Rethinking Measurement of Movement-Evoked Pain with Digital Technology

Madelyn R. Frumkin[1,2], Jingwen Zhang[3], Ziqi Xu[3], Salim Yakdan[4], Braeden Benedict[4], Saad Javeed[4], Justin Zhang[4,5], Kathleen Botterbush[4], Burel R. Goodin[6], Chenyang Lu[3], Wilson Z. Ray[4], & Jacob K. Greenberg[4]

[1]Center for Technology and Behavioral Health, Dartmouth College, Lebanon, NH, USA
[2]Department of Biomedical Data Science, Dartmouth College, Lebanon, NH, USA
[3]Department of Computer Science and Engineering, Washington University, St. Louis, MO, USA
[4]Department of Neurological Surgery, Washington University, St. Louis, MO, USA
[5]Department of Neurological Surgery, University of Utah, Salt Lake City, USA
[6]Department of Anesthesiology, Washington University, St. Louis, MO, USA

Corresponding Author:
Madelyn R. Frumkin, PhD
Center for Technology and Behavioral Health
Department of Biomedical Data Science
Geisel School of Medicine
Dartmouth College

Address: 46 Centerra Parkway, Lebanon, NH 03766
Email: Madelyn.r.frumkin@dartmouth.edu

Abstract

Movement-evoked pain (MEP) may be a useful metric for phenotyping musculoskeletal pain conditions. However, there is significant disagreement over operationalization, and no studies to our knowledge have assessed stability of MEP over time. We collected Fitbit and Ecological Momentary Assessment (EMA) data from adults with moderate-to-severe chronic pain schedule to receive lumbar/thoracolumbar fusion surgery (N=114). On average, participants provided 323 hours of Fitbit data and 74 EMA surveys (84% completion rate). To mimic task-based assessment of MEP using the 6-minute walk test, we extracted EMA pain ratings completed within 3 hours of walking at a speed ≥67spm for at least 6 minutes. Of the full sample, 93 individuals (82%) had any instances of pain ratings following 6-minute activity bouts (*Median*=7, *SD*=11). Post-activity pain scores exhibited good within-person consistency (ICC = .75). However, between-person differences in average pain accounted for >70% of the variance in post-activity pain. MEP change scores defined as the difference between post-activity and pre-activity pain scores had poor reliability (ICC = .08). MEP change scores were not associated with average pain or factors related to the uncontrolled nature of digital assessment (e.g., activity amount, time from activity to pain report). However, MEP change scores tended to be lower when lag-1 pain was elevated ($\beta = -8.49$, $p < .001$), suggesting ceiling effects. We also observed small effects of time of day and prior activity, which could contaminate MEP assessed in the lab or clinic. We recommend the continued development of digital methodologies for assessing MEP.

1. Introduction

There is growing emphasis on assessing musculoskeletal pain experienced with movement, or movement-evoked pain (MEP), as distinct from pain at rest (PAR)[6–8,16,17,48]. MEP tends to be greater than PAR[16,17], and some studies find MEP is a stronger predictor of prospective outcomes, such as 12-month low back pain-related disability and early recovery after spine surgery[19,27]. However, findings are highly mixed. For example, in knee arthroplasty, some studies suggest MEP and PAR are distinct[30,41], while others describe MEP and PAR as substantially overlapping[38].

Mixed findings may be due to heterogeneity in the methods used to assess MEP. MEP can be assessed retrospectively by asking individuals to rate their pain with activity over a recent time period (e.g., the past 24 hours or the past week). However, retrospective recall of pain is known to be influenced by factors including the individual's most intense and most recent pain experiences[42,50,51]. It is therefore suggested that MEP be assessed in response to tasks likely to evoke pain (e.g., six-minute walk test, repeated chair rise)[8,16]. In some cases, a sequential battery of tasks is used to characterize aggregate MEP (e.g., [9,25,46]).

