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

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

[1] *Boston University, Boston, MA, USA*

[2] *Boston University School of Medicine, Boston, MA, USA*

[3] *University of Erlangen, Erlangen, Germany*

[4] *University of Kansas Medical Center, Kansas City, KS, USA*

[5] *The University of Iowa, Iowa City, IA, USA*

[6] *University of Alabama at Birmingham, Birmingham, AL, USA*

[7] *University of California at San Francisco, San Francisco, CA, USA*

[8] *Department of Computer Science and Faculty of Computing & Data Sciences, Boston, MA, USA*

CORRESPONDING AUTHOR

Deepak Kumar, PT, Ph.D.

635 Commonwealth Avenue, Boston, MA 02215

kumard@bu.edu

KEYWORDS

knee, ground reaction force, accelerometer, risk difference

ABSTRACT

Objective: To 1) develop and evaluate a machine learning model incorporating gait and physical activity to predict medial tibiofemoral cartilage worsening over two years in individuals without or with early 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 2-year follow-up from gait, physical activity, clinical and demographic data from the Multicenter Osteoarthritis Study. Model performance was evaluated with Area Under the Curve (AUC) in repeated cross-validations. The top 10 influential predictors of the outcome across 100 held-out test sets were identified by a variable importance measure statistic, and their marginal effect on the outcome was quantified by g-computation.

Results: Of 947 legs included in the analysis, 14% experienced medial cartilage worsening over two years. The median (2.5^{th}-97.5^{th} percentile) AUC across the 100 held-out test sets was 0.73 (0.64-0.80). Presence of baseline cartilage damage, higher Kellgren-Lawrence grade, greater pain during walking, and higher lateral ground reaction force impulse were associated with greater risk of cartilage worsening.

Conclusions: An ensemble machine learning approach incorporating gait, physical activity, and clinical/demographic features showed good performance for predicting cartilage worsening over two years. While identifying potential intervention targets from the machine learning model is challenging, these results suggest that addressing high lateral ground reaction force impulse should be investigated further as a potential target to reduce medial tibiofemoral cartilage worsening in persons without or with early knee osteoarthritis.

SUMMARY BOX

- **What is already known on this topic:** Gait and physical activity features are potential intervention targets to slow knee osteoarthritis (OA) progression but their role in early OA has not been studied and underlying interactions among these features make traditional statistical approaches challenging.
- **What this study adds:** Using machine learning, this study predicted cartilage worsening in persons without or with early knee osteoarthritis from gait, physical activity, and clinical and demographic features with good performance on independent test sets. Further, walking with high lateral ground reaction force impulse was associated with 5.5% higher risk of cartilage worsening over two years compared to lower lateral impulse.
- **How this study might affect research, practice or policy:** Addressing high lateral ground reaction force impulse may be a potential target for interventions to slow cartilage worsening.

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 for OA and many live with pain and poor quality of life for decades. Additionally, existing structural damage and other risk factors (e.g., obesity, malalignment) can drive further degeneration.[2, 3] Addressing this burden will require both early identification of individuals at risk and discovery of potential targets for interventions that can be implemented prior to the onset of extensive damage or other risk factors.

Mechanical loading on the joint is one of the only modifiable risk factors for knee OA[4] and can be manipulated through gait and/or physical activity interventions. While prior research has identified key gait features associated with medial tibiofemoral knee OA progression,[5] these have typically been examined in isolation, in small samples, and/or without accounting for other known clinical/demographic risk factors. Importantly, little is known about gait and physical activity predictors of progression in individuals early in the disease process. Machine learning can help identify features among larger complex constructs that are important to prediction without requiring assumptions about underlying relationships among features, making it useful for exploring gait and physical activity data.[6-8]

The Multicenter Osteoarthritis Study (MOST)[9] is a unique, large, observational cohort of individuals with and without knee OA where data on gait, physical activity, clinical, and demographic measures are available for application of machine learning approaches. Further, MOST includes MRI knee exams at multiple timepoints, providing sensitive measures of early joint structural change, such as 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 or with early knee OA, and (2) identify features that

contributed most to prediction from the machine learning model and quantify their effect on the cartilage worsening 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 enrollment) were invited to return for a clinic visit and concurrently, a new cohort (age 45-69, with or without knee pain, and with Kellgren-Lawrence radiographic grades ≤ 2) was enrolled. Participants with inflammatory disease or stroke were not included in either cohort. The MOST study received institutional review board approval from the four core sites (Boston University, University of Alabama at Birmingham [UAB], University of California San Francisco, University of Iowa [UIowa]). All participants provided informed consent prior to participating, in accordance with the Helsinki Declaration.

We used data from both cohorts for our study baseline (original cohort: 144-month, new cohort: enrollment visit) and 2-year follow-up (original cohort: 168-month, new cohort: 24-month). MRIs were read for one knee per participant (herein referred to as the “study knee”) at both baseline and 2-years. We excluded participants with Kellgren-Lawrence grades (KLG) > 2 in the study knee to focus on early disease (Figure 1). We also excluded participants with history of knee or hip replacement in either leg, steroid or hyaluronic acid injection during the past 6 months in either knee, or regular use of a walking aid. Finally, we excluded participants who did not undergo MRI assessment and participants with data quality issues related to gait and/or physical activity measures (described later).

