 (0.93, 2.39)	0.10	1.23 (0.89, 1.71)	0.21


All UKB N=502, 465
No evidence of knee injury N = 488,422
Evidence of knee injury N = 14,043
Non-genetic Exclusions
Not included:
-Not 18-50 at injury N=4,843
-OA < 6 months after injury N=131
-Other exclusions = N=4,836
Matched cohort time to event analysis
42, 330 non injured age and sex matched controls
4233 knee injured individuals
Analyses
Genetic QC Exclusions
Not included at genetic QC: N=1,159
Genetic QC Exclusions
Genome-wide association studies
1. GWAS time to event analysis, all knee injured cases
3074 knee injured individuals
30, 226 non-injured matched controls
2. GWAS case-control analysis (knee injured vs controls)
90,888 individuals with idiopathic knee OA
263, 673 non-injured controls
3. GWAS case – control analysis (iOA vs controls, for PRS)
Study population selection


Strata
injury=0
injury=1
1.00
0.75
0.50
0.25
0.00
Survival probability
0
10
20
30
40
50
60
Time
Number at risk
Strata
injury=0
injury=1
42330
39420
29633
19233
11295
4889
254
4233
3794
2804
1820
1046
440
20
0
10
20
30
40
50
60
Time


10
8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
11
13
15
18
22
Chromosome


A
Power at p<5e-8
0.0
0.2
0.4
0.6
0.8
1.0
HR=1.25
HR=1.2
HR=1.15
2000
5000
10000
20000
Number of injury cases


B
Power at p<5e-8
0.0
0.2
0.4
0.6
0.8
1.0
OR=1.25
OR=1.2
OR=1.15
500
1000
2000
5000
Number of PTOA cases

Supplementary Material

Structure	Location	Trauma	Supporting Description
1	1	1	Evidence of an acute knee injury or a surgical procedure which is highly suggestive of this (this includes any clinically significant/ injury to the structures of the joint tissues or periarticular bone*, and/or dislocation of the joint)
1	1	0	Evidence of a clinically significant knee joint abnormality which is consistent with injury or a procedure on the knee which implies this, without definite evidence of trauma
1	0	1	Evidence of a clinically significant acute joint injury or procedure which is suggestive of this (but not specific to knee)
1	0	0	Evidence of a clinically significant joint abnormality or procedure on a joint which implies this (but not specific to knee), without definite evidence of trauma
0	1	1	Evidence of a traumatic injury involving the knee, but without definite evidence of clinical significance
0	1	0	Evidence of knee involvement without definite evidence of trauma or clinical significance
0	0	1	Evidence of traumatic episode compatible with a possible knee joint injury without definite evidence of knee involvement or clinical significance to the knee
0	0	0	No evidence of any of the above

Supplementary Table 1. Structure, location and trauma codes supporting case ascertainment of acute knee injury in UK Biobank

*To be included, cases had to achieve a code for structure, location and trauma in a set timeframe, either from a single code or in combination, across EHR and self-report.

Internal structures of the knee are intra-articular structures of the joint tissues include peri-articular bone, and/or dislocation of the joint, or surgical procedures carried out on these structures in keeping with injury to them e.g. reconstruction of a ligament).

Examples lacking sufficient evidence of clinically significant injury to an internal structure include MCL sprains (as opposed to rupture); presence of synovial plica; patello-femoral subluxation (as opposed to dislocation); quadriceps and patellar tendon injuries. Examples lacking sufficient evidence for acute trauma e.g. contusion, superficial wound, knee sprain.

Cohort	Total number of included participants	Cohort type	Code list validation	Number of eligible participants, based on draft criteria (%)
OxKIC	66	Acute knee injury	Inclusion (sensitivity)	36 (54.5%)
MenTOR	93	Degenerative meniscal tear	Inclusion (specificity)	0 (0.0%)
OMB	190	Established knee osteoarthritis	Exclusion/Knee OA outcome	187 (98.4%)

Supplementary Table 2. Validation of inclusion, exclusion and OA outcome code lists

Draft code lists were validated against three existing clinical cohorts for which hospital electronic health care data was gained.

Oxford Knee Injury Cohort (OxKIC) is a clinical cohort of participants (age 18-50 years) with acute knee injury as assessed by an orthopaedic surgeon within 12 weeks of injury. Participants' ICD-10 and OPCS-4 (procedure) codes were used to estimate the sensitivity of the inclusion codes for knee injury. Participants without any relevant episode data for review in export (N=4) were excluded from this analysis. Of the 30 participants not included by the draft code lists, 3/66 (4.5%) cases were not included but were true injury cases with missed knee-specific codes. Using this validation information, code lists were refined to include these relevant codes. A further 27/66 (40.9%) cases were not included appropriately by our criteria (26 cases had codes suggesting degenerative/chronic meniscal injury or codes were not sufficiently specific for knee trauma and 1 case was excluded as there was insufficient evidence for knee joint location). No cases were excluded for insufficient evidence of structural damage.

The Meniscal Tear and Osteoarthritis Risk (MenTOR) study is a longitudinal cohort study of participants with a confirmed meniscal tear likely to be degenerative in nature. Participants' OPCS-4 codes were used to estimate the specificity of the inclusion codes for knee injury. As expected, no participants' codes led to their inclusion based on our code lists. No update to these draft code lists were made based on this.

The OA Pathogenesis study, Oxford Musculoskeletal Biobank (OMB) is sub-collection of bio samples and associated clinical data typically collected from consenting people at the time of a surgical procedure as part of routine care. Data from participants who had taken part at the time of a knee arthroplasty were used to evaluate the sensitivity for OA outcome codes (which were also used as exclusion codes at time of eligibility assessment). Three participants had insufficient information to define knee OA. No updates to the draft code lists were therefore made.

Characteristic	Knee injured group, N=4233 N (%) or median [IQR]	Matched controls, N=42330) N (%) or median [IQR]	P value
Age at injury or pseudo-injury	34.51 [23.80, 43.50]	34.51 [23.80, 43.50]	NA
Sex (female)	1896 (44.8)	18960 (44.8)	NA
Body Mass Index[¥],	26.86 [24.2, 28.89]	26.22 [23.80, 29.17]	0.224
Completeness, n/N[§]	127 (3.0)	1345 (3.2)	
Ethnicity			
White	4005 (94.8)	39283 (93.4)	0.00204
Asian/Asian British	88 (2.1)	1201 (2.8)	
Black/African/Caribbean/ Black British	63 (1.5)	861 (2.0)	
Mixed/Multiple ethnic groups	32 (0.8)	296 (0.7)	
Other ethnic group	35 (0.8)	439 (1.0)	
Days/week walked more than 10 minutes[¥]	6 [4, 7]	6 [4, 7]	0.360
Completeness, n/N[§]	127 (3.0)	1348 (3.2)	
Days/week performed vigorous activity[¥]	2 [1, 3]	2 [0, 3]	0.0679
Completeness, n/N[§]	127 (3.0)	1348 (3.2)	
Pre-injury MET group			0.328
high	45 (45)	489 (44.1)	
medium	42 (42)	412 (37.2)	
low	13 (13)	207 (18.7)	
Age finished education	16 [16, 18]	16 [16, 18]	0.062
Total household income before tax			2.08×10^{-6}
<£18,000	596 (16.0)	6492 (17.6)	
£18,000 to £30,999	807 (21.6)	8543 (23.2)	
£31,000 to £51,999	1014 (27.1)	10229 (27.8)	
£52,000 to £100,000	997 (26.7)	9078 (24.6)	
>£100,000	321 (8.6)	2489 (6.8)	
Townsend Deprivation Index	-2.13 [-3.66, - 0.379]	-2.03 [-3.58, - 0.703]	0.00732
Long-standing illness or disability any time prior to injury[¥]	66 (20.4)	672 (20.5)	0.649
Completeness, n/N[§]	324 (7.7)	3276 (7.7)	
Current smoking[¥]	11 (8.7)	181 (13.4)	0.166
Completeness, n/N[§]	127 (3.0)	1348 (3.2)	
Ever smoked prior to injury	166 (50.2)	1874 (55.9)	0.0486
Completeness, n/N[§]	331 (7.8)	3350 (7.9)	

Supplementary Table 3. Baseline characteristics of knee injured individuals and matched controls

4233 knee injured individuals met eligibility criteria and 10 times the number of age- and sex- matched individuals without knee injury (42330) were identified in UKB. These factors and other clinically relevant characteristics available from UKB (at enrolment to UKB, or within 24 months of injury or pseudo-injury date where appropriate[¥]) are given, for both groups. [§]For those variables only reported within 24 months of injury, there was low completeness and this is noted in the row below.

