OPEN ACCESS

Gait, physical activity and tibiofemoral cartilage damage: a longitudinal machine learning analysis in the Multicenter Osteoarthritis Study

Kerry E Costello ,[1,2,3] David T Felson ,[3] S Reza Jafarzadeh ,[3] Ali Guermazi ,[4] Frank W Roemer ,[5,6] Neil A Segal ,[7,8] Cora E Lewis,[9] Michael C Nevitt,[10] Cara L Lewis ,[2,3] Vijaya B Kolachalama ,[11,12] Deepak Kumar [2,3]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bjsports-2022-106142).

For numbered affiliations see end of article.

Correspondence to
Deepak Kumar, Boston University, Boston 02215, Massachusetts, USA; kumard@bu.edu

Accepted 20 February 2023
Published Online First 3 March 2023

ABSTRACT

Objective To (1) develop and evaluate a machine learning model incorporating gait and physical activity to predict medial tibiofemoral cartilage worsening over 2 years in individuals without advanced knee osteoarthritis and (2) identify influential predictors in the model and quantify their effect on cartilage worsening.

Design An ensemble machine learning model was developed to predict worsened cartilage MRI Osteoarthritis Knee Score at follow-up from gait, physical activity, clinical and demographic data from the Multicenter Osteoarthritis Study. Model performance was evaluated in repeated cross-validations. The top 10 predictors of the outcome across 100 held-out test sets were identified by a variable importance measure. Their effect on the outcome was quantified by g-computation.

Results Of 947 legs in the analysis, 14% experienced medial cartilage worsening at follow-up. The median (2.5–97.5th percentile) area under the receiver operating characteristic curve across the 100 held-out test sets was 0.73 (0.65–0.79). Baseline cartilage damage, higher Kellgren-Lawrence grade, greater pain during walking, higher lateral ground reaction force impulse, greater time spent lying and lower vertical ground reaction force unloading rate were associated with greater risk of cartilage worsening. Similar results were found for the subset of knees with baseline cartilage damage.

Conclusions A machine learning approach incorporating gait, physical activity and clinical/demographic features showed good performance for predicting cartilage worsening over 2 years. While identifying potential intervention targets from the model is challenging, lateral ground reaction force impulse, time spent lying and vertical ground reaction force unloading rate should be investigated further as potential early intervention targets to reduce medial tibiofemoral cartilage worsening.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Gait and physical activity features are potential intervention targets to slow knee osteoarthritis progression, but little is known about their role in early knee osteoarthritis.

⇒ Underlying interactions among gait, physical activity, demographic and clinical features make traditional statistical approaches challenging.

WHAT THIS STUDY ADDS

⇒ Machine learning models predicted cartilage worsening in persons without advanced knee osteoarthritis from gait, physical activity and clinical and demographic characteristics with a median area under the receiver operating characteristic curve of 0.73 across 100 held-out test sets.

⇒ High lateral ground reaction force impulse, more time spent lying and low vertical ground reaction force unloading rate were associated with increased risk of cartilage worsening over 2 years.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Gait and physical activity are some of the only modifiable risk factors for knee osteoarthritis; this study identified three potential intervention targets to slow early knee osteoarthritis progression.

INTRODUCTION

Knee osteoarthritis (OA) is a progressive, painful joint disease and leading cause of disability, affecting over 350 million adults.[1] While some individuals with advanced disease undergo knee replacement, there is no cure and many experience pain and poor quality of life for decades. Existing structural damage and other risk factors (eg, obesity, malalignment) can drive further degeneration.[2 3] Addressing this burden will require early identification of at-risk individuals and discovery of intervention targets that can be addressed before the onset of extensive damage or other risk factors.

Joint loading is one of few modifiable risk factors for knee OA[4] and can be manipulated through gait and physical activity. While prior research has identified gait features associated with medial tibiofemoral knee OA progression,[5] these were typically examined in isolation, in small samples and/or without accounting for other risk factors. Importantly, little is known about gait and physical activity predictors of progression early in the disease process. Machine learning can identify features in complex datasets that are important to prediction without requiring assumptions about underlying relationships among features, making it useful for exploring gait and physical activity.[6–8]

To cite: Costello KE, Felson DT, Jafarzadeh SR, *et al*. *Br J Sports Med* 2023;**57**:1018–1024.

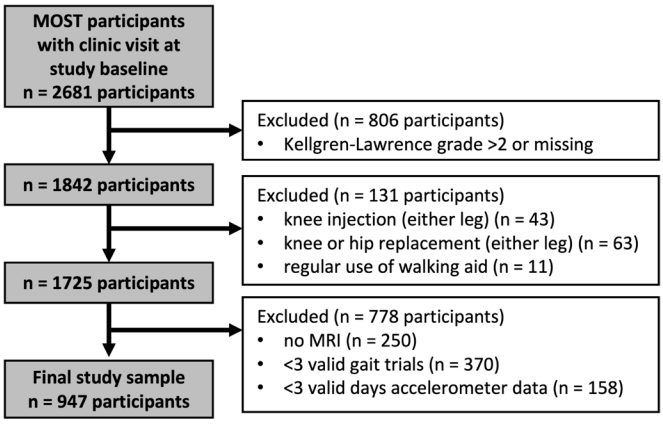


Figure 1 Study sample from the Multicenter Osteoarthritis Study (MOST).

The Multicenter Osteoarthritis Study (MOST)[9] is a large observational cohort of individuals with and without knee OA where data on gait, physical activity, clinical and demographic measures are available for machine learning applications. Further, MOST includes MRI exams at multiple time points, providing sensitive measures of early joint structural change, including worsening cartilage damage.[10] Using MOST data, our objectives were to (1) build and evaluate a machine learning model to predict medial tibiofemoral cartilage worsening over 2 years from gait, physical activity, clinical and demographic features in individuals without advanced knee OA, and (2) identify features that contribute most to model prediction and quantify their effect on the outcome.

METHODS

Study sample

At 144 months, surviving participants from the original MOST cohort (age 50–79, with or at increased risk for developing knee OA at enrolment) were invited for a return visit. Concurrently, a new cohort (age 45–69, Kellgren-Lawrence grades (KLG) ≤2, with or without knee pain) was enrolled. Participants with inflammatory arthritis or stroke were not included in either cohort.

We used data from both cohorts for our baseline (original: 144 months, new: enrolment) and 2-year follow-up (original: 168 months, new: 24 months). MRIs were read for one knee per participant (herein referred to as the 'study knee') at baseline and 2 years. If both knees had readable baseline and follow-up images, the knee with better quality images was read. We excluded participants with KLG >2 in the study knee to focus on early disease (figure 1). We excluded participants with history of knee or hip replacement (either leg), steroid or hyaluronic acid injection during the past 6 months (either knee) or regular use of walking aids. Finally, we excluded participants who did not undergo MRI assessment or with gait or physical activity data quality issues (described later).