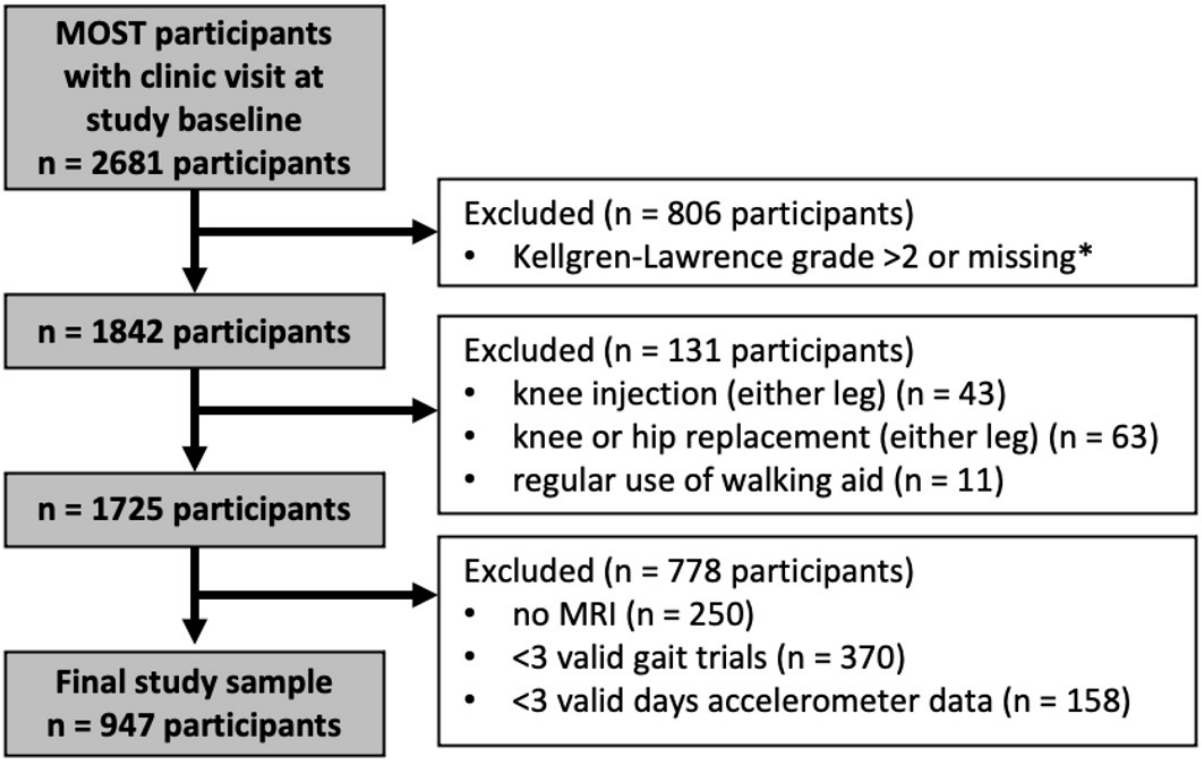

Figure 1. Study sample from the Multicenter Osteoarthritis Study

Patient and public involvement

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

Exposures

Clinical and demographic features

Clinical and demographic factors that are both independent risk factors for OA and affect gait and/or physical activity independent of OA (i.e., confounders) were included as model inputs.[11-18] Sex, age, body mass index (BMI), race, clinic site, and prior history of knee injury or surgery were recorded at baseline. Given small sample sizes in multiple categories of race, particularly at the UIowa clinic site, race and site were combined into a single feature with 3-

levels: UAB non-White, UAB White, and UIowa. Participants also completed the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[19] and Center for Epidemiologic Studies Depression Scale (CES-D), and had posterior-anterior and lateral weight bearing radiographs taken, which were read for KLG [20] at the MOST analysis center. Hip-knee-ankle alignment for the new cohort was read from long-limb radiographs taken at baseline. Long limb radiographs were not acquired for the original cohort at the 144-month visit (current baseline), thus we used hip-knee-ankle alignment read from long limb radiographs taken at the 60-month visit for these participants. WOMAC pain during walking was extracted from the first question of the WOMAC questionnaire, with legs categorized as 'no,' 'mild,' or 'moderate or higher' pain.

Table 1. Baseline demographics and clinical characteristics

Feature	**Frequency, n (%)**	**Mean ± SD**
n participants	947	
Sex:		
Female	551 (58.2%)	
Race:		
American Indian or Alaskan Native	3 (0.3%)	
Asian	8 (0.8%)	
Black or African American	109 (11.5%)	
Don't know/Refused	1 (0.1%)	
More than one race	7 (0.7%)	
Other	11 (1.2%)	
White or 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
Body Mass Index (kg/m^2)		27.8 ± 4.8

Center for Epidemiologic Studies Depression score (/60)			5.8 ± 6.5
Hip-knee-ankle alignment (degrees, negative values indicate varus alignment)			-1.4 ± 2.7
	Study knee	*Contralateral*	
WOMAC 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%)	