A key discrepancy in task-based assessment of MEP is whether MEP should be operationalized as absolute post-task pain versus the degree to which pain increases following activity[7]. Numerous studies have used absolute post-activity pain ratings to quantify MEP, such that individuals who report higher post-activity pain (e.g., pain after walking) are considered to have greater MEP. Absolute post-activity pain tends to be highly correlated with PAR and overall disability[25–27,38]. Prior Ecological Momentary Assessment (EMA) research suggests that pain is highly variable within individuals over time, due to factors including negative cognitive and affective states[14,36], sleep[18], and time of day[24,34]. Because task-based assessments of MEP are typically performed on only one occasion, it is unclear if post-activity pain ratings are similarly variable, whether due to pre-activity pain or independent influences of these factors on post-activity pain ratings.

Alternatively, MEP can be isolated from PAR by calculating the difference between pre- and post-activity pain[7]. Such MEP change scores characterize change in pain following one or multiple tasks and are intended to index the degree to which pain increases with activities expected to evoke pain for a given patient population. Similarly to absolute post-activity pain, it is possible that the degree to which pain increases following activity is variable over time due to factors including pre-activity pain level, time of day, and how much activity the individual has engaged in prior to MEP assessment[29,52,53]. For example, difference scores may suffer from ceiling effects if the individual experiences high resting pain prior to engaging in activity intended to evoke pain.

The goal of the current study was to leverage digital technology to assess stability of MEP assessments over time and under naturalistic conditions. To mimic task-based assessment via the six-minute walk test, we extracted pain ratings collected via EMA that were preceded by at least six minutes of moderate-to-vigorous physical activity as assessed via Fitbit. We examined the reliability of both absolute and change score metrics of MEP. We further considered the degree to which MEP metrics were influenced by average pain, pre-activity pain, factors related to the

uncontrolled nature of digital assessment (e.g., activity amount, time from activity to pain report), and factors that could contaminate MEP assessed in the lab or clinic (e.g., time of day, prior activity). Finally, we assessed the risk of ceiling effects by naturalistically examining PAR following periods of no or low physical activity.

2. Materials and methods

2.1. Participants

This is a secondary analysis of data collected to evaluate the feasibility and utility of preoperative mobile health (mHealth) assessment for improving prediction of lumbar spine surgery outcomes. The methods and primary outcomes have been reported elsewhere[19,20]. Briefly, inclusion criteria included English-speaking adults aged 21 to 85 years old who owned a smartphone, had at least 1 week to complete assessments prior to surgery, and reported a numeric rating scale pain score of at least 3 out of 10 during the previous week. Patients who were undergoing surgery for infection, malignancy, or trauma, those undergoing isolated thoracic fusion, and those undergoing another major surgery within 3 months of data collection were excluded. The study was approved by the institutional review board at Washington University School of Medicine (IRB# 202012139), and all patients provided informed consent. To be included in the current study, participants were also required to have both EMA and Fitbit data available preoperatively.

2.2. Procedure

Participants were recruited over the phone following a recent appointment with a neurosurgery or orthopedic spine surgeon. A research coordinator contacted the patient, explained the study, and assessed interest in further participation. If participants indicated interest in the study, then the research coordinator verbally reviewed the purpose and procedures of the study, the voluntary nature of participation, access to protected health information, and study compensation. If patients indicated they would like to participate, they then provided informed consent and were given instructions to download the LifeData application (LifeData LLC) to complete EMA on their personal smartphone. Participants specified a 12-hour period in which they preferred to receive surveys every 3 hours (i.e., 9am to 9pm). Participants also completed self-report questionnaires via REDCap.