‡Townsend Deprivation Index is a validated measure of social deprivation where higher scores indicate greater levels of deprivation (for reference, in whole of UKB study population mean -1.3, SD 3.09, minimum -6.3 and maximum 11.0).
https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=189

To explore and illustrate any potential differences, variables with multiple categories were compared by Fisher's exact test. Continuous measures were compared by Mann Whitney U test. P values are given.

Abbreviations: UKB, UK Biobank; MET, Metabolic Equivalent of Task

Covariate	All data sources, injured (n = 4,233; OA events = 1,096)		All data sources, non-injured (n = 42,330; OA events = 6,530)		Joint-injury-by-covariate interactions (n = 46,563; OA events = 7,626)	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
Sex (female)	1.15 (1.02, 1.30)	0.02	1.46 (1.39, 1.54)	1.41×10^{-52}	0.77 (0.67,0.87)	4.2×10^{-5}
Age at injury or pseudo-injury	1.07 (1.07, 1.08)	1.4×10^{-82}	1.07 (1.07,1.07)	$<10^{-52}$	1.02 (1.01, 1.02)	8.5×10^{-8}

Supplementary Table 4. Effect of age and sex on time-to-knee osteoarthritis amongst injured and non-injured participants, and interactions between covariates and joint injury status

Cox Proportional Hazard model for effects of age and sex, and age-by-injury and sex-by-injury interactions, on time to knee OA.

Abbreviations: HR, hazard ratio; CI, confidence interval; OA, osteoarthritis

Covariate	HR (95% CI)	P value
Sex (female)	1.35 (1.29, 1.42)	1.16×10^{-32}
Injury	1.87 (1.74, 2.00)	2.64×10^{-70}
Age at injury or pseudo-injury	1.07 (1.07, 1.07)	$<2.2 \times 10^{-308}$
Ethnicity (compared to Asian/Asian British)		
White	1.22 (1.00, 1.48)	0.050
Black/ African/Caribbean/Black British	1.24 (0.94, 1.66)	0.133
Mixed/Multiple ethnic groups	1.09 (0.72, 1.66)	0.675
Other ethnic group	1.16 (0.82, 1.65)	0.402
Townsend Deprivation Index (per unit increase)	1.01 (1.00, 1.02)	0.006
Associated household income compared with <£18,000		
£18,000 to £30,999	0.87 (0.82, 0.94)	8.95×10^{-5}
£31,000 to £51,999	0.78 (0.73, 0.84)	5.61×10^{-12}
£52,000 to £100,000	0.63 (0.58, 0.69)	1.27×10^{-28}
>£100,000	0.58 (0.51, 0.66)	1.17×10^{-15}

Supplementary Table 5. Robustness analysis for the effect of factors additional to age and sex on time-to-knee osteoarthritis amongst injured and non-injured participants

A post-hoc robustness analysis of factors found to be different in the matched cohort study (from Supplementary Table 3: ethnicity, 2 measures of socioeconomic deprivation) adjusting for the effects of these factors in the Cox Proportional Hazard model, on time to knee OA is shown.

P values are shown (**<0.01; ***<0.001)

Abbreviations: HR, hazard ratio; CI, confidence interval

SNP	Chr:Pos	EA	NEA	MAF	Stat	Var	P-value	Hazard ratio (95% CI)
rs2412204	11:3461016	A	G	0.13	48.98	82.24	1.17×10^{-7}	1.76 (1.45-2.13)

Supplementary Table 6. Summary statistics of nominally significant hit for time-to-OA after knee injury GWAS

A single nucleotide polymorphism (SNP) on chromosome 11 showing evidence of association with $P<1x10^{-6}$ is shown. Hazard ratio was calculated using the R function coxph. No SNPs were associated at a level consistent with genome-wide significance, but this requires replication.

Abbreviations: Chr, chromosome; Pos, position; EA, effect allele; NEA, non-effect allele; MAF, minor allele frequency; Stat, score statistic (from SPACox); Var, variance of score statistic

SNP	Chr:Pos	EA	NEA	EAF	OR [95% CI]	P-value	GO Subgroup
rs535914531	1:42186728	T	C	0.9996	0.13 [0.06, 0.29]	3.38×10^{-7}	Knee OA
rs76867045	1:156080110	A	G	0.9942	0.67 [0.57, 0.78]	4.62×10^{-7}	Knee OA
rs12739000	1:61322058	T	G	0.5741	0.95 [0.93, 0.97]	$7.29E\times10^{-7}$	Knee OA
rs538585726	1:153068188	A	G	0.9996	0.26 [0.15, 0.44]	9.39×10^{-7}	Knee OA
rs66906321	2:630070	T	C	0.1715	0.93 [0.91, 0.95]	5.69×10^{-10}	Knee OA
rs715	2:211543055	T	C	0.6996	0.95 [0.93, 0.97]	8.14×10^{-8}	Knee OA
rs553752945	2:209549412	A	C	0.0005	4.91 [2.67, 9.04]	3.09×10^{-7}	Knee OA
rs533824323	2:209793423	A	G	0.9995	0.22 [0.12, 0.40]	8.50×10^{-7}	Knee OA
rs17775871	2:211548674	A	G	0.8657	0.94 [0.91, 0.96]	9.35×10^{-7}	Knee OA
rs10188118	2:653623	C	G	0.7011	1.05 [1.03, 1.07]	9.46×10^{-7}	Knee OA
rs73192408	3:190038737	A	G	0.0233	0.85 [0.80, 0.90]	7.21×10^{-8}	Knee OA
rs147020044	4:151803371	A	C	0.9962	1.93 [1.51, 2.45]	1.09×10^{-7}	Knee OA
rs115730259	4:35920161	A	T	0.9996	0.15 [0.08, 0.31]	1.60×10^{-7}	Knee OA
rs147210849	4:35908469	T	G	0.0004	6.18 [3.06, 12.49]	4.00×10^{-7}	Knee OA
rs10062749	5:141805088	T	G	0.2650	1.06 [1.03, 1.08]	7.48×10^{-8}	Knee OA
rs13183212	5:128037372	A	G	0.1533	1.07 [1.04, 1.09]	2.53×10^{-7}	Knee OA
rs331073	5:127722810	T	G	0.2060	1.06 [1.03, 1.08]	4.32×10^{-7}	Knee OA
rs3817066	5:748477	A	C	0.3202	0.91 [0.88, 0.95]	5.45×10^{-7}	Knee OA
rs17677555	5:127852612	C	G	0.2391	1.05 [1.03, 1.08]	7.82×10^{-7}	Knee OA
rs74929562	7:76554271	T	C	0.2128	0.95 [0.93, 0.97]	7.60×10^{-7}	Knee OA
rs143477522	9:17677351	T	C	0.0394	1.81 [1.44, 2.27]	4.44×10^{-7}	Knee OA
rs12337611	9:17660373	C	G	0.9610	0.56 [0.45, 0.71]	9.30×10^{-7}	Knee OA
rs140642204	10:62494823	A	G	0.0038	1.46 [1.26, 1.70]	6.42×10^{-7}	Knee OA
rs7299016	12:129883226	T	C	0.9989	0.37 [0.26, 0.54]	1.72×10^{-7}	Knee OA
rs188871438	12:1521465	T	G	0.0044	1.77 [1.41, 2.22]	6.81×10^{-7}	Knee OA
rs10842226	12:23959589	A	G	0.4227	1.05 [1.03, 1.08]	6.97×10^{-7}	Knee OA
rs534894292	13:97668192	A	G	0.0013	3.25 [2.16, 4.89]	1.49×10^{-8}	Knee OA
rs181902473	13:97254164	T	C	0.9987	0.43 [0.31, 0.59]	1.10×10^{-7}	Knee OA
rs576724334	13:100035209	T	C	0.0013	3.09 [2.03, 4.69]	1.39×10^{-7}	Knee OA
rs11852372	15:78801394	A	C	0.6550	1.05 [1.03, 1.08]	1.86×10^{-7}	Knee OA
rs9940278	16:53800200	T	C	0.4427	1.07 [1.05, 1.09]	6.93×10^{-13}	Knee OA
rs1477386	16:71600477	C	G	0.3768	0.95 [0.94, 0.97]	1.58×10^{-7}	Knee OA
rs6499509	16:71396921	T	C	0.1817	0.94 [0.92, 0.96]	5.92×10^{-7}	Knee OA
rs4548913	17:2209888	A	G	0.6233	0.95 [0.93, 0.97]	6.62×10^{-9}	Knee OA
rs4429327	17:77331690	T	C	0.0603	1.17 [1.10, 1.24]	1.65×10^{-7}	Knee OA
rs11654663	17:65389305	A	C	0.7675	1.06 [1.03, 1.08]	5.72×10^{-7}	Knee OA
rs1984749	17:2249904	T	C	0.6418	0.95 [0.94, 0.97]	5.95×10^{-7}	Knee OA
rs117661768	17:62567437	T	G	0.9881	1.36 [1.21, 1.54]	6.29×10^{-7}	Knee OA
rs58571470	18:33838549	C	G	0.8902	0.93 [0.90, 0.96]	5.86×10^{-7}	Knee OA
rs143384	20:34025756	A	G	0.5856	1.06 [1.04, 1.08]	5.40×10^{-11}	Knee OA
rs3746410	20:34190870	A	G	0.7902	1.07 [1.04, 1.09]	7.02×10^{-9}	Knee OA
rs143041253	20:62916207	C	G	0.9987	0.32 [0.21, 0.48]	6.64×10^{-8}	Knee OA