Patient and public involvement

Currently, patients and the public are not involved in the design, conduct, reporting or dissemination plans for research projects using MOST data.

Equity, diversity and inclusion statement

The authors include women and men with training in engineering and various clinical specialties from Asia, Europe and North America. This study included participants (58.2% women) from two North American clinical sites with various self-reported racial identities (table 1). Sex, site and race were accounted for in our analyses (described later); however, we did not examine socioeconomic status.

Exposures

Clinical and demographic features

Model inputs included clinical and demographic factors[11–18] that are both independent risk factors for OA and affect gait/physical activity (ie, confounders based on hypothesised directed acyclic graphs[19]). Sex, age, body mass index (BMI), race, clinic site and prior history of knee injury or surgery were recorded at baseline. Given small samples in multiple categories of race (table 1), particularly at UIowa, race and site were combined into a single feature with three levels: UAB non-white (n=117), UAB white (n=221), UIowa (n=609). Participants completed the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[20] and Center for Epidemiologic Studies Depression Scale, and had posterior–anterior and lateral weight-bearing radiographs taken, which were read for KLG.[21] Hip-knee-ankle alignment was read from baseline long-limb radiographs for the new cohort and long-limb radiographs taken at the 60-month visit for the original cohort. Pain during walking was extracted from the first question of WOMAC (categorised as 'no,' 'mild' or 'moderate or higher').

Gait features

Three-dimensional (3D) ground reaction force (GRF) data were recorded (1000 Hz) while participants walked at a self-selected speed across a portable force platform embedded in a 5.3 m walkway (AccuGait, AMTI, Watertown, Massachusetts, USA). At least five trials were acquired per leg, with the first excluded as an acclimatisation trial. Legs with ≥3 remaining trials where the foot landed completely on the force plate were retained for analysis. For each trial, we extracted commonly used 3D GRF metrics (figure 2), 'toe-out' angle defined by Chang *et al*,[22] stance time and walking speed. We normalised all timing features to stance phase (ie, % stance). GRFs were not amplitude-normalised given the inclusion of BMI in the model and to avoid issues with interpreting ratios.[23] We averaged each feature across trials for each leg.

Physical activity features

Participants wore an activity monitor (AX3, Axivity, Newcastle upon Tyne, UK) consisting of a triaxial accelerometer and temperature sensor on the lower back (centred over the midpoint of L5–S1) for 7 days at baseline, with 3D acceleration sampled at 100 Hz with a range of ±8 g. Non-wear was defined as periods ≥10 min with no movement and verified using the temperature sensor.[24] Data for each axis were bandpass filtered (0.2–20 Hz, 4th order Butterworth filter). Summary metrics were calculated for each day: step count, time spent walking, time spent lying and mean 3D signal vector magnitude (overall magnitude of acceleration across all dimensions, Equation 1). Time spent walking and lying were expressed as % wear time to account for differences in wear time among individuals.[25] Metrics were averaged across all valid days (defined as ≥10 hours of wear time/day[26]). We excluded participants with <3 valid days.[27]

$$\textit{Signal vector magnitude} = \sqrt{a_V^2 + a_{AP}^2 + a_{ML}^2} \quad (1)$$

Outcome

Two musculoskeletal radiologists (AG, FWR) scored the severity of cartilage damage in five medial tibiofemoral subregions of the

Table 1 Baseline demographics and clinical characteristics

Feature	Frequency, n (%)		Mean±SD	Median (IQR)
n participants	947			
Sex:				
Female	551 (58.2)			
Race:				
American Indian/Alaskan native	3 (0.3)			
Asian	8 (0.8)			
Black/African American	109 (11.5)			
Don't know/refused	1 (0.1)			
More than one race	7 (0.7)			
Other	11 (1.2)			
White/Caucasian	808 (85.3)			
Clinic site:				
University of Iowa	609 (64.3)			
Cohort:				
New	768 (81.1)			
Previous injury/surgery:				
Yes	178 (18.8)			
Medial tibiofemoral cartilage damage:				
Present	371 (39.2)			
Age (years)			59.2±8.3	59.0 (53.0 to 65.0)
Body mass index (kg/m^2)			27.8±4.8	27.2 (24.3 to 30.8)
CES-D (/60)			5.8±6.5	3.0 (1.0 to 9.0)
Hip-knee-ankle alignment (degrees, negative values indicate varus alignment)			−1.4±2.7	−1.3 (−3.2 to 0.20)
	Study knee	**Contralateral**		
Pain during walking*:				
None	751 (79.3%)	754 (79.6%)		
Mild	165 (17.4%)	161 (17.0%)		
Moderate or higher	31 (3.3%)	32 (3.4%)		
Kellgren-Lawrence Grade (KLG):				
KLG=0	582 (61.5%)	587 (62.0%)		
KLG=1	269 (28.4%)	248 (26.2%)		
KLG=2	96 (10.1%)	112 (11.8%)		

*Extracted from the first question of the Western Ontario and McMaster Universities Osteoarthritis Index.
CES-D, Center for Epidemiologic Studies Depression Scale.

study knee at each time point using the MRI Osteoarthritis Knee Score.[28] We defined medial cartilage worsening as any increase in area and/or depth in at least one of the five subregions over the 2-year period, as done previously.[10 29]

Machine learning model

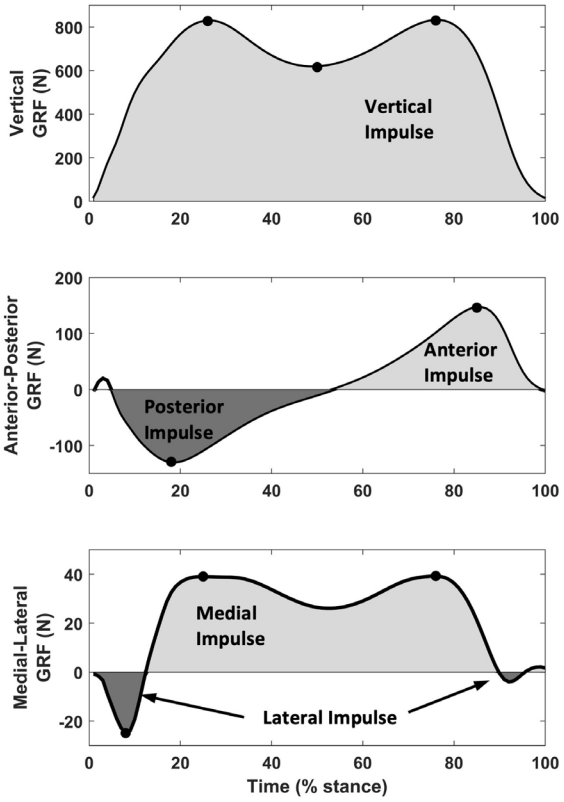
Model development was performed in R (V.4.2.2). We examined Spearman correlations between all continuous features and for near perfect correlations ($\rho > 0.85$), selected one feature to retain for analysis (table 2 shows retained gait and physical activity features). We used the predictive mean matching algorithm within the multiple imputation by chained equations framework (V.3.13.0) to impute missing exposure data (<0.1% dataset).[30] We randomly split the data into 70% train and 30% test, maintaining the same proportion of outcome in both datasets.[31] Continuous features were scaled and centred to have zero mean and unit variance.