After enrolling in the study, participants were mailed a Fitbit Inspire 2 with instructions to wear the tracker as much as possible but at least during the 12-hour EMA period. Participants received 5 EMAs daily for approximately 3 weeks, or until their surgery. Some individuals experienced delays in surgery date due to COVID-19, illness, or other factors. For consistency across participants, we focus our analyses on data collected within 45 days of surgery. Participants could choose the time at which their surveys started. Surveys were administered every 3 hours[1]. Participants had 30 minutes to respond to each survey and were sent 2 reminders at 15-minute

[1] There were 2 individuals for whom errors in EMA set up led to surveys being administered less than 3 hours apart. Surveys received less than 3 hours after the prior survey were removed from the data set so that lag-1 EMA consistently corresponded to surveys completed approximately 3 hours prior.

increments. Participants were paid $1 per completed EMA survey (up to $105), and $20 for using the Fitbit for any duration.

2.3. Measures

2.3.1. Ambulatory movement assessment

Activity data were collected via the Fitbit Inspire 2. Fitbit provides step count and heart rate data extracted every minute. Validity or reliability of Fitbit-based measurements have been assessed in at least 144 studies, and results of a systematic review suggest that Fitbits accurately measure steps[15].

2.3.2. Momentary pain

At each assessment, participants rated pain intensity ("Right now, how intense is your overall pain?") on a scale from 0 (none) to 100 (worst possible). Participants were instructed to respond based on how they were feeling right before they received the notification.

2.3.3. Patient-reported outcome measures

Participants completed computer-adaptive Patient-Reported Outcomes Measurement Information System (PROMIS) pain intensity, pain interference, and physical function measures upon entering the study[1,2,49]. PROMIS scores are reported as t-scores, where 50 corresponds to population average, and 10 is the standard deviation. Higher scores on PROMIS pain intensity and interference and lower scores on physical function suggest impairment[1].

2.4. Data preparation

2.4.1. Data cleaning

Fitbit records step count as zero even if the device is not worn. Therefore, we removed step count observations if heart rate data were not available.

2.4.2. Movement-evoked pain

A schematic overview of ecological MEP assessment is provided in Figure 1A. Pain ratings were marked as indexing MEP if the individual engaged in moderate-to-vigorous physical activity (MVPA) for at least 6 minutes in the 3 hours prior to the pain rating. We chose six minutes to mimic the six-minute walk test (6MWT), which is commonly used to measure MEP in laboratory and clinical settings[8]. The 3-hour window was chosen because pain was measured every 3 hours via EMA. Based on prior literature, we defined MVPA based on a threshold of ≥67 steps per minute (spm)[44]. We considered several step count thresholds to confirm that heart rate was sufficiently elevated to suggest MVPA (see statistical analysis).

MEP is defined in the literature as either absolute post-activity pain ratings or the change in pain from pre- to post-activity. We therefore considered both operationalizations. In the first set of

models, the time-varying outcome is EMA pain ratings following bouts of MVPA (henceforth referred to as "post-activity pain"). In the second set of models, the time-varying outcome is the difference between EMA pain ratings following bouts of MVPA and lag-1 EMA pain ratings made approximately 3 hours earlier by the same individual (henceforth referred to as "MEP change score"). We only calculated MEP change score if the lag-1 EMA pain rating was not associated with a bout of MVPA, so that lag-1 pain more closely mimicked PAR.