rs5756672	22:37800169	A	G	0.3541	0.95 [0.94, 0.97]	3.04×10^{-7}	Knee OA
rs538482408	1:62323572	A	G	0.0019	7.95 [3.48, 18.15]	8.78×10^{-7}	TKR
rs72764904	2:916925	T	C	0.7155	0.91 [0.87, 0.94]	4.81×10^{-7}	TKR
rs112187097	5:111052298	A	G	0.9393	0.84 [0.78, 0.90]	8.33×10^{-7}	TKR
rs9388386	6:125196368	A	G	0.2960	0.90 [0.87, 0.94]	2.66×10^{-7}	TKR
rs1989391	6:1784495	A	G	0.1754	0.89 [0.84, 0.93]	5.45×10^{-7}	TKR
rs2539499	7:76557047	T	C	0.4455	1.09 [1.06, 1.13]	6.55×10^{-7}	TKR
rs192118480	8:15737520	T	G	0.0031	5.01 [2.72, 9.22]	2.39×10^{-7}	TKR
rs10131337	14:37144516	T	C	0.2639	0.90 [0.86, 0.93]	1.42×10^{-7}	TKR
rs4646568	15:58344290	T	C	0.3690	0.91 [0.88, 0.95]	3.60×10^{-7}	TKR
rs141235941	16:78551412	A	G	0.9973	0.33 [0.21, 0.51]	5.01×10^{-7}	TKR
rs4783581	16:69552785	C	G	0.7798	0.90 [0.86, 0.94]	7.15×10^{-7}	TKR

Supplementary Table 7. Summary statistics for SNPs included in PRS for knee-specific OA from GO Consortium

A PRS was constructed from 54 top hits for knee-specific OA in a sub-study of the GO consortium (which reached a nominal level of significance ($P<1\times10^{-6}$), selected those with knee-specific OA (including replacement) and excluded UKB participants who had been included in their main analysis).

Abbreviations: GO, Genetics of Osteoarthritis consortium; PRS, polygenic risk score; SNP, single nucleotide polymorphism; Chr, chromosome; Pos, position; EA, effect allele; NEA, non-effect allele; OR, odds ratio; 95% CI, 95% confidence interval; OA, osteoarthritis; TKR, total knee replacement

Covariate	All sources (n = 3,074; OA events = 377)		EHR only (n = 1,698; OA events = 290)		SR only (n = 1,585; OA events = 58)	
	HR [95% CI]	P-value	HR [95% CI]	P-value	HR [95% CI]	P-value
Sex (female)	1.06 [0.86, 1.30]	0.60	1.02 [0.80, 1.30]	0.86	1.53 [0.89, 2.62]	0.12
Genotyping batch	1.00 [1.00, 1.00]	0.44	1.00 [1.00, 1.00]	0.69	1.00 [0.99, 1.01]	0.48
PC1	1.01 [0.99, 1.04]	0.31	1.01 [0.99, 1.03]	0.41	0.88 [0.75, 1.04]	0.13
PC2	1.01 [0.98, 1.05]	0.47	1.01 [0.97, 1.06]	0.57	1.00 [0.84, 1.18]	0.98
PC3	0.98 [0.93, 1.03]	0.40	0.99 [0.93, 1.06]	0.78	0.87 [0.73, 1.04]	0.12
PC4	0.99 [0.96, 1.02]	0.41	0.99 [0.96, 1.02]	0.48	0.98 [0.86, 1.10]	0.70
PC5	1.00 [0.98, 1.01]	0.78	0.99 [0.98, 1.01]	0.47	1.00 [0.95, 1.05]	0.99
PC6	1.01 [0.95, 1.07]	0.82	1.01 [0.95, 1.08]	0.71	0.96 [0.81, 1.13]	0.63
PC7	1.02 [0.98, 1.05]	0.37	1.01 [0.98, 1.05]	0.47	1.11 [0.97, 1.27]	0.14
PC8	1.00 [0.96, 1.05]	0.90	0.99 [0.94, 1.03]	0.55	1.10 [0.95, 1.26]	0.20
PC9	1.00 [0.98, 1.03]	0.84	0.99 [0.97, 1.02]	0.52	1.00 [0.94, 1.07]	0.92
PC10	1.00 [0.96, 1.05]	0.94	0.98 [0.93, 1.03]	0.43	1.05 [0.93, 1.19]	0.43
Age at injury	1.10 [1.08, 1.11]	3.83×10^{-51}	1.07 [1.05, 1.08]	9.57×10^{-15}	1.09 [1.06, 1.11]	8.47×10^{-10}
PRS[1]	0.43 [0.02, 8.41]	0.58	0.17 [0.01, 4.97]	0.30	3.70 [0.00, 9508.78]	0.74

Supplementary Table 8. Effect of PRS for knee-specific OA from GO Consortium (without UKB) on time to knee-OA amongst participants with knee injury

[1]This PRS was constructed from 54 top hits from the GO consortium data (see Supplementary Table 7).

Abbreviations: OA, osteoarthritis; EHR, electronic healthcare records; SR, self-reported cases; HR, hazard ratio; 95% CI, 95% confidence interval; PC, principal component (as supplied by UK Biobank); PRS, polygenic risk score