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

Gait features

Three-dimensional (3D) ground reaction force (GRF) data were recorded (1000 Hz) while participants walked at self-selected speed across a portable force platform embedded in a 5.3-meter walkway (AccuGait, AMTI Inc., Watertown, MA, USA). At least five trials of data were acquired per leg, with the first excluded as an acclimatization trial. Legs with at least three remaining trials where the foot landed completely on the force plate were retained for analysis. For each trial, we extracted a selection of gait features: commonly used discrete metrics calculated from the 3D GRF waveforms (Figure 2), “toe-out” angle as defined by Chang et al.,[21] stance time, and walking speed. We time-normalized all timing features to the stance phase of gait. We then averaged each feature across trials for each leg (Table 2). GRFs were not amplitude-normalized given the inclusion of BMI in the model and to avoid issues with interpreting ratios.[22]

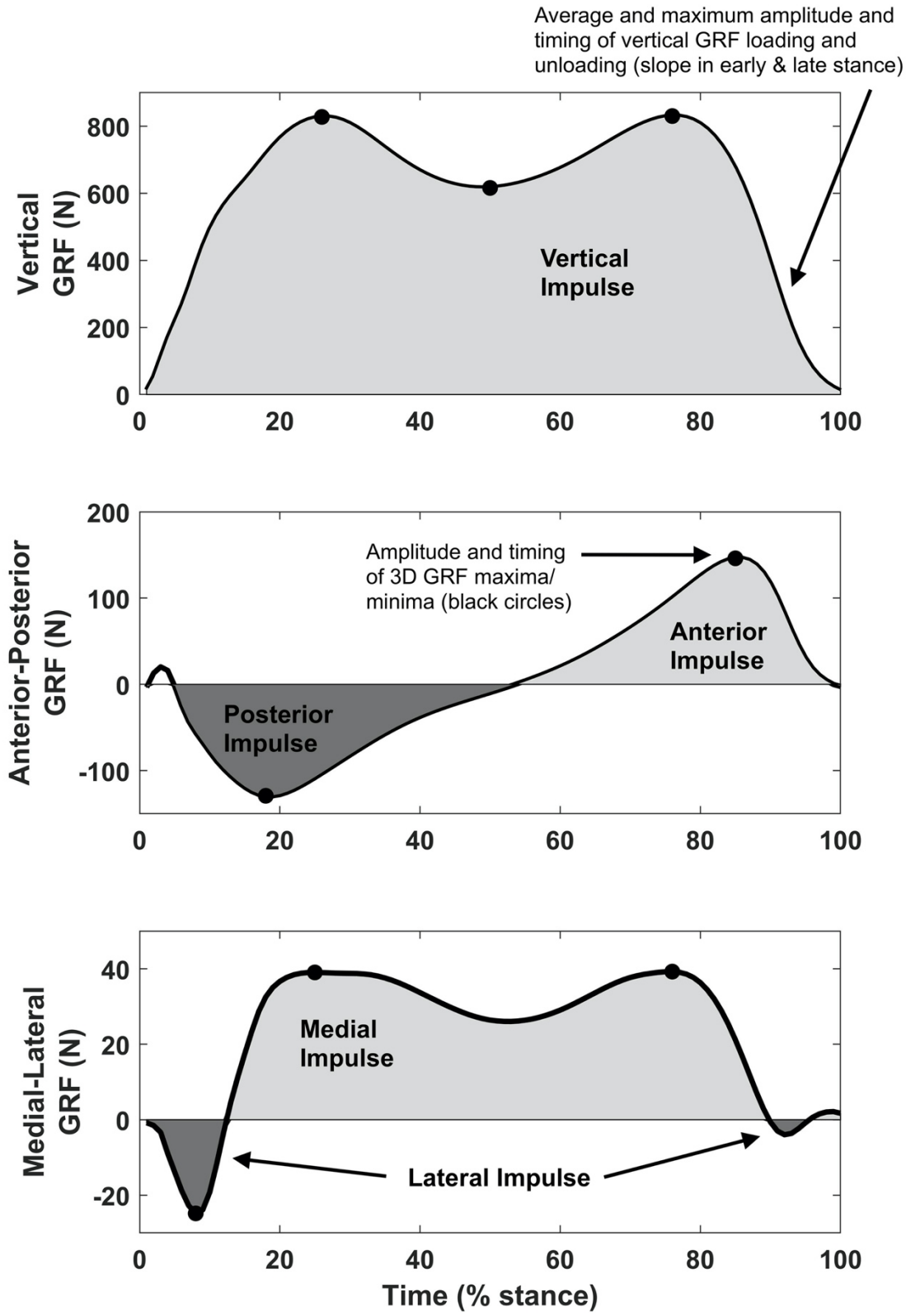

Figure 2. Features extracted from ground reaction force (GRF) data

Physical activity features

Participants wore an activity monitor (AX3, Axivity, Newcastle upon Tyne, UK) consisting of a tri-axial accelerometer and temperature sensor on the lower back (centered over the midpoint of L5-S1) for 7 days at baseline, with 3D acceleration data sampled at 100 Hz with a range of ±8g. Accelerometer data were extracted and processed at a centralized reading center. At least 10 hours of wear time were required for a day to be considered valid.[23] Non-wear was defined as periods ≥10 minutes with no movement (standard deviation < threshold) and verified using the temperature sensor in the device (change from average temperature > threshold). Summary metrics were calculated for each day, including time spent walking, time spent lying, and the mean 3D signal vector magnitude, which describes the overall magnitude of acceleration across all three dimensions (Equation 1). Time spent walking and lying were expressed as percent of wear time to account for possible differences in minutes of accelerometer wear among individuals.[24] Metrics were averaged across all valid days per participant (Table 2). We excluded participants with less than 3 valid days of data.[25]

$$Signal\ vector\ magnitude = \sqrt{a_v^2 + a_{AP}^2 + a_{ML}^2} \qquad \textit{Eq. 1}$$