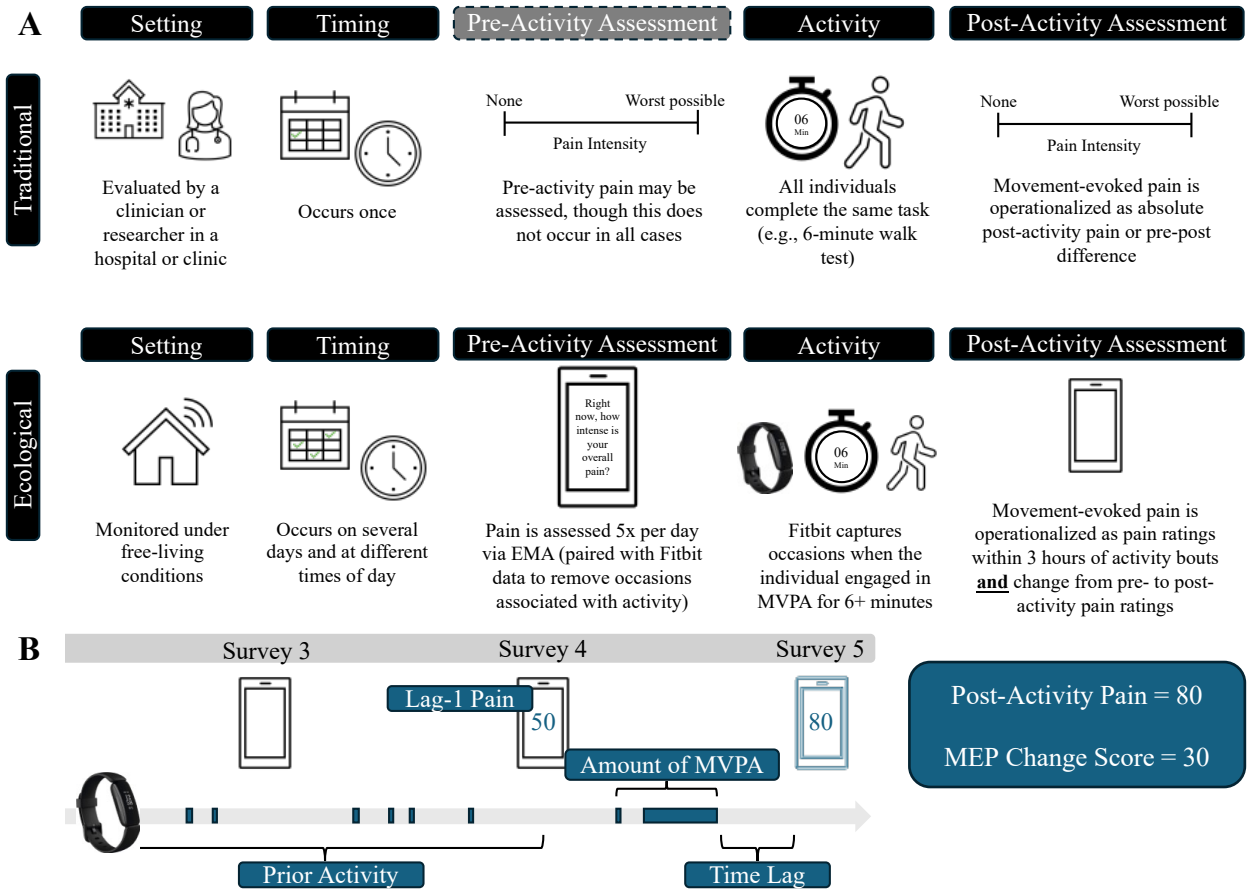

Figure 1. Schematic overview of traditional versus ecological movement-evoked pain assessment (**A**) and time-varying predictors of ecological MEP observations (**B**). EMA = ecological momentary assessment; MVPA = moderate-to-vigorous physical activity.

2.4.3. Predictors of MEP

Raw Fitbit data and EMA pain ratings were used to generate secondary features for predicting post-activity pain. Predictors are described below and in Figure 1B.

Average pain. Average pain for each participant was calculated as the average of all momentary pain responses, regardless of whether Fitbit data were available. Average pain is a time-invariant predictor (one observation per individual) used to predict between-person variability in average MEP. Average pain was group mean centered and scaled, such that the coefficient reflects the expected change in average MEP for an individual whose average pain is 1 SD above the group mean.

Time lag. In this ecological setting, pain ratings were not requested immediately after bouts of MVPA. We therefore calculated time (in hours) from end of MVPA bout to EMA pain rating. The maximum possible time lag was 3 hours. If more than one MVPA bout occurred in the 3-hour window, time lag refers to time since the end of the most recent bout. The coefficient reflects the expected change in MEP per 1-hour increase in time lag.

Survey. Given that pain can fluctuate with circadian rhythms[34,35], we included survey number as a predictor of MEP. Surveys 1-5 were delivered at the same time each day for a given participant. We considered both linear and quadratic ($survey^2$) effects after recoding survey so that 0 referenced the mid-day (e.g., 3pm) survey.