SNP	Chr:Pos	EA	NEA	EAF	OR [95% CI]	P-value
rs71659381	1:113459676	A	G	0.08	1.06 [1.04, 1.08]	3.50×10^{-8}
rs762810932	1:21916694	D	I	0.25	0.96 [0.95, 0.98]	3.91×10^{-8}
rs1256324	1:21910649	A	G	0.25	1.04 [1.02, 1.05]	7.27×10^{-8}
1:174124967_GA_G	1:174124967	D	I	0.29	0.97 [0.95, 0.98]	8.26×10^{-8}
rs3755381	2:70718695	C	T	0.48	1.05 [1.04, 1.06]	6.93×10^{-20}
rs113179593	2:81877733	A	C	0.06	1.09 [1.06, 1.11]	1.80×10^{-12}
rs7581258	2:33453763	T	C	0.49	1.04 [1.02, 1.05]	6.78×10^{-10}
rs11690912	2:20496452	C	T	0.46	0.97 [0.96, 0.98]	6.63×10^{-9}
rs62194158	2:203642155	A	G	0.45	1.03 [1.02, 1.04]	3.76×10^{-8}
rs11434039	3:50080920	D	I	0.50	0.96 [0.94, 0.97]	1.23×10^{-14}
3:49791637_CA_C	3:49791637	D	I	0.49	1.03 [1.02, 1.04]	2.97×10^{-8}
rs13107325	4:103188709	T	C	0.07	1.10 [1.08, 1.12]	3.10×10^{-18}
rs34592089	4:102926923	A	G	0.06	1.08 [1.06, 1.10]	7.03×10^{-12}
rs35236367	4:121593596	D	I	0.24	0.96 [0.95, 0.97]	4.51×10^{-9}
rs4543129	4:130302554	T	C	0.21	1.04 [1.03, 1.05]	9.71×10^{-9}
rs11731421	4:1749160	A	G	0.33	1.03 [1.02, 1.05]	1.78×10^{-8}
rs28405622	4:151148930	T	C	0.30	1.03 [1.02, 1.05]	3.43×10^{-8}
rs56166900	4:122708120	A	T	0.06	0.94 [0.91, 0.96]	7.88×10^{-8}
rs140898346	4:13063846	T	C	0.40	0.97 [0.96, 0.98]	8.82×10^{-8}
rs9277511	6:33054215	C	A	0.23	0.96 [0.94, 0.97]	2.81×10^{-11}
rs10947428	6:33647058	C	T	0.22	1.04 [1.03, 1.06]	5.10×10^{-10}
rs6903171	6:33864137	C	G	0.14	1.05 [1.03, 1.06]	1.16×10^{-8}
rs9277861	6:33105817	T	C	0.16	1.04 [1.03, 1.06]	1.22×10^{-8}
rs6928965	6:6894948	G	A	0.12	0.95 [0.93, 0.97]	1.54×10^{-8}
rs2137058	6:23853635	C	T	0.35	0.97 [0.96, 0.98]	4.60×10^{-8}
rs561029283	6:32958362	D	I	0.21	0.96 [0.95, 0.98]	9.79×10^{-8}
rs1079359	7:157462464	G	A	0.41	0.97 [0.96, 0.98]	3.38×10^{-8}
rs6986032	8:11032240	T	C	0.48	1.04 [1.02, 1.05]	7.59×10^{-10}
rs7819602	8:10726842	G	C	0.39	1.04 [1.02, 1.05]	1.20×10^{-9}
rs10875467	8:142233396	A	C	0.41	0.97 [0.96, 0.98]	3.24×10^{-8}
rs2176630	8:8166321	G	C	0.40	0.97 [0.96, 0.98]	4.61×10^{-8}
rs3101432	8:9164772	G	A	0.47	1.03 [1.02, 1.04]	8.96×10^{-8}
rs59323233	9:116899138	G	C	0.28	1.04 [1.03, 1.06]	6.45×10^{-12}
rs4978572	9:116943302	G	C	0.41	1.04 [1.03, 1.05]	4.15×10^{-10}
rs1574285	9:4283137	T	G	0.41	0.97 [0.96, 0.98]	5.84×10^{-8}
rs7396943	11:13328979	C	G	0.39	0.96 [0.95, 0.97]	7.79×10^{-12}
rs557675	11:65566719	G	T	0.47	0.97 [0.96, 0.98]	5.70×10^{-9}
11:43621868_CT_C	11:43621868	D	I	0.14	1.05 [1.03, 1.07]	1.04×10^{-8}

rs11030104	11:27684517	G	A	0.20	0.96 [0.95, 0.97]	1.16×10^{-8}
rs111591982	11:64905748	A	G	0.11	0.95 [0.93, 0.97]	1.54×10^{-8}
rs382603	11:30877924	C	T	0.34	1.03 [1.02, 1.05]	2.97×10^{-8}
rs158146	11:31210433	A	G	0.37	1.03 [1.02, 1.04]	4.63×10^{-8}
12:123678364_GAC_G	12:123678364	D	I	0.23	0.95 [0.93, 0.96]	8.78×10^{-16}
rs2732441	12:48723324	G	C	0.32	0.97 [0.95, 0.98]	9.88×10^{-9}
rs11168366	12:48421773	C	T	0.36	0.97 [0.96, 0.98]	1.35×10^{-8}
rs61953491	12:123932196	G	A	0.34	1.04 [1.02, 1.05]	2.04×10^{-8}
12:123908878_AT_A	12:123908878	D	I	0.16	1.05 [1.03, 1.07]	5.84×10^{-8}
rs11638576	15:75260387	A	G	0.41	1.04 [1.03, 1.05]	1.38×10^{-12}
rs11634109	15:75643714	C	T	0.26	1.04 [1.02, 1.05]	5.83×10^{-9}
rs147516418	15:75894226	D	I	0.26	1.04 [1.02, 1.05]	9.55×10^{-9}
rs35697691	15:52353498	G	C	0.09	1.06 [1.04, 1.08]	5.38×10^{-8}
rs11852686	15:74715205	C	T	0.48	0.97 [0.96, 0.98]	9.65×10^{-8}
rs4985464	16:69978687	C	T	0.34	0.96 [0.95, 0.97]	1.27×10^{-10}
rs56305262	16:88814958	A	G	0.34	1.04 [1.03, 1.05]	4.96×10^{-10}
rs577551101	16:89897316	D	I	0.40	1.03 [1.02, 1.05]	5.95×10^{-8}
rs216195	17:2203167	G	T	0.31	0.96 [0.95, 0.98]	1.69×10^{-9}
rs144056477	17:30401304	C	T	0.13	0.95 [0.94, 0.97]	2.79×10^{-8}
19:10763214_ATT_A	19:10763214	D	I	0.36	1.05 [1.03, 1.06]	2.36×10^{-14}
rs75621460	19:41833784	A	G	0.03	1.14 [1.10, 1.17]	6.24×10^{-12}
rs12978481	19:10964259	C	G	0.07	0.93 [0.91, 0.95]	2.34×10^{-11}
rs143383	20:34025983	G	A	0.38	0.96 [0.95, 0.97]	2.24×10^{-12}
rs532201406	20:34301588	D	I	0.27	0.96 [0.95, 0.98]	2.12×10^{-8}
rs150116001	20:34370184	T	G	0.10	0.95 [0.93, 0.97]	4.32×10^{-8}
rs4818310	21:18396523	G	A	0.26	0.97 [0.95, 0.98]	3.89×10^{-8}

Supplementary Table 9. Summary statistics for SNPs included in PRS for iOA (non-injured) in UK Biobank

A PRS was constructed from 64 top hits for knee-specific OA, excluding those with knee injury from UKB (which reached nominally suggestive significance ($P<1\times10^{-7}$).

Abbreviations: UKB, UKBiobank; SNP, single nucleotide polymorphism; Chr, chromosome; Pos, position; EA, effect allele; NEA, non-effect allele; OR, odds ratio; 95% CI, 95% confidence interval