Table 2. Baseline gait and physical activity features

Feature	Mean ± SD
GRF impulses (N*s):	
Vertical GRF impulse	454.5 ± 113.7
Medial GRF impulse	18.9 ± 8.1
Lateral GRF impulse	1.6 ± 1.0
Anterior GRF impulse	23.9 ± 7.0
Posterior GRF impulse	24.2 ± 7.2
GRF local maxima (N):	
Vertical GRF 1st peak	850.6 ± 179.5
Posterior GRF peak	134.1 ± 41.1
Early lateral GRF peak	29.5 ± 14.4

GRF loading and unloading rates (N/s):	
Maximum instantaneous vertical GRF loading rate	11555 ± 4077
Maximum instantaneous vertical GRF unloading rate	-10204 ± 2289
Timing of GRF local maxima/minima (% stance):	
Vertical GRF 1st peak	25.9 ± 3.5
Vertical GRF 2nd peak	75.5 ± 2.5
Vertical GRF valley (midstance minimum)	48.3 ± 3.5
Posterior GRF peak	17.5 ± 2.4
Anterior GRF peak	85.4 ± 1.4
Early lateral GRF peak	7.1 ± 1.7
Early medial GRF peak	26.3 ± 5.7
Late medial GRF peak	72.3 ± 5.7
Spatiotemporal parameters:	
Gait speed (m/s)	1.35 ± 0.2
Stance time (s)	0.7 ± 0.1
Angle formed by center of pressure path and direction of travel, "toe-out angle" (degrees, negative values indicate varus)	-3.7 ± 5.3
Accelerometer derived physical activity measures:	
Time spent walking (% total wear time)	12.3 ± 4.8
Time spent lying (% total wear time)	9.2 ± 10.4
Mean signal vector magnitude (milligravity)	4.0 ± 1.4

GRF = ground reaction force; SD = standard deviation

Outcome

Two musculoskeletal radiologists (AG, FWR) scored the severity of cartilage damage in 5 medial tibiofemoral subregions of the study knee at each timepoint using the MRI Osteoarthritis Knee Score (MOAKS)[26]. We defined medial cartilage worsening as any within-grade or ≥ full grade increase in area and/or depth in at least one of the 5 medial tibiofemoral cartilage subregions over the 2-year period, as done previously.[10, 27]

Machine learning model

Data preparation, model development, and evaluation were performed in R (v4.0.5). We examined Pearson correlations between all continuous features and in the case of near perfect

correlations ($r > 0.85$), selected one feature to retain for analysis. We used the predictive mean matching algorithm within the multiple imputation by chained equations (MICE) framework (v3.13.0) to impute missing exposure data (<0.1%).[28] Continuous features were scaled and centered. We randomly split the data into 70% training data and 30% test data without altering the proportion of outcome in both the training and test sets.[29]

Our goal was to predict the binary cartilage worsening outcome from the GRF, accelerometer, and clinical/demographic data. We used “super learning” (v1.4.2),[30] an ensemble machine learning approach that combines several candidate machine learning algorithms to enhance prediction accuracy above and beyond individual algorithms (Figure 3). Our candidate learners included Bayesian adaptive regression trees, generalized linear model, least absolute shrinkage and selection operator, ridge regression, elastic net, random forest, and extreme gradient boosting. Candidate learners were trained through 5-fold cross validation, and the corresponding predictions on the out-of-fold samples were used to develop a meta learner that optimized the weight (i.e., contribution) of each individual learner. We then cross-validated this trained model by applying it to the held-out test set to assess its performance by the area under the receiver operating characteristic curve (AUC).