Lag-1 pain. EMA pain ratings made approximately 3 hours before a post-activity pain observation by the same individual were included as a time-varying predictor. We removed sequential MEP observations so that lag-1 pain more closely mimicked PAR. Lag-1 pain was centered and scaled within persons, such that the coefficient reflects the expected change in MEP when lag-1 pain is 1 SD above average for the individual.

Amount of MVPA. Individuals may have engaged in MVPA for differing amounts of time across bouts, and multiple bouts may have occurred within the 3-hour window. We therefore calculated total number of minutes above the MVPA threshold in the 3 hours prior to a post-activity pain rating. Amount of MVPA was centered and scaled within persons, such that the coefficient reflects the expected change in MEP when amount of MVPA within the 3-hour window is 1 SD above average for the individual.

Prior activity. Because MEP may be additive, we calculated amount of daily activity prior to the MEP occasion. To avoid conflating prior activity with amount of MVPA, we calculated cumulative step count up to 3 hours before the post-activity pain rating (e.g., if the post-activity pain rating was recorded at 9pm, we calculated cumulative step count up to 6pm). To account for times when participants may not have worn the Fitbit, cumulative step count was divided by cumulative wear time. Prior activity was centered and scaled within persons, such that the coefficient reflects expected change in MEP when prior activity is 1 SD above average for the individual.

2.4.4. Resting Pain

To assess potential ceiling effects that can occur when pre-activity pain is high, we extracted EMA pain observations that were not associated with at least 6 minutes of MVPA in the 3 hours prior to the pain rating. We retained observations that were preceded by at least 10 minutes of Fitbit wartime in the 30 minutes prior to EMA.

2.5. Statistical Analysis

2.5.1. Defining MVPA

Various thresholds have been used to define sedentary vs. active time using Fitbits and other wearable devices[22,44]. Fitbit uses a proprietary algorithm to categorize 60-second sampling intervals as sedentary, light, moderate, or vigorous activity. However, active time is over-estimated compared to research-grade devices[37,39]. We therefore opted to quantify MVPA based on step count thresholds. Guidelines suggest that moderate intensity walking involves ≥100 steps per minute (spm), including among older adults[55]. However, such a cadence may be uncommon in individuals with physical disability, and walking speeds of 2-2.5 km/h may be sufficient for expanding 3 metabolic equivalents (METs), a common threshold for MVPA[4,21,44,55]. We therefore considered 50, 67, and 100 spm thresholds[44]. We identified all 6-minute windows when average spm was at or above the threshold. This time period was chosen to mimic the 6MWT.

Given a lack of prior research defining MVPA among adults with moderate-to-severe pain interference and disability, we used Fitbit data to verify that heart rate was elevated at each threshold, and as compared to heart rate when sedentary (0spm or <10spm). Minute-level heart rate data was divided by the individual's maximum heart rate (calculated as 208 - 0.7*Age) and averaged within each 6-minute window, yielding average percent of maximum heart rate for the activity bout[54]. We examined the proportion of activity bouts where the individual's heart rate was elevated to at least 50% of their maximum heart rate, a typical cutoff for moderate physical activity[33].

2.5.2. Describing reliability of MEP

We calculated Intraclass Correlation Coefficient (ICC), which indexes the proportion of variance that is due to between-person differences. We also examined the person-specific standard deviation of each MEP metric (absolute post-activity pain and MEP change scores).

2.5.3. Examining multilevel predictors of absolute post-activity pain

We fit mixed-effects models using the lmerTest package in R version 4.4.1[28]. All models include a random intercept, allowing for individual differences in average MEP. There were two time-varying outcomes: post-activity pain and MEP change score. For each outcome, we fit univariate models to understand the impact of potential predictors on both MEP operationalizations. We report marginal R^2 to describe the proportion of variance accounted for by predictors of interest (i.e., fixed effects).

