



OPEN

Relation of gait measures with mild unilateral knee pain during walking using machine learning

Kathryn L. Bacon¹✉, David T. Felson¹, S. Reza Jafarzadeh¹, Vijaya B. Kolachalama¹, Jeffrey M. Hausdorff², Eran Gazit³, Neil A. Segal⁴, Cora E. Lewis⁵, Michael C. Nevitt⁶, Deepak Kumar¹ & Multicenter Osteoarthritis Study Investigators*

Gait alterations in those with mild unilateral knee pain during walking may provide clues to modifiable alterations that affect progression of knee pain and osteoarthritis (OA). To examine this, we applied machine learning (ML) approaches to gait data from wearable sensors in a large observational knee OA cohort, the Multicenter Osteoarthritis (MOST) study. Participants completed a 20-m walk test wearing sensors on their trunk and ankles. Parameters describing spatiotemporal features of gait and symmetry, variability and complexity were extracted. We used an ensemble ML technique (“super learning”) to identify gait variables in our cross-sectional data associated with the presence/absence of unilateral knee pain. We then used logistic regression to determine the association of selected gait variables with odds of mild knee pain. Of 2066 participants (mean age 63.6 [SD: 10.4] years, 56% female), 21.3% had mild unilateral pain while walking. Gait parameters selected in the ML process as influential included step regularity, sample entropy, gait speed, and amplitude dominant frequency, among others. In adjusted cross-sectional analyses, lower levels of step regularity (i.e., greater gait variability) and lower sample entropy (i.e., lower gait complexity) were associated with increased likelihood of unilateral mild pain while walking [aOR 0.80 (0.64–1.00) and aOR 0.79 (0.66–0.95), respectively].

Individuals with knee osteoarthritis (OA) are known to exhibit multiple gait alterations. These may include alterations in spatio-temporal parameters (e.g., greater stride duration, lower cadence, lower gait speed, lower stride length) and in kinematics and kinetics (e.g., varus thrust, greater knee adduction moment)^{1,2}. These alterations in gait, specifically greater knee adduction moment and varus thrust^{3,4}, in people with knee OA can lead to accelerated disease progression^{5–9}, and hence, interventions to modify gait to slow progression of knee OA have been of significant clinical interest^{10,11}.

While most prior studies have focused on gait alterations related to structural OA pathology^{12,13}, less is known about their relation to knee pain¹⁴, particularly mild knee pain. It is known that cartilage loss in OA progression is not a direct source of pain in mild to moderate OA¹⁵. An indicator of early OA may be unilateral knee pain, hypothesized to alter loading patterns across both knees and contribute to the eventual development of OA in the contralateral knee. Most persons who start with unilateral knee pain from OA eventually develop bilateral OA. Thus understanding the associations between mild unilateral knee pain and gait may provide important mechanistic insights about gait in knee pain and OA, insights that could be used to identify interventions to modify gait to reduce pain or to prevent progression to more severe pain.

Prior gait studies in people with knee OA have additional limitations. Most have relied on small sample sizes, limiting their ability to comprehensively characterize gait and account for confounders. Also, prior studies selected a limited number of gait variables which could lead to important information being missed. Finally, prior studies used 3D optical motion capture to characterize gait. Although optical motion capture provides high accuracy, it is expensive and time consuming to collect and process these data, limiting its application to research laboratories⁶.

¹Boston University Chobanian & Avedisian School of Medicine, 650 Albany Street, Suite X200, Boston, MA 02118, USA. ²Tel Aviv University, Tel Aviv, Israel. ³Tel Aviv Sourasky Medical Center, Tel Aviv, Israel. ⁴University of Kansas Medical Center, Kansas City, USA. ⁵University of Alabama at Birmingham, Birmingham, AL, USA. ⁶University of California, San Francisco, San Francisco, CA, USA. *A list of authors and their affiliations appears at the end of the paper. ✉email: kbacon@bu.edu

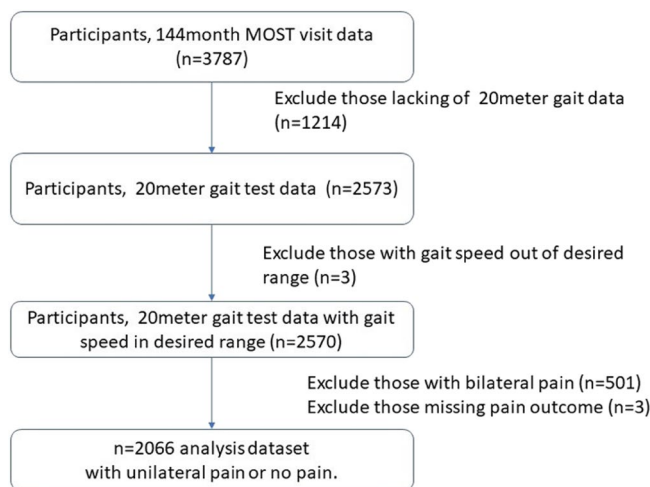


Figure 1. Selection for analysis dataset. Selection of participants for analysis dataset from MOST cohort.

Characteristic	n = 2066
Age, years mean (sd)	63.3 (10.4)
Female, %	55.7
BMI, mean (sd)	29 (5.5)
Whites, %	84.5
CES-D Score (calculated from full twenty questions) mean (sd)	5.8 (6.5)
Radiographic knee OA, % categ: None/unilateral OA/bilateral OA	72.9/ 13.7/ 13.5 (n = 1446/ 271/ 268)
Unilateral knee pain walking, % categ: none/mild/moderate or severe	73.9/ 21.3/ 4.8 (n = 1527/ 440/ 99)
Gait speed, m/s mean (sd)	1.3 (0.2)

Table 1. Participant Characteristics at 144-month MOST visit (baseline).