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

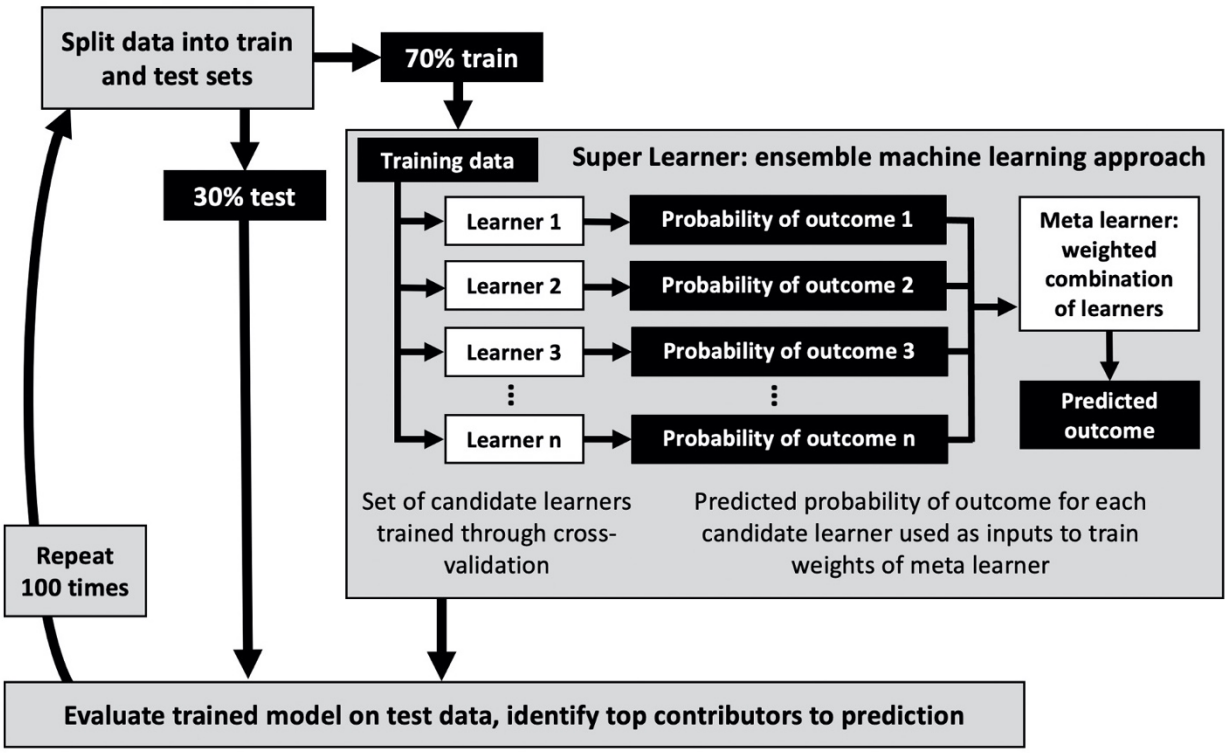

Figure 3. Machine learning model development and evaluation

Identification of influential predictors and marginal causal risk differences

A Variable Importance Measure (VIM) statistic was calculated for each feature for each data split based on the size of the risk difference between the model fit with and without the feature. Thus, for each split 37 VIMs were produced. The top contributors to prediction for each split 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 splits.

To quantify the effect of the influential predictors identified by the super learner model on cartilage worsening, we utilized the parametric g-computation method [31]. Continuous variables were quantized into tertiles. 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 riskCommunicator (v 1.0.0). Different risk factors may be associated with OA initiation versus progression; thus, we explored sensitivity analyses

stratified by baseline cartilage damage (i.e., lesion in ≥ 1 medial tibiofemoral subregion). A small proportion (6%) of knees without baseline damage had cartilage worsening at follow-up, thus we focused our sensitivity analysis on knees with baseline damage (Supplement).

RESULTS

Model performance

Of 947 participants (KLG ≤ 2) included in this study, 133 (14%) experienced medial tibiofemoral cartilage worsening in the study knee over 2-years. Across 100 splits, the median AUC (2.5- and 97.5-th percentiles) on the held-out test sets was 0.73 (0.64-0.80).

Influential predictors and marginal risk differences

The features most frequently appearing as top contributors to prediction across 100 data splits (and frequency of appearance) were baseline medial tibiofemoral cartilage damage (100), KLG (97), lateral GRF impulse (43), WOMAC pain during walking (39), time spent walking (33), vertical GRF impulse (32), gait speed (29), vertical GRF 1st peak (28), time spent lying (27), and timing of early lateral GRF peak (26). Marginal risk differences (Figure 4) can be interpreted as the difference in risk of cartilage worsening per 100 individuals in the given category compared to the referent category. Presence of baseline cartilage damage, higher KLG, greater lateral GRF impulse, and greater pain during walking were associated with increased risk of cartilage worsening over 2-years. Results were similar in the sensitivity analysis (Supplement).