Our goal was to predict the binary cartilage worsening outcome from baseline GRF, accelerometer and clinical/demographic data. We used 'super learning' (V.1.4.2),[32] an ensemble machine learning approach that combines several candidate algorithms to enhance prediction accuracy above and beyond individual algorithms (figure 3). We selected candidate learners to include diverse learning strategies while being computationally feasible, as recommended by Phillips *et al*.[33] Using the training dataset, candidate learners were trained through fivefold cross-validation. Corresponding predictions on out-of-fold samples were used to develop a meta learner that optimised the weight (ie, contribution) of each individual learner. We then applied this model to the held-out test set to assess its performance by area under the receiver operating characteristic curve (AUC) and mean squared error (MSE).

To test robustness and reproducibility of the model training and testing, we used repeated cross-validation, that is, repeated the process of randomly splitting the data into train and test, training the super learner and evaluating its performance on the held-out test set. Here, we report median (ie, 50th percentile), 2.5th and 97.5th percentile AUC and MSE across 100 iterations.

Identification of influential predictors

To assess the contribution of each feature to model prediction, for each of the 100 iterations, 35 additional models were trained on the training set (each excluding one of the features included in the full model). These models were applied to the test set and a variable importance measure (VIM) statistic was calculated for each feature for each iteration based on the size of the risk



Extracted Features:

- Local maxima/minima for each dimension (in N) [black circles]
- Timing of local maxima/minima (in % stance)
- Impulse for each dimension (in N*s) [gray shaded areas under the curves]
- Average amplitude of vertical GRF loading (in N/s) [average slope in early stance]
- Maximum amplitude of vertical GRF loading (in N/s) [maximum slope in early stance]
- Timing of maximum vertical GRF loading (in % stance)
- Average amplitude of vertical GRF unloading (in N/s) [average slope in late stance]
- Maximum amplitude of vertical GRF unloading (in N/s) [absolute maximum slope in late stance]
- Timing of maximum vertical GRF unloading (in % stance)

Figure 2 Features extracted from ground reaction force (GRF) data.

difference between the full model and the model fit without the feature. Thus, 35 VIMs were produced per iteration. The top contributors to prediction for each iteration were identified as the 10 features with the highest VIMs. We defined 'influential predictors' as the 10 features that most frequently appeared as top contributors across the 100 iterations.

Marginal causal risk differences

To quantify the effect of influential predictors identified from the super learner model on cartilage worsening, we used parametric g-computation.[34] Continuous variables were quantised into tertiles. For each predictor, we calculated the marginal causal risk difference of each category of the predictor on cartilage worsening, compared with the corresponding reference category, using risk Communicator (V.1.0.0); 95% CIs were calculated using 1000 bootstrap samples.[35] Different risk factors may be associated with OA initiation vs progression; thus, we explored sensitivity analyses stratified by baseline cartilage damage (ie, lesion in ≥1 subregion). Only 6% of knees without baseline damage had cartilage worsening at follow-up, thus, we focused our sensitivity analysis on those with baseline damage (online supplemental file).

RESULTS

Model performance

Of 947 participants, 133 (14%) experienced cartilage worsening in the study knee over 2 years. Across 100 iterations, the median (2.5th and 97.5th percentiles) AUC and MSE on the held-out test sets were 0.73 (0.65–0.79) and 0.11 (0.09–0.13), respectively.

Influential predictors

The features most frequently appearing as top contributors to prediction across 100 iterations (and frequency of appearance) were baseline medial tibiofemoral cartilage damage (100), KLG (98), lateral GRF impulse (46), pain during walking (45), time spent lying (35), timing of the vertical GRF first peak (31), vertical GRF impulse (30), early medial GRF peak (29), timing of the vertical GRF second peak (28) and the maximum instantaneous vertical GRF unloading rate (28).

Marginal risk differences

Marginal risk differences from the g-computation analyses (figure 4) can be interpreted as the difference in risk of cartilage worsening per 100 individuals in the given category compared with the referent category. Presence of cartilage damage, higher KLG, greater lateral GRF impulse, greater pain during walking, greater time spent lying and lower vertical GRF unloading rate at baseline were associated with increased risk of cartilage worsening (figure 4). Point estimates were similar in the sensitivity analysis (online supplemental file).

DISCUSSION

An ensemble machine learning approach incorporating baseline gait, physical activity and clinical/demographic features showed good performance predicting medial tibiofemoral cartilage worsening over 2 years in knees with KLG ≤2. While determining the relationships among predictors and outcomes in machine learning models is challenging, our analysis suggests that high lateral GRF impulse, high time spent lying, and low vertical GRF unloading should be investigated further as potential targets to reduce cartilage worsening.