Covariates	All sources (n = 3,074; OA events = 377)		EHR only (n = 1,698; OA events = 290)		SR only (n = 1,585; OA events = 58)	
	HR [95% CI]	P-value	HR [95% CI]	P-value	HR [95% CI]	P-value
Sex (female)	1.07 [0.87, 1.31]	0.53	1.02 [0.80, 131]	0.85	1.53 [0.90, 2.63]	0.12
Genotyping batch	1.00 [1.00, 1.00]	0.41	1.00 [1.00, 1.00]	0.66	1.00 [0.99, 1.01]	0.50
PC1	1.01 [0.99, 1.04]	0.34	1.01 [0.99, 1.04]	0.39	0.88 [0.75, 1.03]	0.10
PC2	1.01 [0.98, 1.05]	0.50	1.01 [0.97, 1.06]	0.57	0.99 [0.84, 1.17]	0.90
PC3	0.98 [0.93, 1.03]	0.39	0.99 [0.93, 1.06]	0.83	0.87 [0.73, 1.03]	0.11
PC4	0.99 [0.96, 1.02]	0.38	0.99 [0.96, 1.02]	0.48	0.98 [0.86, 1.10]	0.70
PC5	1.00 [0.98, 1.01]	0.67	0.99 [0.98, 1.01]	0.43	1.00 [0.95, 1.05]	0.97
PC6	1.01 [0.96, 1.07]	0.83	1.01 [0.95, 1.08]	0.73	0.97 [0.82, 1.14]	0.70
PC7	1.02 [0.98, 1.05]	0.38	1.01 [0.98, 1.05]	0.49	1.10 [0.96, 1.27]	0.16
PC8	1.00 [0.96, 1.05]	0.85	0.99 [0.94, 1.03]	0.56	1.08 [0.94, 1.24]	0.25
PC9	1.00 [0.98, 1.02]	0.88	0.99 [0.97, 1.02]	0.53	1.00 [0.94, 1.07]	0.93
PC10	1.00 [0.96, 1.05]	0.9	0.98 [0.93, 1.03]	0.44	1.05 [0.93, 1.19]	0.42
Age at injury	1.10 [1.09, 1.11]	1.86×10^{-51}	1.07 [1.05, 1.08]	9.87×10^{-15}	1.09 [1.06, 1.12]	5.06×10^{-10}
PRS[1]	18.87 [1.71, 208.69]	0.02	2.66 [0.17, 40.84]	0.48	322.8 [0.61, 169593.55]	0.07

Supplementary Table 10. Effect of PRS for iOA from on time to knee-OA amongst participants with knee injury

[1]This PRS was constructed from 64 top hits from OA cases in UKB with no evidence of past history of knee injury (see Supplementary Table 9).

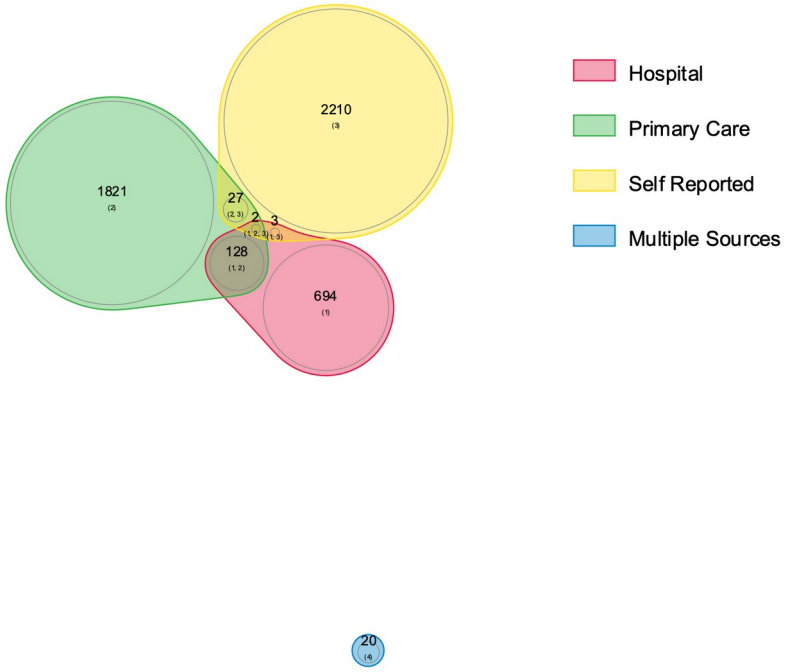
Abbreviations: OA, osteoarthritis; UKB, UK Biobank; EHR, electronic healthcare records; SR, self-reported cases; HR, hazard ratio; 95% CI, 95% confidence interval; PC, principal components (as supplied by UKB); PRS, polygenic risk score

SNP	Chr:Pos	EA	NEA	EAF	OR [95% CI]	P-value
7:106299784_AT_A	7:106299784	D	I	0.04	1.36 [1.24, 1.48]	5.31×10^{-7}
rs113767068	8:124890802	C	T	0.03	1.41 [1.27, 1.55]	8.52×10^{-7}
rs3025406	9:136511019	C	T	0.05	1.33 [1.21, 1.45]	7.39×10^{-7}
rs17105816	14:37433851	T	C	0.08	1.28 [1.18, 1.38]	3.39×10^{-7}

Supplementary Table 11. Summary statistics of nominally significant SNPs for knee injury GWAS

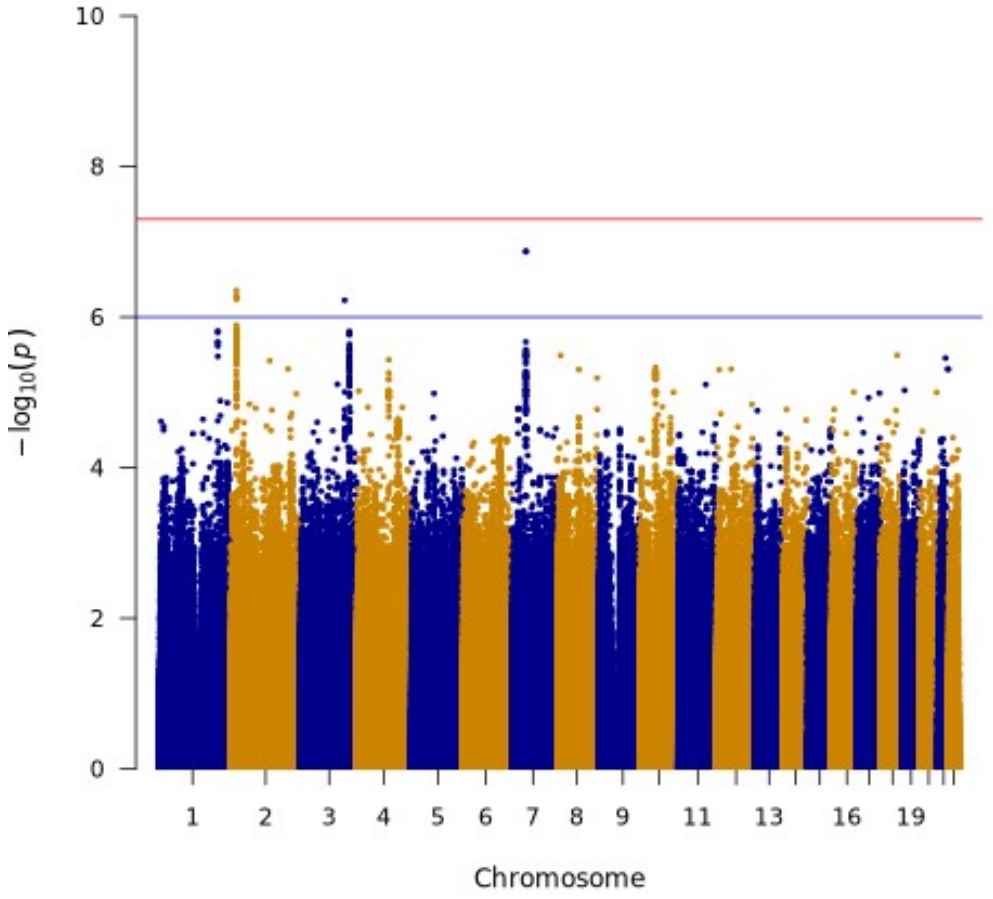
Single nucleotide polymorphisms (SNPs) showing evidence of associations with $P<1x10^{-6}$ are shown.
*No SNPs were associated at a level consistent with genome-wide significance, but this requires replication.

Abbreviations: Chr, chromosome; Pos, position; EA, effect allele, NEA, non-effect allele; EAF, effect allele frequency; OR, odds ratio; 95% CI, 95% confidence interval. * tri-allelic variant



Supplementary Figure 1. Venn diagram of case identification source
Codes used from multiple sources only identified an additional 20 cases.