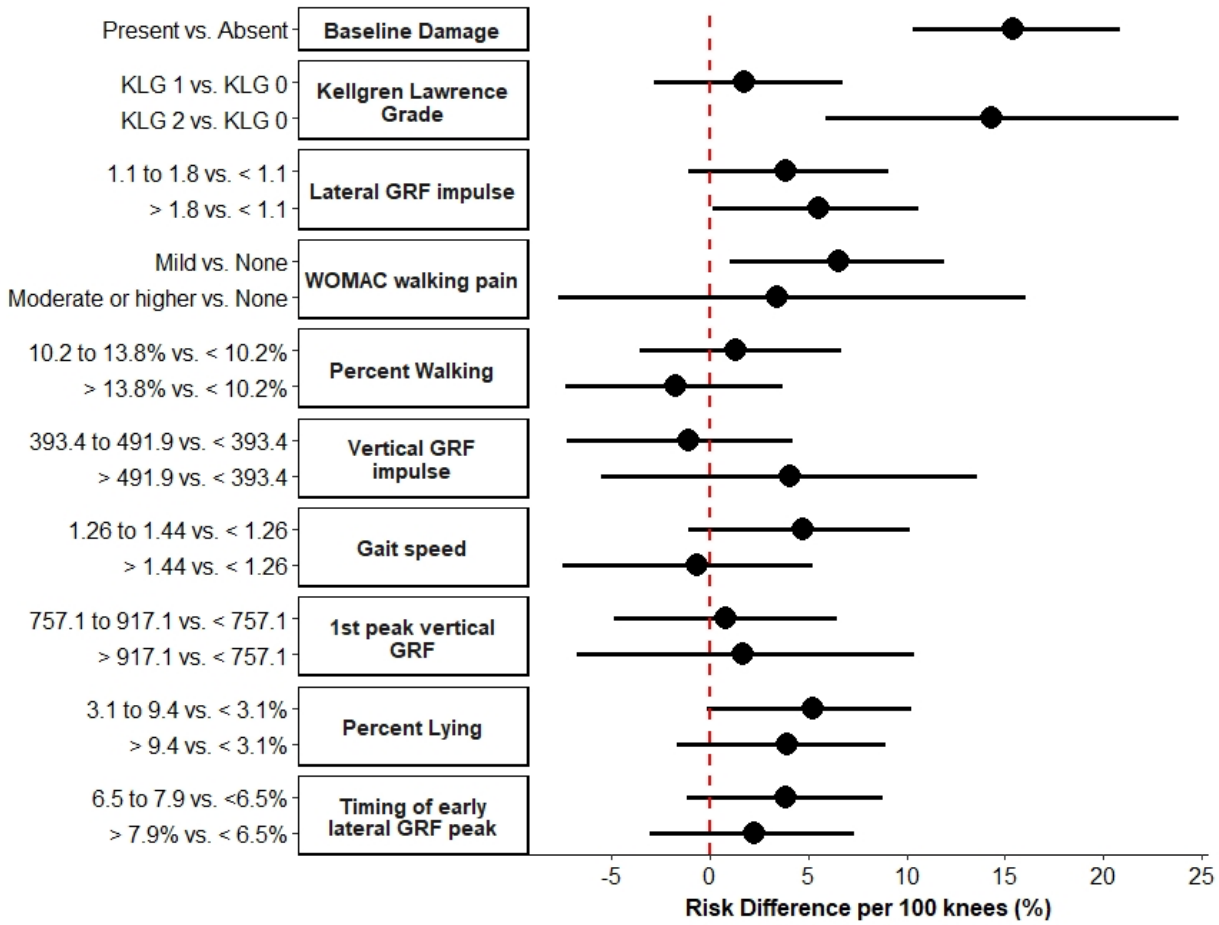

Figure 4. Causal risk differences for influential predictors identified from the machine learning model

DISCUSSION

An ensemble machine learning approach incorporating gait, physical activity, and clinical/demographic features showed good performance (median AUC = 0.73) predicting medial tibiofemoral cartilage worsening over 2-years in knees without or with early radiographic OA. While determining the relationships among predictors and outcomes in machine learning models is challenging, our analysis suggests that addressing high lateral GRF impulse may be a potential target to reduce cartilage worsening.

Model performance

The current 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 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.[32] 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.[33] 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 studies of gait typically examined knee-specific loading (e.g., knee adduction moment) rather than GRFs, often with samples of 15 to 300 knees.[5] Correspondingly, few addressed clinical/demographic confounders, incorporated physical activity, or used cross-validation to examine performance in independent test sets. Further, many were conducted in samples with established OA (KLG ≥ 2), limiting their potential to identify those at risk of progression early in the disease process or identify targets for early intervention. Our sample included 947 individuals with KLG ≤ 2, who predominantly had no or had mild pain during walking, and thus were younger with lower BMI than previously reported samples.[18]

Predictors of OA progression

The machine learning model identified multiple influential gait and physical activity predictors of cartilage worsening in knees without or with early radiographic disease. The only significant result from the g-computation analyses, however, was for lateral GRF impulse, where there was 5.5% increased risk of cartilage worsening for every 100 individuals in the highest compared to lowest tertile. 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 to limbs without radiographic OA or pain.[34] The current results suggest lateral GRF may also play a role in progression.

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 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 15.4% increased risk of cartilage worsening for every 100 individuals with baseline damage compared to no damage, and a 14.3% increased risk for every 100 individuals with KLG 2 versus 0. The lack of difference for KLG 1 versus 0 may highlight limitations of the KLG scoring system, which does not reflect tissue-level damage well, particularly in early disease.[35, 36] Along with these structural measures, WOMAC pain during walking appeared as an influential predictor. Knees with mild pain had an increased risk of cartilage worsening (6.6%) compared to those with no pain, but knees with moderate or higher pain did not have a significantly higher risk. The large confidence interval 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 easier and faster to collect than joint moments, collecting GRFs requires specialized equipment (force platform). Future advances in wearable technologies may facilitate capture of gait information during everyday life, including estimates of GRFs,[37, 38] 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 ≥ 97% of the data splits while the others appeared less consistently (<50%), and only one gait and physical activity predictor had a significant risk difference among tertiles. While we removed highly correlated features, this may result in part from predictors that capture similar constructs (e.g., four features collectively describing an important construct could each appear as top contributors 25% of the time). Similarly, our g-computation approach provides insight into causal pathways but does not account for potential concurrent changes in several risk factors. An important motivation for using machine learning was to address potential interactions among gait, physical activity, and clinical/demographic factors. While it is challenging to identify these relationships from the model, the lack of consistency in top contributors and null g-computation results could speak to a need for simultaneous intervention on several gait and physical activity features rather than a single feature, opening interesting avenues for future study.