Advances in wearable movement sensors allows rapid assessment of gait in large cohorts. Few studies of OA or of knee pain have used wearable inertial sensors in “native” knees (i.e. no TKR)^{16,17}. Furthermore, work in Parkinson’s disease^{18–20} has demonstrated the potential of state-of-the-art machine learning (ML) analytical techniques, particularly when combined with wearable inertial sensors for the collection of gait data, either in or outside gait labs. Machine learning techniques allow the use of the computer to “learn” connections within the data using few assumptions. These approaches, however, have not yet been systematically applied to gait in knee pain or OA^{21,22}.

Our objective was to determine gait alterations associated with mild unilateral knee pain using gait data collected with inertial sensors in a large cohort of participants with or at risk of knee OA from the Multicenter Osteoarthritis (MOST) study. We first used ML approaches to select gait parameters related to mild knee pain. We then determined the associations of these gait parameters with unilateral knee pain while adjusting for common covariates for pain and OA.

Results

Our study sample (Fig. 1, Table 1) included 2066 participants (mean age 63.3 [SD: 10.4] years, 56% female) from the 144-month visit of the MOST study. In this cohort, 21.3% (n = 440) of participants had unilateral mild knee pain while walking; 4.8% (n = 99) had moderate/severe unilateral knee pain while walking. Of those with unilateral walking pain, 15.0% (n = 81) had unilateral radiographic OA; 20.4% (n = 110) had bilateral radiographic OA. Thus although the focus of this analysis is on unilateral knee pain while walking, there are participants in this sample with bilateral radiographic OA.

Machine learning for variable selection. The median area under the receiver operating curve (AUC) across 100 runs was 0.75 (2.5–97.5th percentiles = 0.72–0.78). The top contributing gait parameters for unilateral knee pain based on the variables importance (VIM) statistic are shown in Table 2. Non-gait-related variables chosen by the ML process as “influential” include age, BMI, Center for Epidemiological Studies Depression (CES-D) Scale, and radiographic OA. These variables, plus sex and race/site, had already been chosen as confounders to be included in the adjusted models.

Predictors selected from machine learning	Importance (% runs, variables found to be important)
BMI	100
CES-D	100
Radiographic OA	100
Gait Speed (m/s)	100
Age	99
Sample Entropy	99
Step Regularity	83
Amplitude Dominant Frequency (psd)	68
Step Symmetry	40
Left Swing Time (s)	35

Table 2. Variables selected from machine learning.

Predictors selected from machine learning	Unadjusted Mild pain versus no pain OR (95% CI) ^{ad}	Adjusted Mild pain versus no pain aOR ^d (95% CI) ^{bd}
Gait Speed (m/s) ^c	1.03 (0.84–1.27)	1.16 (0.92–1.46)
Sample Entropy ^c	0.83 (0.70–0.99)	0.79 (0.66–0.95)
Step Regularity ^c	0.78 (0.64–0.95)	0.80 (0.64–1.00)
Amplitude Dominant Frequency (psd) ^{cd}	0.94 (0.79–1.11)	0.95 (0.79–1.14)
Step Symmetry ^c	1.02 (0.89–1.17)	0.98 (0.85–1.14)
Left Swing Time ^c (s)	0.90 (0.80–1.02)	0.98 (0.85–1.12)

Table 3. Associations with unilateral mild pain while walking. ^aModel 1: Unadjusted model including only gait variables, listed above. ^bModel 2: Adjusted for age, race, site, sex, BMI, CES-D, and radiographic OA status of ipsilateral and contralateral knees. ^cAll continuous variables are standardized. ^daOR adjusted odds ratio, CI confidence interval, psd power spectral density.

Predictors selected from machine learning	Mild pain versus no pain aOR ^c (95% CI) ^{ac}	Moderate/severe pain versus no pain aOR ^c (95% CI) ^{ac}
Gait Speed (m/s) ^b	1.16 (0.92–1.46)	0.63 (0.39–1.01)
Sample Entropy ^b	0.79 (0.66–0.95)	0.92 (0.65–1.30)
Step Regularity ^b	0.79 (0.64–0.98)	0.91 (0.60–1.39)
Amplitude Dominant Frequency (psd) ^{bc}	0.95 (0.79–1.14)	1.31 (0.91–1.87)
Step Symmetry ^b	0.99 (0.85–1.15)	0.84 (0.62–1.14)
Left Swing Time ^b (s)	0.98 (0.85–1.13)	0.86 (0.66–1.11)

Table 4. Associations with unilateral pain while walking—sensitivity analysis with 3-category pain outcome. ^aAdjusted for age, race, site, sex, BMI, CES-D, and radiographic OA status of ipsilateral and contralateral knees. ^bAll continuous variables are standardized. ^caOR adjusted odds ratio, CI confidence interval, psd power spectral density.

In our testing for multicollinearity among model variables, there were no correlations greater than 0.80, no variables with tolerance values less than 0.10, and no variables with variable inflation factors greater than 10. Inspection of eigenvalues and condition values gave no indication of issues. Based on these collinearity diagnostics, we felt there was no evidence for collinearity sufficient to warrant dropping variables from the model.

Gait alterations related to knee pain during walking. Among those with mild unilateral knee pain during walking ($n = 440$), lower step regularity and lower sample entropy, but not gait speed, were related to greater odds of pain (Table 3) in both unadjusted and adjusted models. The unadjusted model included only gait-related variables as we consider these collectively to be the “exposure”.

The sensitivity analysis with the 3-category pain outcome including those with moderate or severe unilateral pain during walking (Table 4) largely confirmed the findings of our primary analysis. However, among those with moderate or severe unilateral knee pain during walking ($n = 99$), lower gait speed, but not sample entropy or gait regularity, was associated with greater odds of pain. An additional sensitivity analysis excluding those with gait speed greater than 1.8 m/s gave essentially the same results as our primary analyses.