Model performance

Our model performance is comparable to other machine learning models predicting OA progression from clinical/demographic data. Du *et al* reported AUCs of 0.70–0.79 for

Table 2 Baseline gait and physical activity features

Feature	Mean±SD	Median (IQR)
GRF impulses (N×s):		
Vertical GRF impulse	454.5±113.7	439.5 (370.5 to 523.5)
Medial GRF impulse	18.9±8.1	17.6 (13.0 to 23.1)
Lateral GRF impulse	1.6±1.0	1.5 (0.9 to 2.1)
Anterior GRF impulse	23.9±7.0	23.1 (18.5 to 28.3)
GRF local maxima (N):		
Vertical GRF first peak	850.6±179.5	826.9 (718.2 to 962.9)
Posterior GRF peak	134.1±41.1	128.5 (103.5 to 156.4)
Early medial GRF peak	47.6±15.8	45.7 (36.3 to 57.5)
GRF loading and unloading rates (N/s):		
Maximum instantaneous vertical GRF loading rate	11555±4077	10981 (8776 to 13379)
Maximum instantaneous vertical GRF unloading rate	10204±2289	9983 (8564 to 11536)
Timing of GRF local maxima/minima (% stance):		
Vertical GRF first peak	25.9±3.5	25.7 (23.8 to 28.0)
Vertical GRF second peak	75.5±2.5	75.9 (74.5 to 77.0)
Vertical GRF valley (midstance minimum)	48.3±3.5	48.3 (46.0 to 50.8)
Posterior GRF peak	17.5±2.4	17.5 (16.1 to 18.9)
Anterior GRF peak	85.4±1.4	85.5 (84.6 to 86.4)
Early lateral GRF peak	7.1±1.7	7.2 (6.1 to 8.2)
Early medial GRF peak	26.3±5.7	26.0 (22.2 to 30.1)
Late medial GRF peak	72.3±5.7	73.5 (70.0 to 76.3)
Spatiotemporal parameters:		
Gait speed (m/s)	1.35±0.20	1.34 (1.21 to 1.48)
Stance time (s)	0.7±0.1	0.7 (0.7 to 0.8)
Angle formed by centre of pressure path and direction of travel, 'toe-out angle' (degrees, negative values indicate varus)	−3.7±5.3	−3.5 (−7.2 to −0.2)
Accelerometer derived physical activity measures:		
Step count	9554±3919	9149 (6812 to 11533)
Time spent lying (% total wear time)	9.2±10.4	5.6 (2.0 to 12.5)
Mean signal vector magnitude (milligravity)	4.0±1.4	3.8 (3.1 to 4.7)

GRF, ground reaction force.

predicting radiographic worsening (increase in KLG, medial or lateral joint space narrowing) over 2 years from baseline cartilage damage MRI features in those with KLG 0 to 4.[36] Tiulpin *et al* reported AUCs of 0.73–0.75 for predicting worsening (increase in KLG or knee joint replacement) over 7 years from baseline age, sex, BMI, injury, surgery, WOMAC and KLG in individuals with KLG <2.[37] The current model achieved similar AUC for predicting cartilage worsening over 2 years in individuals with KLG ≤2, with the added benefit of providing information about potentially modifiable gait and physical activity predictors.

Prior longitudinal gait studies typically examined knee-specific loading (eg, knee adduction moment) rather than GRFs, often in samples of 15–300 knees.[5] Correspondingly, few addressed clinical/demographic confounders, incorporated physical activity or examined performance in held-out test sets. Further, many were conducted in samples with established OA (KLG ≥2), limiting their potential to identify at-risk individuals early in the disease process or identify early intervention targets. Our sample included 947 individuals with KLG ≤2, who predominantly had no or mild pain during walking, and thus were younger with lower BMI than previously reported samples (mean age 59.2 vs 62.0 years, BMI 27.8 vs 29.2 kg/m^2).[18]

Predictors of OA progression

The super learner identified multiple influential gait and physical activity predictors of cartilage worsening. Of these, the g-computation analyses found baseline lateral GRF impulse, time spent lying and vertical GRF unloading rate were associated with cartilage worsening. The 7.2% estimate of risk difference for lateral GRF impulse suggests that for every 100 individuals in the highest tertile, there are 7.2 individuals who experience cartilage worsening who would not experience worsening in they were in the lowest tertile. Accordingly, approximately 14 (ie, 1/0.072) individuals with lateral GRF impulse at the highest tertile would be needed to observe an increase in the number of individuals with cartilage worsening by 1 person. In a cross-sectional study in the same cohort, we previously reported that limbs with radiographic OA, with or without knee pain, have higher lateral GRFs in early stance compared with limbs without radiographic OA or pain.[38] The current results suggest lateral GRF may also play a role in progression. The 5.4% increased risk of cartilage worsening for the middle versus lowest tertile of time spent lying, along with prior research showing greater sedentary time is associated with future functional decline[39] and lower quality of life,[40] suggests reducing sedentary time should be investigated as a potential intervention target. The physiological reason for the 6.6% decreased risk of cartilage worsening in

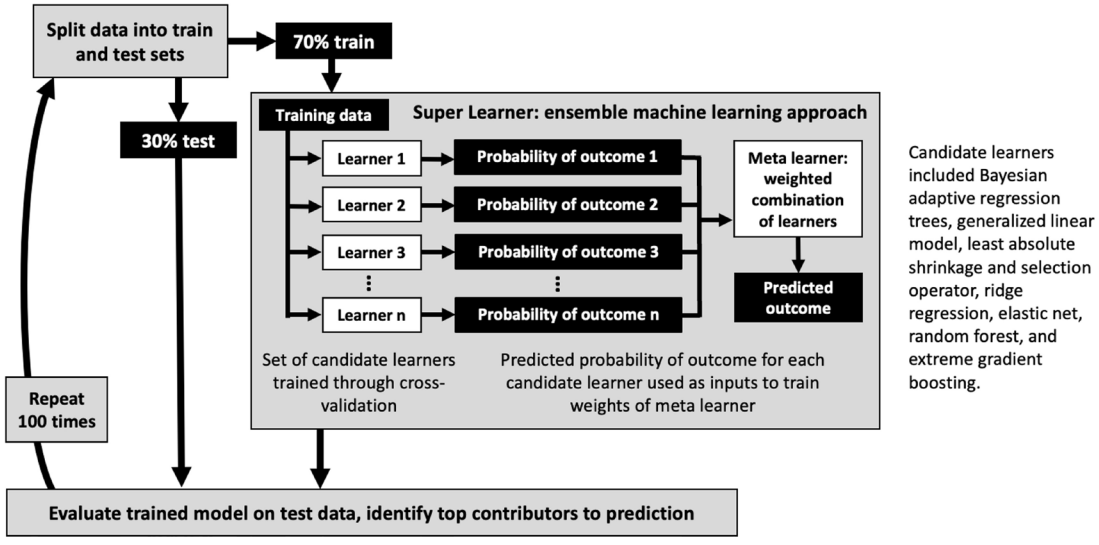


Figure 3 Machine learning model development and evaluation.