Abbreviations: HES, Hospital Episode Statistics (either ICD10 or procedure codes); GP, General Practice (Primary care Read codes); SR, Self-reported data collected as part of UK Biobank case report data; Multi, Multiple source

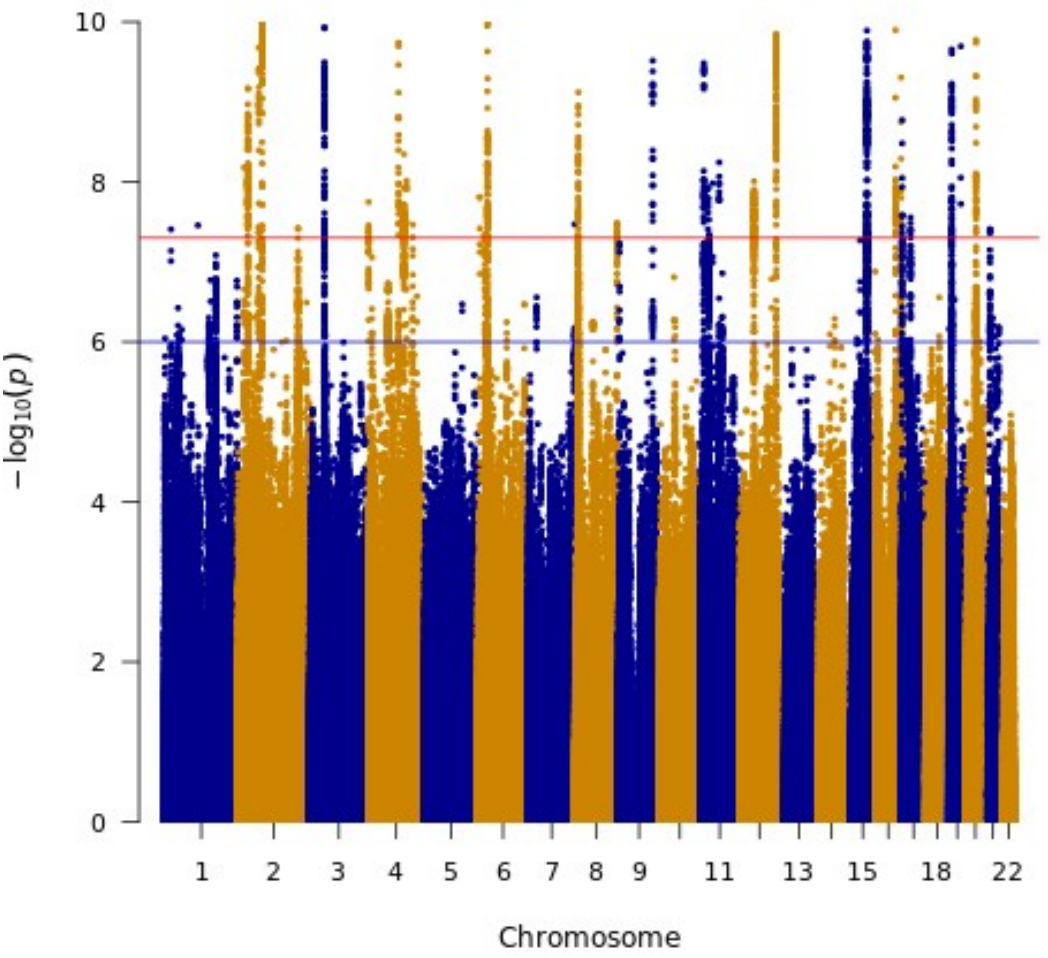


identification

Supplementary Figure 2. Manhattan plot for time-to-OA GWAS of knee injured cases identified by EHR codes only

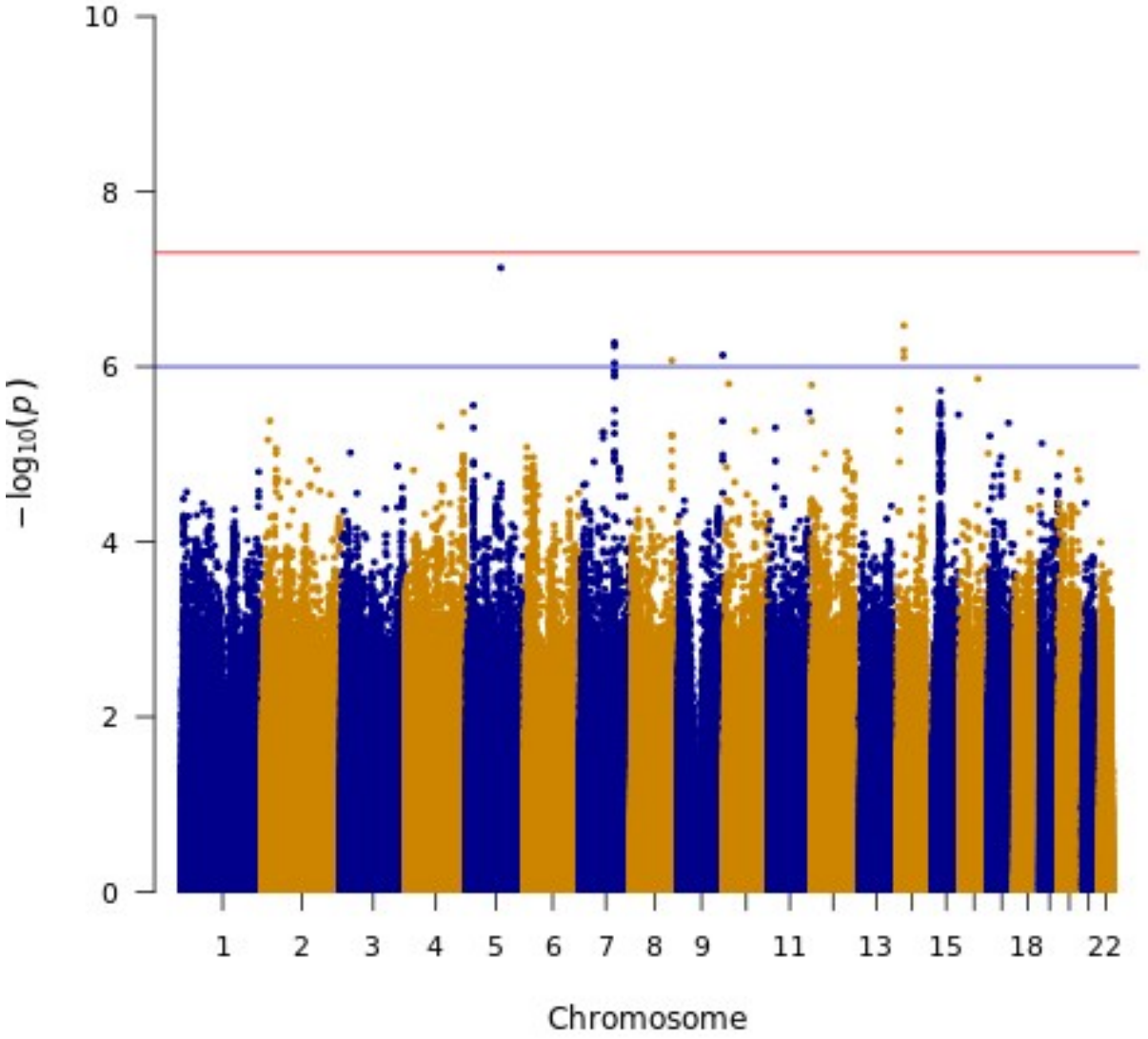
A post-hoc sensitivity analysis was run of the primary time-to-event GWAS, including only N=1698 participants who had injury ascertainment and OA outcome by electronic health record (EHR) data only.

The red line represents threshold for genome-wide significance ($P < 5 \times 10^{-8}$), blue line represents nominal significance level ($P < 1 \times 10^{-6}$).



Supplementary Figure 3. Manhattan plot for GWAS of idiopathic (non-injured) osteoarthritis in UK Biobank

The red line represents threshold for genome-wide significance ($P < 5 \times 10^{-8}$), blue line represents nominal significance level ($P < 1 \times 10^{-6}$).



Supplementary Figure 4. Manhattan plot for GWAS of traumatic knee injury in UK Biobank

The red line represents threshold for genome-wide significance ($P < 5 \times 10^{-8}$), blue line represents nominal significance level ($P < 1 \times 10^{-6}$).

Supplementary Methods

Ethics

UK Biobank (UKB) has approval from the North West Multi-centre Research Ethics Committee (MREC), Haydock, UK as a Research Tissue Bank (RTB) approval. https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics This approval means that individual researchers do not require separate ethical clearance and can operate under the UKB's RTB approval, following approval of a project by them. We followed this process for this project.