Discussion

In this cross-sectional analysis of a community-dwelling population with or at risk of knee OA, we observed that lower step regularity (i.e., greater gait variability) and lower sample entropy (i.e., lower gait complexity) were related to greater odds of mild unilateral knee pain. Using a large cohort and robust machine learning approaches, our results provide information on gait alterations that may be specifically related to mild knee pain during walking. It is of note that the majority of this sample did not have severe radiographic OA; they had walking pain in what is likely early OA. Given the age of this sample, work by Cibere et al.²³ using MRIs has shown many persons in this age range with chronic knee pain have early OA even if it is not seen in x-rays.

It is challenging to compare the results from our work with prior studies given the paucity of research on gait alterations in those with mild knee pain. The studies available focused on kinematic and kinetic gait parameters^{14,24–27} and report conflicting findings on the relation of measures of knee joint loading (e.g., knee adduction moment) and severity of knee pain.

Lower step regularity reflects greater variability of the center of mass motion. In the absence of a significant difference in stride regularity, lower step regularity also reflects asymmetry of center of mass motion during gait²⁸. While measures of gait variability have been reported to be sensitive to mild to moderate gait impairments in people with neurological impairments (for example, Parkinson's disease and multiple sclerosis)^{29,30}, evidence exists for lower step regularity in people with knee OA^{31,32}. Our findings suggest altered neuromotor control of center of mass motion in the presence of mild unilateral knee pain during walking and may reflect an early adaptation of the nervous system to knee pain. This hypothesis aligns with theories^{33–35} suggesting a link between neurobiological mechanisms underlying chronic pain and control of movement. Interventions to improve gait symmetry in individuals with mild pain may improve step regularity and could be investigated in future studies³⁶.

The idea of altered neuromotor control in the presence of pain is further supported by our finding of lower sample entropy (i.e., lower gait complexity) being related to mild pain. Importantly, the association of sample entropy with knee pain was present after adjusting for age given that with aging, gait becomes less complex³⁷. Hence, the association of knee pain with sample entropy may reflect changes in neuromotor control beyond those due to aging. Both greater variability and lower complexity are considered detrimental gait adaptations and are associated with worse mobility outcomes including greater fall risk^{29,30,38}. Lower gait complexity indicates reduced adaptability of neuromotor control to external perturbations and may reflect increased attention to pain and has been found in persons with OA who have knee buckling episodes^{39,40}. Interventions to increase gait adaptability or to reduce attention to pain (e.g., mindfulness meditation) could increase gait complexity in people with knee OA^{41–43}.

While a few spatio-temporal measures of gait (e.g., swing time) were highlighted in the machine learning model, they were not found to be significant in the logistic regression model. This may be because they were related to covariates added later such as age and BMI or because their relation to mild knee pain was mathematically complex and not captured well by our logistic regression approach.

Given the cross-sectional nature of our study, it is not possible to determine the direction of causality. For example, individuals with knee pain may adapt their gait patterns in response to pain and thereby reduce their pain during walking^{25,44}. Our findings of lower step regularity and complexity may reflect these adaptations. This is supported by findings of reduction in gait variability after administration of an opioid analgesic in people with knee OA⁴⁵. Longitudinal studies are needed to further understand the associations between mild pain and neuromotor control of gait. If these studies show that step regularity and complexity contribute to pain and mobility limitations, they could be targets of therapeutic interventions.

In sensitivity analyses, we confirmed the primary findings of lower step regularity and lower sample entropy being related to greater odds of unilateral mild knee pain during walking. In our analyses of moderate-severe knee pain versus no pain during walking, slower gait speed emerged as important factor. This latter finding aligns with prior studies reporting reduced walking speed in people with advanced structural knee OA¹. Although our results are cross-sectional, slow gait speed in older adults may have long-term consequences such as increased disability, morbidity, and mortality^{46–48}. With greater pain severity, individuals may adopt the simplest strategy to reduce joint loading i.e., reducing gait speed. Another possibility is that brain function of different regions may be altered in those with moderate-severe pain. In older adults, the ability to sustain attention has been reported to be related to gait variability, whereas executive functioning is related to gait speed⁴⁹. Hence, in people with moderate-severe pain, brain regions related to executive functioning may exhibit altered function, whereas in people with mild pain, brain regions related to attention are altered. Longitudinal studies are needed to confirm these cross-sectional observations. An additional sensitivity analysis excluding those with gait speed greater than 1.8 m/s gave essentially the same results as our primary analyses.

The MOST study has several key strengths for this type of analysis: it is a large community-dwelling cohort of men and women with data on risk factors and characteristics of OA, and gait data from inertial sensors. We were able to evaluate people with mild pain. There were also limitations to our study. This is a cross-sectional analysis, so we cannot rule out reverse-causation. Also as in any observational study, residual confounding may exist.

One important strength of our study was the use of wearable sensors which allowed efficient collection of complex data from a large number of persons. Advances in wearable sensors and computing could allow for rapid and easy assessment of these gait outcomes in clinical and real-world settings⁵⁰. These gait alterations, if shown to be important in longitudinal studies of knee pain, could provide clues about interventions to reduce pain in people with knee OA.

In summary, in this cross-sectional study of persons with or at risk of knee OA, measures of step regularity and complexity, derived using wearable inertial sensors, are altered in those with mild unilateral knee pain and may provide new insights into gait abnormalities that occur even with mild pain and that may have implications for strategies to use to prevent mild pain from progressing.