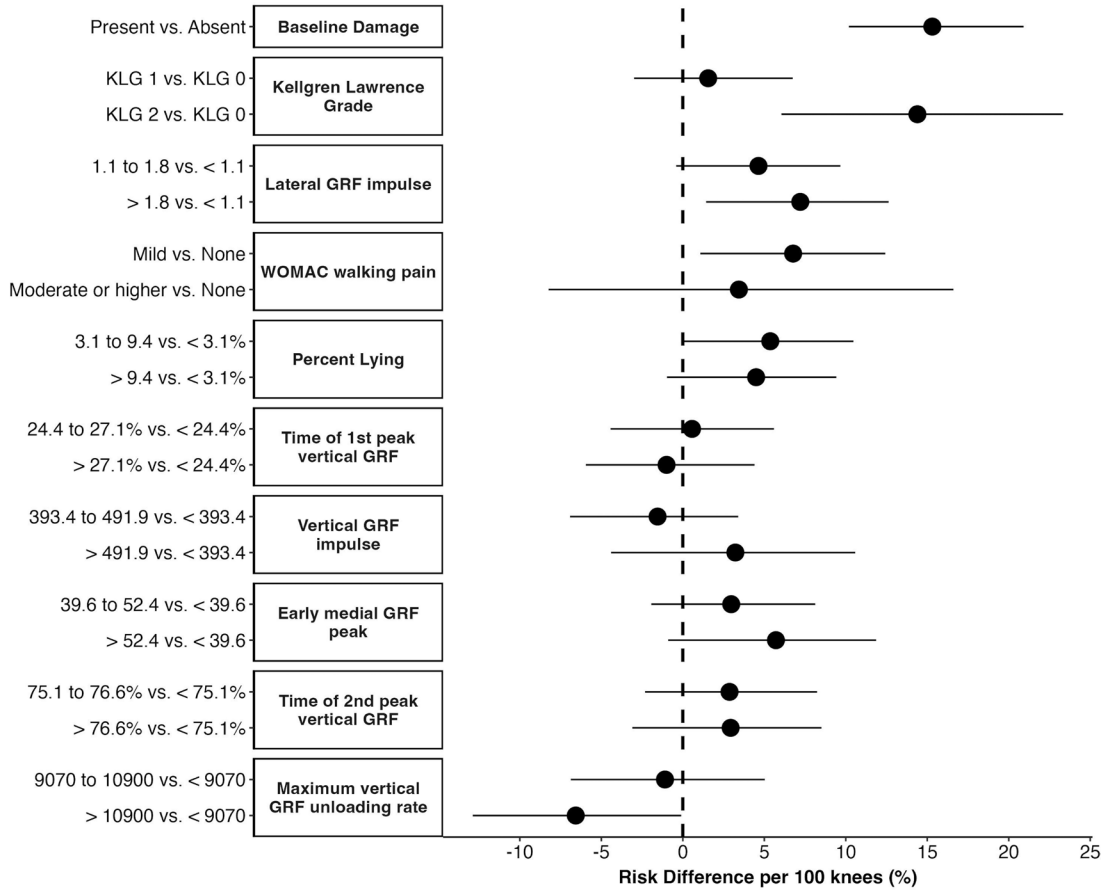


Figure 4 Causal risk differences for influential predictors identified from the machine learning model. GRF, ground reaction force; KLG, Kellgren-Lawrence grades; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

the highest vs lowest tertile of vertical GRF unloading rate is not clear and warrants further exploration.

The appearance of structure and symptom features as influential predictors is not surprising, given that these are established risk factors. Of note, despite only 10.1% of the sample having what is traditionally considered established radiographic OA (KLG=2), 39.2% had baseline cartilage damage and both appeared as influential predictors in the model. The g-computation analysis identified a 15.3% increased risk of cartilage worsening for every 100 individuals with baseline damage compared with no damage, and a 14.4% increased risk for KLG 2 vs 0. The lack of difference for KLG 1 vs 0 may highlight limitations of the KLG scoring system, which does not reflect tissue-level damage well, particularly in early disease.[41 42] Along with these structural measures, knees with mild pain had an increased risk of cartilage worsening (6.8%) compared with those with no pain. The large CI for moderate and higher pain could stem from the small proportion of knees (3% of sample) and/or heterogeneity in this category.

Clinical implications

The utility of this model for risk screening is debatable, as it requires GRF data. While faster to collect than joint moments, collecting GRFs requires specialised equipment (force platform). Future advances in wearable technologies may facilitate gait data capture during everyday life, including estimates of GRFs,[43 44] improving the potential of this type of model as a risk screening tool.

This model identified potential gait and physical activity intervention targets for further study. Interestingly, two influential predictors (baseline damage, KLG) appeared as top contributors in ≥98% of the iterations while others appeared less consistently (<50%). While we removed highly correlated features, this may result in part from predictors that capture similar constructs (eg, four features collectively describing an important construct could each appear 25% of the time). Similarly, our g-computation approach provides insight into causal pathways but does not account for concurrent changes in several risk factors. An important motivation for using machine learning was to address potential interactions among predictors. While it is challenging to identify these relationships from the model, the lack of consistency in top contributors could indicate a need for simultaneous intervention on several features rather than a single feature, opening interesting avenues for future study.

Strengths and limitations

Strengths include the large sample, investigation of gait and physical activity in early disease, use of machine learning to address inter-related predictors and use of g-computation to quantify their effects. These strengths expand existing literature by accounting for demographics and clinical characteristics, examining multiple gait and physical activity features, and testing the model on held-out data. Our sample was largely white with little to no pain during walking; these results may not generalise to diverse populations or those with severe symptoms. Lateral or patellofemoral worsening could have been present in both

outcome groups, causing noise in the model. While we adjusted for a diverse set of confounders, as in any observational study, there may be residual unmeasured confounding. Knee loading (eg, knee adduction moment) may provide additional predictive power but kinematics are not available in MOST, limiting comparison to prior gait studies and insight into mechanisms by which features such as lateral GRF impulse affect structure. We are unaware of other large datasets with gait, physical activity and MR outcomes that could be used for external validation, however, we assessed reproducibility with repeated cross-validation. Better characterisation of dynamic physical activity patterns may also improve model performance and identification of relevant intervention targets.

CONCLUSION

Using an ensemble machine learning approach, we predicted medial tibiofemoral cartilage worsening over 2 years in persons without and with early radiographic OA with good performance on held-out samples. Additionally, we identified baseline gait and physical activity measures associated with cartilage worsening that may be potential early intervention targets, including lateral GRF impulse, time spent lying and vertical GRF unloading rate.