Strengths and limitations

Strengths include the large sample, investigation of risk factors in early disease, incorporation of both gait and physical activity, use of machine learning to address potential relationships among predictors, and use of g-computation to quantify the effects of these predictors. These strengths allowed us to expand existing literature by accounting for patient demographics and clinical characteristics, examining multiple gait and physical activity features, and testing our model in independent samples. While this study was performed in the large MOST dataset, our sample was limited to persons with KLG ≤ 2 and was largely White with little to no pain during walking, thus these results may not generalize 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. Knee joint specific loading (e.g., knee adduction moment) is not available in MOST, limiting comparisons to prior longitudinal gait studies. Additionally, we are unaware of

other large datasets with gait, physical activity, and MR outcome data that could be used for external validation. However, we utilized repeated cross-validation to provide information on reproducibility. While 3D GRFs were well characterized (21 features), physical activity was described by only 3 features. Better characterization of dynamic physical activity patterns may improve model performance and ability to identify relevant intervention targets. Last, while cartilage loss in those without baseline damage (i.e., incidence) is of interest, we did not have enough knees with incident damage to evaluate this subsample.

Conclusion

Using an ensemble machine learning approach, we predicted medial tibiofemoral cartilage worsening over 2-years in persons without and with early radiographic osteoarthritis with good performance on independent samples. Additionally, we identified gait and physical activity measures associated with cartilage worsening that may be potential early intervention targets, in particular reducing high lateral ground reaction force impulse.

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.

ETHICS STATEMENT

Patient consent for publication

Not applicable.

Ethics approval

Ethical approval for the Multicenter Osteoarthritis Study was obtained at each of the four core sites (Boston University, University of Alabama at Birmingham, University of California San Francisco, University of Iowa) and all participants provided informed consent prior to participating.

ACKNOWLEDGEMENTS

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

TWITTER

@kecostello, @srjafarz, @ali_guermazi, @FrankRoemer1, @ProfCaraLewis, @vkola_lab, @ProfDeepakKumar

CONTRIBUTORS

Responsibility for the integrity of the work as a whole: K.E. Costello, D. Kumar

Conception and design: K.E. Costello, D. Kumar, D.T. Felson

Analysis and interpretation of the data: All authors

Drafting of the article: K.E. Costello

Critical revision of the article for important intellectual content: All authors

Final approval of the article: All authors

Provision of study materials or patients: D.T. Felson, C.E. Lewis, M.D. Nevitt, N.A. Segal

FUNDING

MOST is comprised of four cooperative grants [D.T. Felson (BU) – AG18820, J.C. Torner (UI) – AG18832, C.E. Lewis (UAB) – AG18947, and M.C. Nevitt (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 (K.E. Costello), P30AR072571 (D.T. Felson), 1UL1TR001430 (D.T. Felson), R21AR074578 (S.R. Jafarzadeh), R03AG060272 (S.R. Jafarzadeh), K23AR063235 (C.L. Lewis), R01HL159620 (V.B. Kolachalama), R21CA253498 (V.B. Kolachalama), and K01AR069720 (D. Kumar) from the National Institutes of Health and 17SDG33670323 (V.B. Kolachalama) and 20SFRN35460031 (V.B. Kolachalama) from the American Heart Association. This manuscript was prepared using MOST data and does not necessarily represent the official views of MOST investigators or the National Institutes of Health. The National Institutes of Health was not involved in study design, collection, analysis or interpretation of data, or the decision to submit this manuscript for publication.

COMPETING INTERESTS

NAS reports personal fees from Tenex Health and grants from Pacira Bioscience, Inc., 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.

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. *The Lancet*. 2020;396(10258):1204-22. doi: 10.1016/s0140-6736(20)30925-9

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 and rheumatism*. 2002;46(10):2632-6. doi: 10.1002/art.10530

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 and Cartilage*. 2011;19(6):683-8. doi: 10.1016/j.joca.2011.02.008

4. Griffin TM, Guilak F. The role of mechanical loading in the onset and progression of osteoarthritis. *Exerc Sport Sci Rev*. 2005;33(4):195-200. doi:

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 and cartilage / OARS, Osteoarthritis Research Society*. 2021. doi: 10.1016/j.joca.2021.10.010

6. Fan J, Han F, Liu H. Challenges of Big Data Analysis. *Natl Sci Rev*. 2014;1(2):293-314. doi: 10.1093/nsr/nwt032

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. doi: 10.1002/acr2.11319

8. Jamshidi A, Pelletier JP, Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. *Nature reviews Rheumatology*. 2019;15(1):49-60. doi: 10.1038/s41584-018-0130-5