Participating GO cohorts have individual ethical permissions which permit this work and are listed in the GO consortium author block and in related references.

Ethical approval numbers for the cohorts in the validation study are given in the relevant section below.

Exclusion criteria

The presence of any of the following, diagnosed at **any time PRIOR** to the date of the knee injury[#]:

1. Osteoarthritis (or non-specified arthropathy which could be osteoarthritis[a]) of either knee, hip or ankle
2. Partial or total knee arthroplasty of either knee, hip or ankle
3. Clinically significant injuries to the hip including hip and pelvic fracture.

The presence of any of the following, diagnosed **6 months or more PRIOR** to the date of the knee injury[#]:

4. Chronic instability or recurrent patellofemoral dislocation of the knee without evidence of a prior acute knee injury of known timing
5. Acquired structural abnormalities of the knee, without evidence of a prior acute knee injury of known timing (e.g. degenerative meniscus included cystic change, chondrocalcinosis) or which are not attributable to an acute knee injury that meets the inclusion criteria[b].

The presence of any of the following **at ANY TIME** (prior, during or after the date of knee injury[#] and up until the first date of osteoarthritis or equivalent):

6. Inflammatory arthritis[c] or associated connective tissue disorder (e.g. rheumatoid arthritis, psoriatic arthritis, juvenile idiopathic arthritis, enteropathic arthritis, inflammatory spondyloarthropathy, Lupus or Sjogren's syndrome; generalised or lower

limb gout, CPPD, reactive arthritis or septic arthritis); neuropathic arthritis or inherited collagen vascular disease affecting joints.

7. Clinically significant congenital structural abnormality of the knee, or clinically significant congenital or acquired structural abnormality of the hip or ankle (e.g. discoid meniscus, absence of anterior cruciate ligament, congenital dislocation of hip, protrusio acetabuli)
8. Septic arthritis or osteomyelitis of the knee
9. Individuals with non-normal weight bearing (e.g. wheelchair user, those with spina bifida or cerebral palsy, lower limb amputations)

Evidence of more extensive musculoskeletal trauma **within the 'injury window'** for the acute knee injury[#]:

10. Open or closed fractures of the lower limb which are not likely to be intra-articular in origin (e.g. proximal or mid femoral fractures, proximal or distal tibial fractures, distal fibular fractures).
11. Clinically significant neurovascular or nerve palsies of the lower limb likely caused by lower limb trauma e.g. peroneal nerve, sciatic nerve or femoral nerve palsies.
12. Multisite musculoskeletal trauma, outside of the knee.

[#]where the date of the knee injury is the start of a three month 'injury window', which begins at the first evidence of injury

[a] This includes climacteric arthritis, HPOA, chondrolysis and inherited or metabolic disorders which always give rise to OA

[b] Osgood Schlatters disease (osteochondritis of the tibial tuberosity) and functional disorders of the patella including chondromalacia patellae are not sufficient to fulfil this criterion. Operative procedures to the joint such as arthroscopy are not sufficient to meet this criterion, unless definitely attributable to a relevant structural abnormality which was not due to an acute injury.

[c]Allergic arthritis, transient arthritis, palindromic arthritis, post immunisation arthritis, intermittent hydrarthrosis are not considered sufficient/specific.

Establishment, cross-mapping and validation of inclusion and exclusion code lists

Supplementary Table 1 outlines the principles applied to code review for inclusion code lists. A clinical expert panel (a rheumatologist FW, a knee orthopaedic surgeon AW, a sports and exercise medicine physician SK and a hospital coder CW), advised on interpretation of codes, to ensure these were sufficient evidence of clinically significant structural injury, and specific enough to suggest

acute knee trauma. Manual search of relevant code trees was carried out to identify inclusion codes (LJD, FW). Additional publications and data sources were searched for valid code lists for knee injury[1-2]. To ensure consistency and to identify codes that may be been missed, cross mapping of inclusion codes between different coding schemes (ICD-10, OPCS-4, Read2 and Read3) was carried out using code maps provided by NHS England's Technology Reference Update Distribution (TRUD) service, which led to review and addition of relevant codes. Inclusion codes lists were then compiled and a script written to search across these to identify individuals who met the inclusion criteria (https://github.com/dr-bench/PTOA).

Exclusion criteria are given in full above. Briefly these were: 1) Evidence prior to the injury of OA (or partial or total arthroplasty) of the knee, hip or ankle, or significant injuries to the hip; 2) diagnoses 6 months or more prior to the injury of chronic knee instability or acquired structural abnormalities of the knee without evidence of a knee injury; 3) the presence at any time of inflammatory arthritis, associated connective tissue disorder, neuropathic arthritis or collagen vascular disease, congenital structural abnormality of the knee, congenital or acquired structural abnormality of the hip or ankle, septic arthritis or osteomyelitis of the knee or non-normal weight bearing; and 4) evidence within the 'injury window' of more extensive musculoskeletal trauma (including lower limb fractures not likely to be intra-articular), lower limb neurovascular injury or multisite musculoskeletal trauma. Exclusion code lists were then compiled for all types of code lists, with similar searches, review and cross mapping approaches as inclusions (https://github.com/dr-bench/PTOA).

An outcome of subsequent earliest documented knee OA code (henceforth referred to as first knee OA code) was defined as diagnostic codes compatible with definite, probable or possible knee OA and/or total knee replacement/revision consistent with knee OA. Additional data sources were searched for valid code lists for OA[3-14]. Again all available EHR codes (ICD-10, OPCS-4, Read2, Read3) were reviewed and excluded where they suggested OA of the lower limb but not of the knee i.e. hip or ankle location, or OA that was definitely not of the lower limb. Knee, hip and ankle location codes

were employed in OPCS-4, Read2 and Read3 to improve specificity. Knee OA code lists were then compiled (https://github.com/dr-bench/PTOA).

Validation study of eligibility and outcome code lists

To test the sensitivity of the EHR-based inclusion, exclusion and outcome classifiers, we applied our code lists to three cohorts of known knee injury/OA status with local ethical approvals: the Oxford Knee Injury Cohort (OxKIC), of which all participants are known to meet our inclusion criteria and none of our exclusion criteria, the Meniscal Tear and Osteoarthritis Risk (MenTOR) cohort, of which all participants are known to either not meet our inclusion criteria or meet our exclusion criteria, a subproject in the Oxford Musculoskeletal Biobank (OMB), in which all participants should meet our knee OA outcome criteria. We used OxKIC to test our inclusion/exclusion sensitivity (proportion of true knee injury cases identified), MenTOR to test our inclusion/exclusion specificity (one minus the proportion of false knee injury cases identified) and OMB to test our outcome sensitivity (proportion of true knee OA cases identified) (Supplementary Table 2).

Participants in these cohorts had previously given written informed consent and provided information as part of taking part in these cohorts. Their consent in all cases allowed access to related medical record data. All data exports of their linked NHS data used in this validation study were generated and provided in an anonymous way by Oxford University Hospitals NHS Foundation Trust to researchers, following appropriate internal approvals. All three studies have received ethical approval granted by Research Ethics Committees (RECs) in the UK as follows:

- OxKIC: 14/WA/1108; Wales REC 7, UK
- MenTOR: 15/SC/0551; South Central - Oxford REC B, UK
- OMB (also see acknowledgements, due to approvals): 09/H0606/11; Oxford REC C, UK

Description of knee injury group

The number and demographics of knee injury cases were calculated, based on the data source of their identification. Data sources were defined as 1) Hospital codes; 2) Primary care codes; 3) SR codes; 4) EHR codes (Hospital and Primary care codes combined); and 5) all sources combined. It was

possible that SR data would be less reliable and could bias results. In view of this, the number and % of OA cases (those knee injured with subsequent knee OA) and the median time to first OA diagnostic code after knee injury, the median follow up time (i.e. time from date of injury to date of last possible observation, henceforth referred to as observation period) and the related effects sizes were calculated for each data source separately as well as combined data sources, as a sensitivity analysis. IQRs were calculated for medians and SDs for means, after assessing the normality of the data using histograms.