Author affiliations
[1]Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida, USA
[2]Physical Therapy, Boston University, Boston, Massachusetts, USA
[3]Rheumatology, Boston University School of Medicine, Boston, Massachusetts, USA
[4]Radiology, VA Boston Healthcare System, West Roxbury, Massachusetts, USA
[5]Radiology, Universitatsklinikum Erlangen, Erlangen, Germany
[6]Radiology, Boston University School of Medicine, Boston, Massachusetts, USA
[7]Rehabilitation Medicine, The University of Kansas Medical Center, Kansas City, Kansas, USA
[8]Epidemiology, The University of Iowa, Iowa City, Iowa, USA
[9]Epidemiology, The University of Alabama, Birmingham, Alabama, USA
[10]Epidemiology & Biostatistics, University of California, San Francisco, California, USA
[11]Computational Biomedicine, Boston University School of Medicine, Boston, Massachusetts, USA
[12]Computer Science, Boston University, Boston, Massachusetts, USA

Twitter Kerry E Costello @kecostello, David T Felson @david_felson, S Reza Jafarzadeh @srjafarz, Ali Guermazi @ali_guermazi, Frank W Roemer @FrankRoemer1, Cara L Lewis @ProfCaraLewis, Vijaya B Kolachalama @vkola_lab and Deepak Kumar @ProfDeepakKumar

Acknowledgements The authors would like to acknowledge the contributions of the MOST participants and clinic staff.

Contributors Guarantors for the integrity of the work as a whole: KEC and DK. Conception and design: KEC, DK and DTF. Analysis and interpretation of the data: all authors. Drafting of the article: KEC. Critical revision of the article for important intellectual content: all authors. Final approval of the article: all authors. Provision of study materials or patients: DTF, CEL, MCN and NAS.

Funding MOST is composed of four cooperative grants (DTF (BU)—AG18820, J.C. Torner (UI)—AG18832, CEL (UAB)—AG18947, and MCN (UCSF)—AG19069] funded by the National Institutes of Health (NIH), a branch of the Department of Health and Human Services, and conducted by MOST study investigators. Research reported in this publication was also supported under award numbers F32AR076907 (KEC), P30AR072571 (DTF), 1UL1TR001430 (DTF), R21AR074578 (SRJ), R03AG060272 (S.R. Jafarzadeh), K23AR063235 (CLL), R01HL159620 (VBK), R21CA253498 (VBK), and K01AR069720 (DK) from the National Institutes of Health and 17SDG33670323 (VBK) and 20SFRN35460031 (VBK) from the American Heart Association. KEC was also supported by an Investigator Award from the Rheumatology Research Foundation. The MOST publications committee reviewed this manuscript before submission. The National Institutes of Health, the American Heart Association, and the Rheumatology Research Foundation were not involved in study design, collection, analysis or interpretation of data, or the decision to submit this manuscript for publication.

Disclaimer This manuscript was prepared using MOST data and does not necessarily represent the official views of MOST investigators, the National Institutes of Health, the American Heart Association, or the Rheumatology Research Foundation.

Competing interests NAS reports personal fees from Tenex Health and grants from Pacira Bioscience, outside of the submitted work. AG is shareholder of BICL, LLC and consultant to Pfizer, AstraZeneca, Novartis, TissueGene, Regeneron and MerckSerono. FWR is shareholder of BICL, LLC and consultant to Grünenthal. All other authors have no competing interests to report.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Ethical approval for the Multicenter Osteoarthritis Study was obtained from the Institutional Review Board at each of the four core sites: Boston University (H-32956), University of Alabama at Birmingham (IRB-000329007), University of California San Francisco (301480) and University of Iowa (201511711). All participants provided informed consent before participating, in accordance with the Helsinki Declaration. Approval for secondary data analysis using MOST data was obtained from the Institutional Review Board at the University of Florida (IRB202201899).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data from the Multicenter Osteoarthritis Study are available through the National Institute on Aging Research Biobank: https://agingresearchbiobank.nia.nih.gov/. Public use datasets from the 144- and 168-month visits of the Multicenter Osteoarthritis Study will be available at this website in 2023.

ORCID iDs
Kerry E Costello http://orcid.org/0000-0002-5928-1471
David T Felson http://orcid.org/0000-0002-2668-2447
S Reza Jafarzadeh http://orcid.org/0000-0002-1099-9175
Ali Guermazi http://orcid.org/0000-0002-9374-8266
Frank W Roemer http://orcid.org/0000-0001-9238-7350
Neil A Segal http://orcid.org/0000-0002-8294-080X
Cara L Lewis http://orcid.org/0000-0002-9888-4902
Vijaya B Kolachalama http://orcid.org/0000-0002-5312-8644
Deepak Kumar http://orcid.org/0000-0003-1728-484X

REFERENCES

1 Vos T, Lim SS, Abbafati C, *et al*. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 2020;396:1204–22.
2 Cerejo R, Dunlop DD, Cahue S, *et al*. The influence of alignment on risk of knee osteoarthritis progression according to baseline stage of disease. *Arthritis Rheum* 2002;46:2632–6.
3 Cibere J, Sayre EC, Guermazi A, *et al*. Natural history of cartilage damage and osteoarthritis progression on magnetic resonance imaging in a population-based cohort with knee pain. *Osteoarthritis Cartilage* 2011;19:683–8.
4 Griffin TM, Guilak F. The role of mechanical loading in the onset and progression of osteoarthritis. *Exerc Sport Sci Rev* 2005;33:195–200.
5 D'Souza N, Charlton J, Grayson J, *et al*. Are biomechanics during gait associated with the structural disease onset and progression of lower limb osteoarthritis? A systematic review and meta-analysis. *Osteoarthritis Cartilage* 2022;30:381–94.
6 Fan J, Han F, Liu H. Challenges of big data analysis. *Natl Sci Rev* 2014;1:293–314.
7 Costello KE, Astephen Wilson JL, Hubley-Kozey CL. Association of low physical activity levels with gait patterns considered at risk for clinical knee osteoarthritis progression. *ACR Open Rheumatol* 2021;3:753–63.
8 Jamshidi A, Pelletier JP, Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. *Nat Rev Rheumatol* 2019;15:49–60.
9 Segal NA, Nevitt MC, Gross KD, *et al*. The multicenter osteoarthritis study: opportunities for rehabilitation research. *PM R* 2013;5:647–54.
10 Roemer FW, Nevitt MC, Felson DT, *et al*. Predictive validity of within-grade scoring of longitudinal changes of MRI-based cartilage morphology and bone marrow lesion