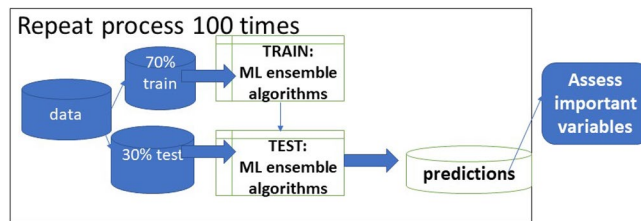


Figure 2. Machine learning process. Representation of the machine learning process repeated 100 times: data is split into 70% training set, 30% test set. Ensemble of ML algorithms is trained on the training set, then testing is done on the testing set and predictions are evaluated. Influential variables are assessed and saved for that run. At the end of all runs influential variables across all runs are assessed.

Methods

Study population. The Multicenter Osteoarthritis (MOST) study is an NIH-funded cohort study of men and women between 45 to 90 years of age at risk of knee osteoarthritis. Study participants were recruited in Birmingham, Alabama and Iowa City, Iowa⁵¹. The study started in 2003 when study participants were interviewed by telephone and attended clinic visits and participants with or at risk of knee OA were recruited. Further details of inclusion and exclusion criteria have been published^{51,52}. In addition to the original cohort, a new cohort of 1500 persons with at most mild radiographic OA and moderate knee pain was recruited at the 144-month clinic visit (2016–2018) of the MOST study. All participants in the 144-month clinic visit (original and new cohort) underwent weight bearing posteroanterior (PA) and lateral fixed flexion radiographs, filled out the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)⁵³ and Center for Epidemiological Studies Depression (CES-D) Scale surveys and had weight and height measured according to the MOST protocol.

Combining the original cohort with the newly recruited one at the 144 month clinic visit, we carried out a cross-sectional analysis comparing individuals with unilateral knee pain while walking to those without knee pain while walking. We included participants who had inertial sensor gait data from a 20-m walk test at the 144-month clinic visit, self-report data on the WOMAC questions on pain while walking, and a score for OA severity based on the Kellgren and Lawrence (KL)⁵⁴ grade.

Gait parameters. An inertial sensor system (OPAL, APDM Inc) was used to collect spatial and temporal measures during over-ground walking. During their 144-month clinic visit, participants completed two trials of a 20-m walk test in an obstacle-free laboratory setting, during which they wore inertial sensors on their trunk and bilateral ankles. Gait parameters available from the MOST database included spatiotemporal features and measures of gait symmetry, variability, and complexity (Table 5). These variables were extracted from the raw vertical acceleration signal from the trunk sensor using published algorithms. For each gait parameter for a participant, the mean of the two trials was used in the analyses⁵⁵. The analysis sample was restricted to those participants with baseline gait speed in the range 0.3 to 2.3 m/second to exclude potentially invalid gait speed measurements (Fig. 1).

Pain while walking. We used the item asking about pain during level walking from the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)⁵³ Pain survey. Participants rated the pain during walking over the past 30 days in their right and left knees (separately) on a Likert scale (none, mild, moderate, severe, extreme). Because we were interested in gait changes that were likely to be asymmetrical, participants who reported bilateral pain while walking were excluded (Fig. 1). We then created a per-person unilateral pain outcome from the left and right knee values for each person (i.e. if an individual has pain, it will be in only one of the legs and these values can map directly into the per-person variable). For the primary per-person dichotomous pain outcome, participants were categorized into those with no pain during walking, versus those with mild unilateral pain during walking (n = 1967); participants with greater than mild pain are marked missing for this outcome. A 3-category pain outcome was then created for a sensitivity analysis including those with more pain, which grouped participants into (1) those with no pain while walking, (2) mild unilateral pain while walking, or (3) moderate or severe unilateral pain while walking (total n = 2066).

Covariates. Age at baseline in years, and BMI (weight kg/ height m²) were used as continuous variables. Depressive symptoms were measured by the CES-D⁵⁶ (range 0–60). A categorical covariate was created to combine study site (Iowa or Alabama sites) and race.

Radiographic OA status. We created a per-person radiographic OA status indicating the number of knees with a Kellgren-Lawrence (KL)⁵⁴ score of 2 or greater: 0, 1 or 2 knees. According to the MOST protocol, once participants developed KL scores of 3 or greater or had a TKR, they no longer were eligible to obtain x-rays on that knee. KL scores from the 144-month MOST visit, if missing, were back-filled with a non-missing value from the most recent prior exam with a non-missing value, if available.

Variable	Explanation
Gait speed (m/s)	Total distance walked/total walking time
Cadence (steps/minute)	Number of steps taken per minute ⁶⁶
Step Length (m)	Total distance walked/number of steps taken
Stride Time (s)	Mean duration of gait cycle (i.e., stride)
Left Stance Time (s)	Mean duration of stance phase for left and right legs
Right Stance Time (s)	
Left Swing Time (s)	Mean duration of swing phase for left and right legs
Right Swing Time (s)	
Left stance percent (%)	Mean duration of stance phase for left and right legs expressed in percentage of the total gait cycle
Right stance percent (%)	
Left swing percent (%)	Mean duration of swing phase for left and right legs expressed in percentage of the total gait cycle
Right swing percent (%)	
Step symmetry	Expresses the symmetry of the acceleration between left and right limbs ^{28,67}
Gait asymmetry	Calculated as $-100 * \ln(\text{Right Swing Time}/\text{Left Swing Time}) $. Values of 0.0 reflect perfect symmetry and higher values reflect greater degrees of asymmetry ⁶⁸
Stride time coefficient of variation (CV)	Reflects the magnitude of the stride-to-stride variability of stride time ²⁹
Left stance time CV	Reflects the magnitude of the stride-to-stride variability of the left or right stance time
Right stance time CV	
Step regularity	Measure of regularity of the vertical acceleration measured from the trunk sensor between consecutive steps using autocorrelation. Low step regularity indicates that there is a low regularity between steps or a systematic asymmetry between left and right leg ^{28,67}
Stride regularity	Expresses the regularity of the vertical acceleration measured from the trunk sensor between consecutive strides using autocorrelation ^{28,67,69}
Sample entropy	Entropy is a measure that quantifies regularity in time series: the more predictable and less complex a series is, the lower the entropy value ^{70,71}
Phase coordination index (PCI)	A measure of bilateral coordination of gait assessed by quantifying the phase relationship between the step timing of the left and right legs ⁷²
Phase CV	Coefficient of variation of the phase ⁷²
Phase absolute difference	Measure of how close the phases are to 180°
Dominant frequency	Frequency with the largest amplitude in the power spectral density in the 0.5–3 Hz band of the vertical acceleration signal from the trunk sensor ⁷³
Amplitude of dominant frequency	The peak of the power spectral density of the vertical acceleration signal from the trunks sensor ⁷³
Width of dominant frequency	Width of the peak at half of its peak amplitude ⁷³
Lyapunov exponent (short divergent)	Estimate of local stability during gait ⁷⁴