Injury case ascertainment from different data sources was visualised using Venn diagrams. Specifically, the number of cases identified within each data source, whether these cases were identified from other data sources as well and the additional cases identified when ascertaining from combined data sources were all calculated.

Baseline characteristics were described: in addition to the pre-defined covariates (age at injury, sex), ethnicity, presence of comorbidity prior to the injury, BMI, physical activity levels and smoking in the 24 months prior to the injury, and socioeconomic factors (Townsend Deprivation Index, Age finished education, total household income before tax) were listed, as factors of relevance to osteoarthritis that might vary between cases and controls.

Genome-wide association studies

Samples were removed from genetic analyses if they were not included in the pre-defined QC-passing unrelated samples UK Biobank field (used.in.pca.calculation), if they self-declared an ethnicity other than White, White British or White Irish, had a detected aneuploidy or other chromosomal abnormality, or had withdrawn consent. Variants were removed from the analysis if they had a low minor allele frequency (MAF < 0.01) or a low imputation quality score (info < 0.4). The genetic studies in UK Biobank were performed in a pre-defined subset (provided by UK Biobank) that self-defined as “White British” or “White Irish”, had similar genetic ancestry based on principal components, and were not closely related to one another. The number of ethnic minority joint injury

and PTOA patients was too small to analyse separately (N=300, of which 72 developed OA), and were removed to reduce the risk of population stratification.

In addition to the time-to-event GWAS (methods and populations described in main manuscript Methods and above), two predefined additional related case-control GWAS were employed

i) Case-control GWAS for knee iOA

This was conducted to enable the PRS studies. All UKB participants passing sample QC were eligible for inclusion. In this instance, cases were those participants identified as having knee OA from our code lists, excluding those with knee injury. Controls were all remaining UKB participants without knee OA and without knee injury.

ii) Case-control GWAS for knee injury

This was conducted on the same matched knee injury group as was used in the survival analysis, excluding samples not passing QC. Controls were those without injury.

For these additional two GWAS, a logistic regression model was run with sex, age at injury, genotype chip and 10 previously genetically determined principal components (PCs) received from UKB included as covariates. PLINK (v1.9) was used for whole genome analyses. R (version 4.1.2) was used for all other analyses.

iOA Polygenic Risk Scores

Polygenic risk scores (PRS) for iOA were generated by two methods: 1) using the summary statistics for knee-specific OA hits from the GO consortium data set[15], using two subgroups ('Knee OA' and 'Total Knee Replacement') and excluding UK Biobank participants, approved subproject January 2022; this generated summary statistic data for 80,697 cases and 2) using the lead variants for the knee iOA GWAS in 90, 888 UKB participants described above. Distance and LD-based clumping was performed to identify lead variants at nominally suggestive significance thresholds of $P < 1 \times 10^{-6}$ (250 kb, LD $r^2 > 0.1$) for knee-specific OA hits from GO and $P < 1 \times 10^{-7}$ (250 kb, LD $r^2 > 0.1$) for knee iOA case-control GWAS in UKB. PRS were constructed by weighting the total number of risk alleles by the effect sizes (log odds ratios) for each of the two methods.

These two different 'iOA PRS' were then included as a covariate in two different Cox Proportional Hazards regression models in UKB participants with knee injury, assessing their association with outcome, PTOA. Exclusions and other covariates were as in the primary survival analysis.

Power calculations

At the outset, based on best assumptions on rates of PTOA and case finding within 500,000 individuals, power calculations were made. >5000 knee injury cases would give sufficient power for a full GWAS (approximately 2000 knee injury cases were anticipated in UKB in feasibility studies, so this was not planned). However if 700 individuals with subsequent PTOA were identified in the injury cases (based on 35% of those developing OA over a median 5 year period[16]) we noted at least 50% power to detect associations with an OR of ≥1.2 for 10 or more selected variants.

A further power calculation was planned, based on data generated by this project to enable powering of a full meta-analysis for a PTOA GWAS. This power calculation was carried out using simulations, using the estimated time-to-event and time-to-censuring distributions to simulate risk variants with different hazard ratios, with power established by measuring the proportion of variants meeting the defined p-value threshold in a Cox Proportional Hazards test.

References

1. Khan T, Alvand A, Prieto-Alhambra D, Culliford DJ, Judge A, Jackson WF, et al. ACL and meniscal injuries increase the risk of primary total knee replacement for osteoarthritis: a matched case-control study using the Clinical Practice Research Datalink (CPRD). Br J Sports Med. 2018.

2. Snoeker B, Turkiewicz A, Magnusson K, Frobell R, Yu D, Peat G, et al. Risk of knee osteoarthritis after different types of knee injuries in young adults: a population-based cohort study. Br J Sports Med. 2020;54(12):725-30.

3. Zengini, E., Hatzikotoulas, K., Tachmazidou, I. et al. Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. Nat Genet 2018. 50, 549–558

4. Funck-Brentano T, Nethander M, Movérare-Skrtic S, Richette P, Ohlsson C. Causal Factors for Knee, Hip, and Hand Osteoarthritis: A Mendelian Randomization Study in the UK Biobank. Arthritis Rheumatol. 2019 Oct;71(10):1634-1641

5. Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, Johnson T, Koprulu M, Zengini E, Steinberg J, Wilkinson JM, Bhatnagar S, Hoffman JD, Buchan N, Süveges D; arcOGEN Consortium, Yerges-Armstrong L, Smith GD, Gaunt TR, Scott RA, McCarthy LC, Zeggini E. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat Genet. 2019 Feb;51(2):230-236

6. Styrkarsdottir, U., Lund, S.H., Thorleifsson, G. et al. Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis. Nat Genet 50, 1681–1687 (2018)

7. Keele OA code list (supplied on request)

8. Ford, E., Nicholson, A., Koeling, R. et al. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text?. BMC Med Res Methodol 13, 105 (2013)

9. Carreira H, Williams R, Strongman H, et al Identification of mental health and quality of life outcomes in primary care databases in the UK: a systematic review,BMJ Open 2019

10. Arden N, Altman D, Beard D, Carr A, Clarke N, Collins G, Cooper C, Culliford D, Delmestri A, Garden S, Griffin T, Javaid K, Judge A, Latham J, Mullee M, Murray D, Ogundimu E, Pinedo-Villanueva R, Price A, Prieto-Alhambra D, Raftery J. Lower limb arthroplasty: can we produce a tool to predict outcome and failure, and is it cost-effective? An epidemiological study. Southampton (UK): NIHR Journals Library; 2017 Jun

11. Metcalfe D, Masters J, Delmestri A, et al. Coding algorithms for defining Charlson and Elixhauser co-morbidities in Read-coded databases. BMC Med Res Methodol. 2019;19(1):115. Published 2019 Jun 6. doi:10.1186/s12874-019-0753-5

12. Graves N, Wloch C, Wilson J, et al. A cost-effectiveness modelling study of strategies to reduce risk of infection following primary hip replacement based on a systematic review. Southampton (UK): NIHR Journals Library; 2016 Jul. (Health Technology Assessment, No. 20.54.) Appendix 4, Primary total hip replacement OPCS codes
13. Abram SGF, Palmer AJR, Judge A, Beard DJ, Price AJ. Rates of knee arthroplasty in patients with a history of arthroscopic chondroplasty: results from a retrospective cohort study utilising the National Hospital Episode Statistics for England. *BMJ Open*. 2020;10(4):e030609. Published 2020 Apr 16. doi:10.1136/bmjopen-2019-030609
14. Sibanda N, Copley LP, Lewsey JD, et al. Revision rates after primary hip and knee replacement in England between 2003 and 2006. *PLoS Med*. 2008
15. Boer CG, Hatzikotoulas K, Southam L, Stefansdottir L, Zhang Y, Coutinho de Almeida R, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell. 2021;184(24):6003-5.
16. 2. Frobell RB, Roos HP, Roos EM, Roemer FW, Ranstam J, Lohmander LS. Treatment for acute anterior cruciate ligament tear: five year outcome of randomised trial. BMJ. 2013;346:f232.