9. Segal NA, Nevitt MC, Gross KD, et al. The Multicenter Osteoarthritis Study: opportunities for rehabilitation research. *PM R*. 2013;5(8):647-54. doi: 10.1016/j.pmrj.2013.04.014

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 and cartilage / OARS, Osteoarthritis Research Society*. 2012;20(11):1391-8. doi: 10.1016/j.joca.2012.07.012

11. McKean KA, Landry SC, Hubley-Kozey CL, et al. Gender differences exist in osteoarthritic gait. *Clinical Biomechanics*. 2007;22(4):400-9. doi: 10.1016/j.clinbiomech.2006.11.006

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. doi: 10.1016/j.gaitpost.2017.06.015

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 and cartilage / OARS, Osteoarthritis Research Society*. 2012;20(11):1234-42. doi: 10.1016/j.joca.2012.08.004

14. Sims EL, Carland JM, Keefe FJ, et al. Sex differences in biomechanics associated with knee osteoarthritis. *Journal of women & aging*. 2009;21(3):159-70. doi: 10.1080/08952840903054856

15. Robbins SM, Pelletier JP, Abram F, et al. Gait risk factors for disease progression differ between non-traumatic and post-traumatic knee osteoarthritis. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*. 2021;29(11):1487-97. doi: 10.1016/j.joca.2021.07.014

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. *Seminars in Arthritis and Rheumatism*. 2018;47(5):683-8. doi: 10.1016/j.semarthrit.2017.10.005

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(2):195-202. doi: 10.1002/acr.21803

18. Dunlop DD, Song J, Semanik PA, et al. Objective physical activity measurement in the osteoarthritis initiative: Are guidelines being met? *Arthritis & Rheumatism*. 2011;63(11):3372-82. doi: 10.1002/art.30562

19. 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(12):1833-40. doi:

20. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis*. 1957;16(4):494-502. doi: 10.1136/ard.16.4.494

21. 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(10):1271-5. doi: 10.1136/ard.2006.062927

22. Kronmal RA. Spurious correlation and the fallacy of the ratio standard revisited. *Journal of the Royal Statistical Society Series A (Statistics in Society)*. 1993;156(3):379-92. doi:

23. 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(12):1724-32. doi: 10.1002/acr.20305

24. Aadland E, Ylvisaker E. Reliability of Objectively Measured Sedentary Time and Physical Activity in Adults. *PLoS One*. 2015;10(7). doi: 10.1371/journal.pone.0133296

25. 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(8):1376-81. doi:

26. Hunter DJ, Guermazi A, Lo GH, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*. 2011;19(8):990-1002. doi: 10.1016/j.joca.2011.05.004

27. 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 & rheumatology*. 2020;72(6):957-65. doi: 10.1002/art.41181

28. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45(3). doi: 10.18637/jss.v045.i03

29. 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. arXiv [Preprint]. February 18, 2022 [cited 2022 Jun 20] https://doi.org/10.48550/arXiv.2202.09101.

30. Coyle JR, Hejazi NS, Malenica I, et al. sl3: modern pipelines for machine learning and super learning. https://github.com/tlverse/sl32021. p. R package version 1.4.2.

31. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *American journal of epidemiology*. 2011;173(7):731-8. doi: 10.1093/aje/kwq472

32. Du Y, Almajalid R, Shan J, et al. A novel method to predict knee osteoarthritis progression on MRI using machine learning methods. *IEEE Transactions on Nanobioscience*. 2018;17(3):228--36. doi: 10.1109/TNB.2018.2840082

33. 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(1):20038. doi: 10.1038/s41598-019-56527-3

34. 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). *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*. 2021. doi: 10.1016/j.joca.2021.03.009

35. Luyten FP, Denti M, Filardo G, et al. Definition and classification of early osteoarthritis of the knee. *Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA*. 2012;20(3):401-6. doi: 10.1007/s00167-011-1743-2

36. 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 & Rheumatism*. 1991;34(11):1381-6. doi:

37. Jacobs DA, Ferris DP. Estimation of ground reaction forces and ankle moment with multiple, low-cost sensors. *Journal of neuroengineering and rehabilitation*. 2015;12(1):90. doi: 10.1186/s12984-015-0081-x

38. 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. doi: 10.1016/j.jbiomech.2019.109416

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	5-6
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	6
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-11
Bias	9	Describe any efforts to address potential sources of bias	7-12
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-12
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	11-12
		(*b*) Describe any methods used to examine subgroups and interactions	11-12
		(*c*) Explain how missing data were addressed	11
		(*d*) If applicable, explain how loss to follow-up was addressed	
		(*e*) Describe any sensitivity analyses	12
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	12
		(b) Give reasons for non-participation at each stage	6-9
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Tables 1-2
		(b) Indicate number of participants with missing data for each variable of interest	Tables 1-2
		(c) Summarise follow-up time (eg, average and total amount)	6
Outcome data	15*	Report numbers of outcome events or summary measures over time	12

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	12-13
		(*b*) Report category boundaries when continuous variables were categorized	12-13
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12-13
Discussion			
Key results	18	Summarise key results with reference to study objectives	13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	16-17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13-17
Generalisability	21	Discuss the generalisability (external validity) of the study results	16
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	18-19

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.