Table 5. Gait variables available in the MOST database.

Machine learning process for variable selection. Our goal was to identify gait characteristics which may be associated with mild unilateral knee pain. As a first step towards this goal, we used an ensemble ML technique (“super learning”)^{57,58} as a feature reduction approach to identify important variables associated with the presence/absence of unilateral knee pain during walking. Super learner uses a multi-fold cross-validation to select the optimal combination of algorithms^{58,59} that theoretically achieves an accuracy superior to any single ML method. Our super learner configuration with fivefold cross-validation included a stacked ensemble^{58,60} of the following individual algorithms⁵⁸ appropriate for binomial outcomes: discrete Bayesian additive regression trees, xgboost: extreme gradient boosting, generalized linear models (GLM) with convex penalties (that consisted of least absolute shrinkage and selection operator [LASSO], GLM ridge regression, and GLM elastic net, logistic regression, random forest(ranger: Fast(er) Random Forests)⁶¹, and support vector machine^{62,63}. Variables included for possible selection by the super learning process included all inertial sensor gait variables, age, BMI, CES-D, sex, and radiographic OA status. Missing values were first imputed in the data used for the ML process, using multivariate imputation by chained equations (MICE)⁶⁴ and data were then randomly split into 70% training (development) and 30% test (evaluation) sets. To increase robustness, the random data split and model training and testing were repeated 100 times as part of the ML process (see Fig. 2). A variable importance measure (VIM) statistic based on loss squared error⁵⁸ identified variables which contributed to the prediction of unilateral knee pain in each run; we then took the 10 variables most frequently identified across 100 runs to use in logistic regression models. The area under the curve (AUC) was calculated for each run, and the median AUC across all runs was calculated.

Logistic regression for direction and magnitude of association between gait variables and knee pain. The second and final step in our process to determining the direction and magnitude of association between gait variables and mild unilateral knee pain during walking was a logistic regression model built using the variables chosen by the ML process. Continuous variables were standardized. As our “exposure” is

the set of gait variables, the unadjusted model shown (Table 3) includes the gait variables chosen with the ML process. Confounders found to be associated with knee pain in other studies, including age, sex, race, BMI, depressive symptoms, and person-level radiographic OA status were included in adjusted models, even if they were not selected by the ML process.

We evaluated the model variables for collinearity by reviewing correlations, tolerance values, and variable inflation factors as follows⁶⁵. We first checked correlations on all variables, looking for any which might be greater than 0.8. In the case of a high correlation between model variables we would choose to drop (one or more) variables. We then examined tolerance values, where a value over 0.10 could indicate multicollinearity issues, and evaluated variable inflation factors to ensure that all were less than 10 for our model variables. Last we examined eigenvalues and condition values in the collinearity diagnostics from SAS. For any indication of multicollinearity we would choose to drop one or more variables and test the remaining variables again (Table 2).

For sensitivity analyses using the 3-category walking pain outcome, we first tested for proportional odds. As this assumption did not hold, we used multinomial logistic regression with a generalized logit model to evaluate the 3-category walking pain outcome with the same variables used in the dichotomous model. We also tested our models in an analysis sample further restricted on gait speed (gait speed < 1.8 m/s).

Analyses were performed using SAS software version 9.4 and R version 4.0.2.

Ethics approval. All participants underwent an informed consent process approved by the Institutional Review Board Committee on Human Research at the participating institutions as listed in the Acknowledgements statement. Written consent is obtained from all participants at the beginning of a clinic visit at the clinical center. The consent covers all data collection scheduled for each grant cycle. Verbal consent is obtained for telephone interviews. Participants give written permission for clinical centers to obtain medical records needed for documentation of joint replacement surgery. There were no participants under age 16.

Data availability

All data used in this project is publicly available through the Multicenter Osteoarthritis Study (MOST), now at the NIA AgingResearchBiobank. Multicenter Osteoarthritis Study (nih.gov) is at: <https://agingresearchbiobank.nia.nih.gov/studies/most/>.