assessment in the tibio-femoral joint -- the most study. *Osteoarthritis Cartilage* 2012;20:1391–8.
11 McKean KA, Landry SC, Hubley-Kozey CL, *et al*. Gender differences exist in osteoarthritic gait. *Clin Biomech (Bristol, Avon)* 2007;22:400–9.
12 Duffell LD, Jordan SJ, Cobb JP, *et al*. Gait adaptations with aging in healthy participants and people with knee-joint osteoarthritis. *Gait Posture* 2017;57:246–51.
13 Harding GT, Hubley-Kozey CL, Dunbar MJ, *et al*. Body mass index affects knee joint mechanics during gait differently with and without moderate knee osteoarthritis. *Osteoarthritis Cartilage* 2012;20:1234–42.
14 Sims EL, Carland JM, Keefe FJ, *et al*. Sex differences in biomechanics associated with knee osteoarthritis. *J Women Aging* 2009;21:159–70.
15 Robbins SM, Pelletier J-P, Abram F, *et al*. Gait risk factors for disease progression differ between non-traumatic and post-traumatic knee osteoarthritis. *Osteoarthritis Cartilage* 2021;29:1487–97.
16 Song J, Chang AH, Chang RW, *et al*. Relationship of knee pain to time in moderate and light physical activities: data from osteoarthritis initiative. *Semin Arthritis Rheum* 2018;47:683–8.
17 Song J, Hochberg MC, Chang RW, *et al*. Racial and ethnic differences in physical activity guidelines attainment among people at high risk of or having knee osteoarthritis. *Arthritis Care Res (Hoboken)* 2013;65:195–202.
18 Dunlop DD, Song J, Semanik PA, *et al*. Objective physical activity measurement in the osteoarthritis initiative: are guidelines being Met? *Arthritis Rheum* 2011;63:3372–82.
19 Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–25.
20 Bellamy N, Buchanan WW, Goldsmith CH, *et al*. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833–40.
21 Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957;16:494–502.
22 Chang A, Hurwitz D, Dunlop D, *et al*. The relationship between toe-out angle during gait and progression of medial tibiofemoral osteoarthritis. *Ann Rheum Dis* 2007;66:1271–5.
23 Kronmal RA. Spurious correlation and the fallacy of the ratio standard revisited. *J R Stat Soc Ser A Stat Soc* 1993;156:379.
24 Zhou S-M, Hill RA, Morgan K, *et al*. Classification of accelerometer wear and non-wear events in seconds for monitoring free-living physical activity. *BMJ Open* 2015;5:e007447.
25 Aadland E, Ylvisåker E. Reliability of objectively measured sedentary time and physical activity in adults. *PLoS One* 2015;10:e0133296.
26 Song J, Semanik P, Sharma L, *et al*. Assessing physical activity in persons with knee osteoarthritis using accelerometers: data from the osteoarthritis initiative. *Arthritis Care Res (Hoboken)* 2010;62:1724–32.
27 Matthews CE, Ainsworth BE, Thompson RW, *et al*. Sources of variance in daily physical activity levels as measured by an accelerometer. *Med Sci Sports Exerc* 2002;34:1376–81.
28 Hunter DJ, Guermazi A, Lo GH, *et al*. Evolution of semi-quantitative whole joint assessment of knee oa: MOAKS (MRI osteoarthritis knee score). *Osteoarthritis Cartilage* 2011;19:990–1002.
29 Voinier D, Neogi T, Stefanik JJ, *et al*. Using cumulative load to explain how body mass index and daily walking relate to worsening knee cartilage damage over two years: the most study. *Arthritis Rheumatol* 2020;72:957–65.
30 van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011;45.
31 van den Goorbergh R, van Smeden M, Timmerman D, *et al*. The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. *J Am Med Inform Assoc* 2022;29:1525–34.
32 Coyle JR, Hejazi NS, Malenica I, *et al*. SI3: modern pipelines for machine learning and super learning, p. R package version 1.4.2. n.d. Available: https://github.com/tlverse/sl32021
33 Phillips RV, van der Laan MJ, Lee H, *et al*. Practical considerations for specifying a super learner. *ArXiv* 2022.
34 Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol* 2011;173:731–8.
35 Grembi JA, Rogawski McQuade ET. Introducing riskcommunicator: an R package to obtain interpretable effect estimates for public health. *PLoS One* 2022;17:e0265368.
36 Du Y, Almajalid R, Shan J, *et al*. A novel method to predict knee osteoarthritis progression on MRI using machine learning methods. *IEEE Trans Nanobioscience* 2018;17:228–36.
37 Tiulpin A, Klein S, Bierma-Zeinstra SMA, *et al*. Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Sci Rep* 2019;9:20038.
38 Costello KE, Felson DT, Neogi T, *et al*. Ground reaction force patterns in knees with and without radiographic osteoarthritis and pain: descriptive analyses of a large cohort (the multicenter osteoarthritis study). *Osteoarthr Cartil* 2021;29:1138–46.
39 Semanik PA, Lee J, Song J, *et al*. Accelerometer-monitored sedentary behavior and observed physical function loss. *Am J Public Health* 2015;105:560–6.
40 Pinto D, Song J, Lee J, *et al*. Association between sedentary time and quality of life from the osteoarthritis initiative: who might benefit most from treatment? *Arch Phys Med Rehabil* 2017;98:2485–90.
41 Luyten FP, Denti M, Filardo G, *et al*. Definition and classification of early osteoarthritis of the knee. *Knee Surg Sports Traumatol Arthrosc* 2012;20:401–6.
42 Brandt KD, Fife RS, Braunstein EM, *et al*. Radiographic grading of the severity of knee osteoarthritis: relation of the kellgren and Lawrence grade to a grade based on joint space narrowing, and correlation with arthroscopic evidence of articular cartilage degeneration. *Arthritis Rheum* 1991;34:1381–6.
43 Jacobs DA, Ferris DP. Estimation of ground reaction forces and ankle moment with multiple, low-cost sensors. *J Neuroeng Rehabil* 2015;12:90.
44 Tan T, Chiasson DP, Hu H, *et al*. Influence of IMU position and orientation placement errors on ground reaction force estimation. *J Biomech* 2019;97:109416.

Supplement for:

Gait, physical activity, and tibiofemoral cartilage damage: A longitudinal machine learning analysis in the Multicenter Osteoarthritis Study

Kerry E. Costello[1,2,3], David T. Felson[2], S. Reza Jafarzadeh[2], Ali Guermazi[2], Frank W. Roemer[2,4], Neil A. Segal[5,6], Cora E. Lewis[7], Michael C. Nevitt[8], Cara L. Lewis[1,2], Vijaya B. Kolachalama[1,2], Deepak Kumar[1,2]

[1] Boston University, Boston, MA, USA
[2] Boston University School of Medicine, Boston, MA, USA
[3] University of Florida, Gainesville, FL, USA
[4] University of Erlangen, Erlangen, Germany
[5] University of Kansas Medical Center, Kansas City, KS, USA
[6] The University of Iowa, Iowa City, IA, USA
[7] University of Alabama at Birmingham, Birmingham, AL, USA
[8] University of California at San Francisco, San Francisco, CA, USA

Corresponding author:

Deepak Kumar, PT, Ph.D.
635 Commonwealth Avenue, Boston, MA 02215
kumard@bu.edu

Sensitivity analysis examining subsample with baseline cartilage damage

In the subsample of knees with baseline cartilage damage (Table S1), 26% had cartilage worsening at 2-year follow-up. For each predictor, we calculated the marginal causal risk difference of each category of the predictor on cartilage worsening, compared to the corresponding reference category using g-computation. Continuous variables were categorized using tertile cutpoints calculated from the full sample (as detailed in the main manuscript). The models included the same predictors as in the main manuscript except for baseline cartilage damage (i.e., 9 total predictors included).