Received: 7 July 2022; Accepted: 22 September 2022

Published online: 23 December 2022

References

- Mills, K., Hunt, M. A. & Ferber, R. Biomechanical deviations during level walking associated with knee osteoarthritis: A systematic review and meta-analysis. *Arthritis Care Res.* **65**, 1643–1665. <https://doi.org/10.1002/acr.22015> (2013).
- Iijima, H., Shimoura, K., Ono, T., Aoyama, T. & Takahashi, M. Proximal gait adaptations in individuals with knee osteoarthritis: A systematic review and meta-analysis. *J. Biomech.* **87**, 127–141. <https://doi.org/10.1016/j.jbiomech.2019.02.027> (2019).
- Sharma, L. *et al.* Varus thrust and incident and progressive knee osteoarthritis. *Arthritis Rheumatol* **69**, 2136–2143. <https://doi.org/10.1002/art.40224> (2017).
- Wink, A. E. *et al.* Association of Varus Knee Thrust during walking with worsening WOMAC knee pain: The multicenter osteoarthritis study. *Arthritis Care Res.* **71**, 1353 (2019).
- Bennell, K. L. *et al.* Higher dynamic medial knee load predicts greater cartilage loss over 12 months in medial knee osteoarthritis. *Ann. Rheum. Dis.* **70**, 1770–1774 (2011).
- Chang, A. H. *et al.* External knee adduction and flexion moments during gait and medial tibiofemoral disease progression in knee osteoarthritis. *Osteoarthritis Cartil.* **23**, 1099–1106 (2015).
- Chehab, E. F., Favre, J., Erhart-Hledik, J. C. & Andriacchi, T. P. Baseline knee adduction and flexion moments during walking are both associated with 5 year cartilage changes in patients with medial knee osteoarthritis. *Osteoarthritis Cartil.* **22**, 1833–1839 (2014).
- Hatfield, G. L., Stanish, W. D. & Hubley-Kozey, C. L. Three-dimensional biomechanical gait characteristics at baseline are associated with progression to total knee arthroplasty. *Arthritis Care Res.* **67**, 1004–1014 (2015).
- Miyazaki, T. *et al.* Dynamic load at baseline can predict radiographic disease progression in medial compartment knee osteoarthritis. *Ann. Rheum. Dis.* **61**, 617–622 (2002).
- Bowd, J., Biggs, P., Holt, C. & Whatling, G. Does gait retraining have the potential to reduce medial compartmental loading in individuals with knee osteoarthritis while not adversely affecting the other lower limb joints? A systematic review. *Arch. Rehabil. Res. Clin. Transl.* **1**, 100022. <https://doi.org/10.1016/j.arrct.2019.100022> (2019).
- Richards, R., van den Noort, J. C., Dekker, J. & Harlaar, J. Gait retraining with real-time biofeedback to reduce knee adduction moment: systematic review of effects and methods used. *Arch. Phys. Med. Rehabil.* **98**, 137–150. <https://doi.org/10.1016/j.apmr.2016.07.006> (2017).
- Kwon, S. B. *et al.* Identifying key gait features associated with the radiological grade of knee osteoarthritis. *Osteoarthritis Cartil.* **27**, 1755–1760. <https://doi.org/10.1016/j.joca.2019.07.014> (2019).
- Naili, J. E., Broström, E. W., Clausen, B. & Holsgaard-Larsen, A. Measures of knee and gait function and radiographic severity of knee osteoarthritis—A cross-sectional study. *Gait Posture* **74**, 20–26. <https://doi.org/10.1016/j.gaitpost.2019.08.003> (2019).
- Birmingham, T. B. *et al.* Association between knee load and pain: Within-patient, between-knees, case-control study in patients with knee osteoarthritis. *Arthritis Care Res.* **71**, 647–650. <https://doi.org/10.1002/acr.23704> (2019).
- O'Neill, T. W. & Felson, D. T. Mechanisms of osteoarthritis (OA) pain. *Curr. Osteoporos. Rep.* **16**, 611–616. <https://doi.org/10.1007/s11914-018-0477-1> (2018).
- McCarthy, I., Hodgins, D., Mor, A., Elbaz, A. & Segal, G. Analysis of knee flexion characteristics and how they alter with the onset of knee osteoarthritis: A case control study. *BMC Musculoskelet. Disord.* **14**(1), 1–7 (2013).
- Tadano, S., Takeda, R., Sasaki, K., Fujisawa, T. & Tohyama, H. Gait characterization for osteoarthritis patients using wearable gait sensors (H-Gait systems). *J. Biomech.* **49**, 684–690 (2016).
- Bernad-Elazari, H. *et al.* Objective characterization of daily living transitions in patients with Parkinson's disease using a single body-fixed sensor. *J. Neurol.* **263**, 1544–1551 (2016).
- Gao, C. *et al.* Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. *Sci. Rep.* **8**, 7129 (2018).

20. Hannink, J. *et al.* Sensor-based gait parameter extraction with deep convolutional neural networks. *IEEE J. Biomed. Health Inform.* **21**, 85–93 (2017).
21. Sun, R., Tomkins-Lane, C., Muaremi, A., Kuwabara, A. & Smuck, M. Physical activity thresholds for predicting longitudinal gait decline in adults with knee osteoarthritis. *Osteoarthritis Cartil.* **29**, 965–972 (2021).
22. Kotti, M., Duffell, L. D., Faisal, A. A. & McGregor, A. H. Detecting knee osteoarthritis and its discriminating parameters using random forests. *Med. Eng. Phys.* **43**, 19–29 (2017).
23. Cibere, J. *et al.* Association of clinical findings with pre-radiographic and radiographic knee osteoarthritis in a population-based study. *Arthritis Care Res.* **62**, 1691–1698 (2010).
24. Maly, M. R., Costigan, P. A. & Olney, S. J. Mechanical factors relate to pain in knee osteoarthritis. *Clin. Biomech. (Bristol, Avon)* **23**, 796–805. <https://doi.org/10.1016/j.clinbiomech.2008.01.014> (2008).
25. O'Connell, M., Farrokhi, S. & Fitzgerald, G. K. The role of knee joint moments and knee impairments on self-reported knee pain during gait in patients with knee osteoarthritis. *Clin. Biomech. (Bristol, Avon)* **31**, 40–46. <https://doi.org/10.1016/j.clinbiomech.2015.10.003> (2016).
26. Thorp, L. E., Sumner, D. R., Wimmer, M. A. & Block, J. A. Relationship between pain and medial knee joint loading in mild radiographic knee osteoarthritis. *Arthritis Care Res.* **57**, 1254–1260. <https://doi.org/10.1002/art.22991> (2007).
27. Hall, M. *et al.* The knee adduction moment and knee osteoarthritis symptoms: relationships according to radiographic disease severity. *Osteoarthritis Cartil.* **25**, 34–41. <https://doi.org/10.1016/j.joca.2016.08.014> (2017).
28. Moe-Nilssen, R. & Helbostad, J. L. Estimation of gait cycle characteristics by trunk accelerometry. *J. Biomech.* **37**, 121–126 (2004).
29. Hausdorff, J. M. Gait dynamics in Parkinson's disease: Common and distinct behavior among stride length, gait variability, and fractal-like scaling. *Chaos* **19**, 026113. <https://doi.org/10.1063/1.3147408> (2009).
30. Hausdorff, J. M., Rios, D. A. & Edelberg, H. K. Gait variability and fall risk in community-living older adults: a 1-year prospective study. *Arch. Phys. Med. Rehabil.* **82**, 1050–1056 (2001).
31. Barden, J. M., Clermont, C. A., Kobsar, D. & Beauchet, O. Accelerometer-based step regularity is lower in older adults with bilateral knee osteoarthritis. *Front. Hum. Neurosci.* **10**, 625 (2016).
32. Kiss, R. M. Effect of severity of knee osteoarthritis on the variability of gait parameters. *J. Electromyogr. Kinesiol.* **21**, 695–703. <https://doi.org/10.1016/j.jelekin.2011.07.011> (2011).
33. Butera, K. A., Fox, E. J. & George, S. Z. Toward a transformed understanding: from pain and movement to pain with movement. *Phys. Ther.* **96**, 1503–1507 (2016).
34. Hodges, P. W. & Tucker, K. Moving differently in pain: a new theory to explain the adaptation to pain. *Pain* **152**, S90–S98 (2011).
35. Perini, I. & Bergstrand, S. Where pain meets action in the human brain. *J. Neurosci.* **33**, 15930–15939 (2013).
36. Meder, K. G., Lojaccono, C. T. & Rhea, C. K. A systematic review of non-pharmacological interventions to improve gait asymmetries in neurological populations. *Symmetry* **14**, 281 (2022).
37. Kosse, N. M., Vuillerme, N., Hortobágyi, T. & Lamoth, C. J. C. Multiple gait parameters derived from iPod accelerometry predict age-related gait changes. *Gait Post.* **46**, 112–117. <https://doi.org/10.1016/j.gaitpost.2016.02.022> (2016).
38. Hausdorff, J. M. Gait variability: methods, modeling and meaning. *J. NeuroEng. Rehabil.* **20**, 1–9 (2005).
39. Segal, N. A. *et al.* Relationship of gait complexity to subsequent knee buckling and falls: the most study. *Osteoarthritis Cartil.* **28**, S429–S430 (2020).
40. Homs, A. F., Dupeyron, A. & Torre, K. Relationship between gait complexity and pain attention in chronic low back pain. *Pain* **163**, e31–e39. <https://doi.org/10.1097/j.pain.0000000000002303> (2022).
41. Zeidan, F., Grant, J. A., Brown, C. A., McHaffie, J. G. & Coghill, R. C. Mindfulness meditation-related pain relief: Evidence for unique brain mechanisms in the regulation of pain. *Neurosci. Lett.* **520**, 165–173. <https://doi.org/10.1016/j.neulet.2012.03.082> (2012).
42. Nørgaard, J. E. *et al.* Effects of gait adaptability training on falls and fall-related fractures in older adults: A systematic review and meta-analysis. *Age Ageing* **50**, 1914–1924. <https://doi.org/10.1093/ageing/afab105> (2021).
43. Kumar, D. *et al.* Mindful walking in knee osteoarthritis: a feasibility randomized controlled trial. *Osteoarthritis Cartil.* **30**, S72–S73 (2022).
44. Henriksen, M., Graven-Nielsen, T., Aaboe, J., Andriacchi, T. P. & Bliddal, H. Gait changes in patients with knee osteoarthritis are replicated by experimental knee pain. *Arthritis Care Res.* **62**, 501–509. <https://doi.org/10.1002/acr.20033> (2010).
45. Henriksen, M. *et al.* Opioid-induced reductions in gait variability in healthy volunteers and individuals with knee osteoarthritis. *Pain Med.* **20**, 2106–2114. <https://doi.org/10.1093/pm/pny286> (2019).
46. Guralnik, J. M. *et al.* Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J. Gerontol. A Biol. Sci. Med. Sci.* **55**, M221–M231 (2000).
47. Perera, S. *et al.* Gait speed predicts incident disability: a pooled analysis. *J. Gerontol. Ser. A Biomed. Sci. Med. Sci.* **71**, 63–71 (2016).
48. Studenski, S. *et al.* Gait speed and survival in older adults. *JAMA* **305**, 50–58 (2011).
49. Lo, O.-Y. *et al.* Gait speed and gait variability are associated with different functional brain networks. *Front. Aging Neurosci.* <https://doi.org/10.3389/fnagi.2017.00390> (2017).
50. Soangra, R. & Lockhart, T. E. Inertial sensor-based variables are indicators of frailty and adverse post-operative outcomes in cardiovascular disease patients. *Sensors* **18**, 1792 (2018).
51. Felson, D. T. *et al.* Correlation of the development of knee pain with enlarging bone marrow lesions on magnetic resonance imaging. *Arthritis Rheum.* **56**, 2986–2992 (2007).
52. Segal, N. A. *et al.* The Multicenter Osteoarthritis Study (MOST): Opportunities for Rehabilitation Research. *PM & R J. Injury Funct. Rehabil.* **5** (2013).
53. Bellamy, N. *WOMAC Osteoarthritis Index* (University of Western Ontario, 1995).
54. Kellgren, J. & Lawrence, J. Radiological assessment of osteo-arthrosis. *Ann. Rheum. Dis.* **16**, 494 (1957).
55. Bautmans, I., Jansen, B., Van Keymolten, B. & Mets, T. Reliability and clinical correlates of 3D-accelerometry based gait analysis outcomes according to age and fall-risk. *Gait Posture* **33**, 366–372 (2011).
56. Radloff, L. S. The CES-D scale: A self-report depression scale for research in the general population. *Appl. Psychol. Meas.* **1**, 385–401 (1977).
57. Van der Laan, M. J. & Rose, S. *Targeted Learning: Causal Inference for Observational and Experimental Data* (Springer, 2011).
58. sl3: Pipelines for Machine Learning and Super Learning. <https://doi.org/10.5281/zenodo.1342293> (2021).
59. Wolpert, D. H. Stacked generalization. *Neural Netw.* **5**, 241–259 (1992).
60. Breiman, L. Stacked regressions. *Mach. Learn.* **24**, 49–64 (1996).
61. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
62. Chang, C.-C. & Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011).
63. Meyer, D. *et al.* e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien [R Package Version 1.7–9]. *Comprehensive R Archive Network (CRAN)* (2021).
64. van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67. <https://doi.org/10.18637/jss.v045.i03> (2011).
65. Schreiber-Gregory, D. N. Paper 1404–2017 Multicollinearity: What Is It, Why Should We Care, and How Can It Be Controlled? (SAS, 2017).