As in the main analysis, in the subsample with baseline cartilage damage, the g-computation analysis identified an increased risk of cartilage worsening for individuals with KLG 2 versus 0 (17.9% per 100 individuals) and for pain during walking of mild versus none (16.3% per 100 individuals) (Figure S1). In the main analysis a lateral ground reaction force (GRF) impulse of 1.8 N*s or higher compared to <1.1 N*s had a higher risk of cartilage worsening. In the subsample, the point estimate was similar to the main analysis (6.1% versus 7.2% per 100 individuals), but the 95% confidence interval included zero. Similarly, point estimates were similar for the middle versus lowest tertile of time spent lying (7.1% versus 5.4% per 100 individuals) and for the highest versus lowest tertile of maximum vertical GRF unloading rate (7.8% versus 6.6%) but both 95% confidence intervals included zero. These wider confidence

intervals could be related to the smaller sample size of this subsample or heterogeneity within the subsample.

Table S1. Baseline demographics and clinical characteristics for subsample with baseline cartilage damage

Feature	**Frequency, n (%)**		**Mean ± SD**
n participants	371		
Sex:			
Female	183 (49.3%)		
Race:			
American Indian or Alaskan Native	1 (0.3%)		
Asian	3 (0.8%)		
Black or African American	36 (9.7%)		
Don't know/Refused	0 (0.0%)		
More than one race	3 (0.8%)		
Other	3 (0.8%)		
White or Caucasian	325 (87.6%)		
Clinic Site:			
University of Iowa	254 (66.0%)		
Cohort:			
New	268 (72.2%)		
Previous injury/surgery:			
Yes	89 (32.8%)		
Age (years)			61.2 ± 8.6
Body Mass Index (kg/m^2)			28.2 ± 4.8
Center for Epidemiologic Studies Depression score (/60)			5.3 ± 5.4
Hip-knee-ankle alignment (degrees, negative values indicate varus alignment)			-1.9 ± 2.7
	Study knee	*Contralateral*	
WOMAC pain during walking:			
None	293 (79.0%)	294 (79.2%)	
Mild	62 (16.7%)	62 (16.7%)	
Moderate or higher	16 (4.3%)	15 (4.0%)	
Kellgren-Lawrence Grade (KLG):			
KLG = 0	156 (42.0%)	171 (46.1%)	
KLG = 1	136 (36.7%)	130 (35.0%)	
KLG = 2	79 (21.3%)	70 (18.9%)	

SD = standard deviation; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index

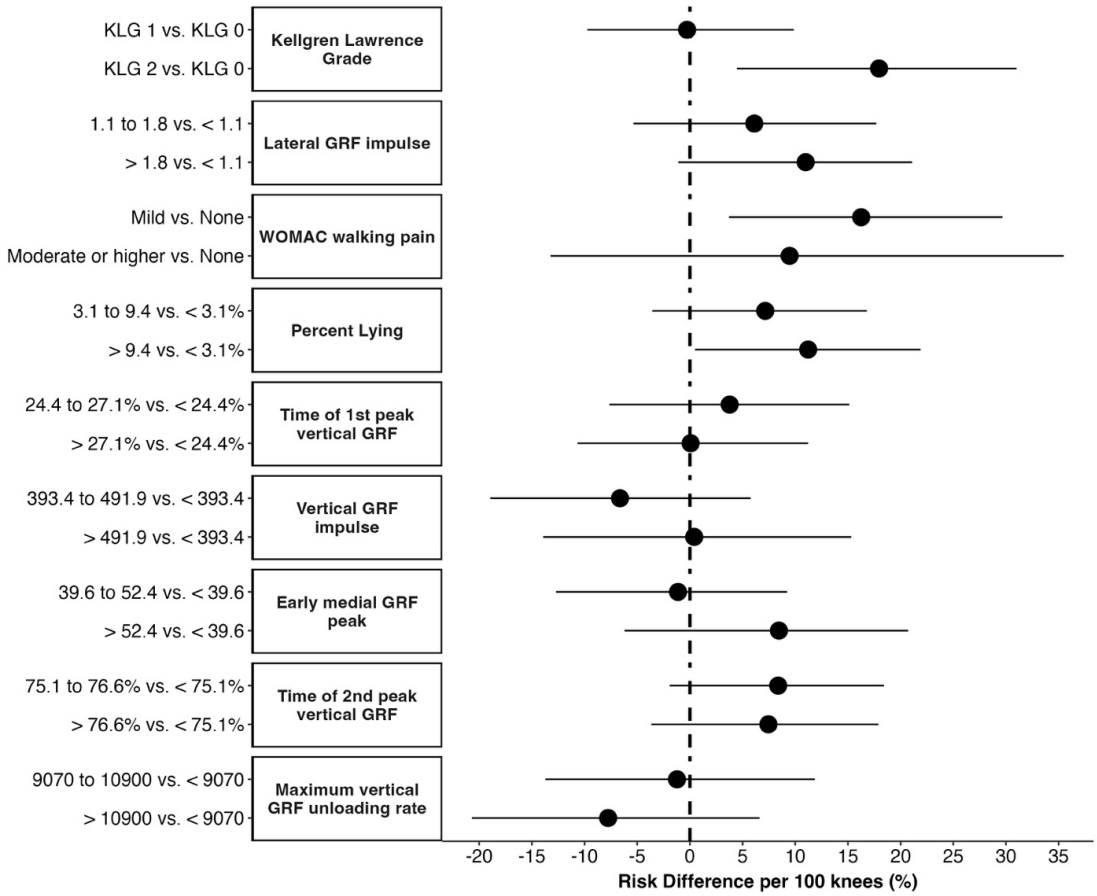


Figure S1. Causal risk differences in the subsample with baseline cartilage damage for influential predictors identified from the machine learning model