66. Slaght, J., Sénéchal, M., Hrubeniuk, T., Mayo, A. & Bouchard, D. Walking cadence to exercise at moderate intensity for adults: a systematic review. *J. Sports Med.* **2017** (2017).
67. Tura, A., Raggi, M., Rocchi, L., Cutti, A. G. & Chiari, L. Gait symmetry and regularity in transfemoral amputees assessed by trunk accelerations. *J. Neuroeng. Rehabil.* **7**, 1–10 (2010).
68. Yogeve, G., Plotnik, M., Peretz, C., Giladi, N. & Hausdorff, J. M. Gait asymmetry in patients with Parkinson's disease and elderly fallers: when does the bilateral coordination of gait require attention?. *Exp. Brain Res.* **177**, 336–346 (2007).
69. Kobsar, D., Olson, C., Paranjape, R., Hadjistavropoulos, T. & Barden, J. M. Evaluation of age-related differences in the stride-to-stride fluctuations, regularity and symmetry of gait using a waist-mounted tri-axial accelerometer. *Gait Posture* **39**, 553–557 (2014).
70. Goldberger, A. L., Peng, C.-K. & Lipsitz, L. A. What is physiologic complexity and how does it change with aging and disease?. *Neurobiol. Aging* **23**, 23–26 (2002).
71. Torres, B. D. L. C., López, M. S., Cachadaña, E. S. & Orellana, J. N. Entropy in the analysis of gait complexity: A state of the art. *Br. J. Appl. Sci. Technol.* **3**, 1097 (2013).
72. Plotnik, M., Giladi, N. & Hausdorff, J. M. A new measure for quantifying the bilateral coordination of human gait: effects of aging and Parkinson's disease. *Exp. Brain Res.* **181**, 561–570 (2007).
73. Weiss, A. *et al.* Toward automated, at-home assessment of mobility among patients with Parkinson disease, using a body-worn accelerometer. *Neurorehabil. Neural Repair* **25**, 810–818 (2011).
74. Bruijn, S. M., Bregman, D. J., Meijer, O. G., Beek, P. J. & van Dieën, J. H. Maximum Lyapunov exponents as predictors of global gait stability: A modelling approach. *Med. Eng. Phys.* **34**, 428–436 (2012).

Acknowledgements

MOST is comprised of four cooperative grants (D Felson—AG18820; J Torner—AG18832, C Lewis—AG18947, and M Nevitt—AG19069) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by MOST study investigators. Financial Support for this work was also provided by grants (NIAMS—P30 AR0702571, K01 AR06972), and (Bacon Rheumatology Research Foundation Investigator Award). University of California, San Francisco; University of Iowa; University of Alabama at Birmingham; Boston University; National Institute on Aging (NIA). MOST has been conducted in accordance with U.S. Department of Health and Human Services Protection of Human Subjects regulations (45 CFR part 46) and the Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA) of 1996. Research data and image sets are de-identified in accordance with regulation 45 CFR 164.514(e). Study participants were recruited under local institutional review board (IRB) approval through the collaborating institutions: grant AG18820, Boston University IRB FWA00000301; grant AG18832 University of Iowa IRB FWA00003007; grant AG18947 University of Alabama at Birmingham IRB FWA00005960; grant AG19069 University of California at San Francisco IRB FWA00000068.

Author contributions

Study coauthors contributed substantially to the conception or design of the work (K.L.B., D.T.F., D.K.), data acquisition (D.T.F., N.A.S., C.E.L., M.C.N., J.M.H., E.G.), data analysis (K.L.B., D.T.F., S.R.J., V.B.K., D.K.), interpretation of results (K.L.B., D.T.F., D.K.); drafting the work (K.L.B., D.T.F., D.K.), revising the work critically and providing final approval of the manuscript (K.L.B., D.T.F., S.R.J., V.B.K., J.M.H., E.G., N.A.S., C.E.L., M.C.N., D.K.). K.L.B. takes responsibility for the integrity of the work as a whole.

Funding

Funding sources are not associated with the scientific contents of the study.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.L.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Multicenter Osteoarthritis Study Investigators

David T. Felson¹, Neil A. Segal⁴, Cora E. Lewis⁵ & Michael C. Nevitt⁶