Univariate models use complete case analysis. Most predictors had no missing data, because they were process variables (e.g., time of day), aggregate variables (e.g., person-level average pain), or necessary for inclusion in the data set of MEP observations (e.g., Fitbit data). The exception is lagged pain, which was sometimes missing at the prior assessment (e.g., approximately 3 hours prior) and always missing at the first survey of the day. We opted not to impute missing observations due to problems with parameter estimation when data are not missing at random and a high proportion of missing data are present[47]. Instead, we employed Bayesian multilevel modeling in Mplus to fit multivariate models. Bayesian inference facilitated the analysis of incomplete data and allowed us to incorporate uncertainty into the model estimates via posterior distributions[10].

2.5.4. Describing variability in resting pain

To explore potential ceiling effects in MEP change scores that may occur when pre-activity or resting pain is high, we further examined the distribution of resting pain observations. Because individuals were not instructed to rest before completing EMA, we compared resting pain distributions at differing levels of pre-EMA activity (defined as cumulative steps in the 30 minutes and 3 hours prior to EMA divided by cumulative wear time). We report median PAR at each level of pre-EMA activity, as well as the proportion of observations where resting pain was reported as ≥ 80 on a 0-100 scale indicating severe pain[5].

3. Results

Participants were 114 adults with moderate-to-severe pain interference and disability scheduled to receive lumbar spine surgery (see Table 1). On average, participants provided 323 hours (*SD* = 172) of Fitbit data over 18 days (*SD* = 8). When considering days when participants wore the Fitbit for at least 12 hours during waking hours (e.g., excluding 12am-6am), average daily step count was 6453 (*SD* = 3596, *Min* = 258, *Max* = 15788). Participants completed 84% of EMA prompts on average (*SD* = 15%), yielding an average of 74 completed prompts per person (*SD* = 24).

3.1. Defining MVPA

Maximum steps per minute ranged from 50 to 158, with 11% of participants never exceeding 100 steps/minute (see Figure 2A). As shown in Figure 2B, average heart rate was elevated at the 50spm (*M* = 102.55, *SD* = 9.97), 67spm (*M* = 105.11, *SD* = 10.56), and 100spm (*M* = 111.80, *SD* = 13.34) activity thresholds, compared to windows when participants were unmoving (*M* = 74.48, *SD* = 11.27) or averaged <10spm (*M* = 76.74, *SD* = 11.29). Only 54 participants (47%) had any 6-minute windows with ≥100spm. We therefore retained the 67spm threshold, as 95% of observations corresponded to ≥50% of maximum heart rate, as compared to 92% of observations obtained with the 50spm threshold. Activity bouts characterized by <50% of maximum heart rate were removed (*t* = 55, 4.9%).

Table 1. *Sample descriptives (N = 114)*

Characteristic	**No. (%)**
Sex	
Male	54 (47.4)
Female	60 (52.6)
Ethnicity	
Non-Hispanic	107 (93.86)
Hispanic	7 (6.14)
Race	
White	106 (92.98)
African American	7 (6.14)
Other	1 (0.88)
Education Level	
Graduate or Professional School Degree	36 (31.58)
College Degree	38 (33.33)
High School Degree	36 (31.58)
Did not Graduate High School	4 (3.51)
Work Status	
Actively Working	46 (40.35)
Homemaker	6 (5.26)
On Disability	17 (14.91)
Retired	41 (35.96)
Unemployed	4 (3.51)
Decompression only (no fusion)	7 (6.14)
	M (SD)
Age (years)	58.19 (12.48)
BMI	29.5 (4.8)
PROMIS Pain Intensity	66.39 (6.65)
PROMIS Pain Interference	66.92 (5.12)
PROMIS Physical Function	33.95 (4.93)
Levels fused	2.4 (2.48)

Note. **M**: Mean; **SD**: Standard Deviation; **BMI**: Body Mass Index; **PROMIS**: Patient-Reported Outcomes Measurement Information System

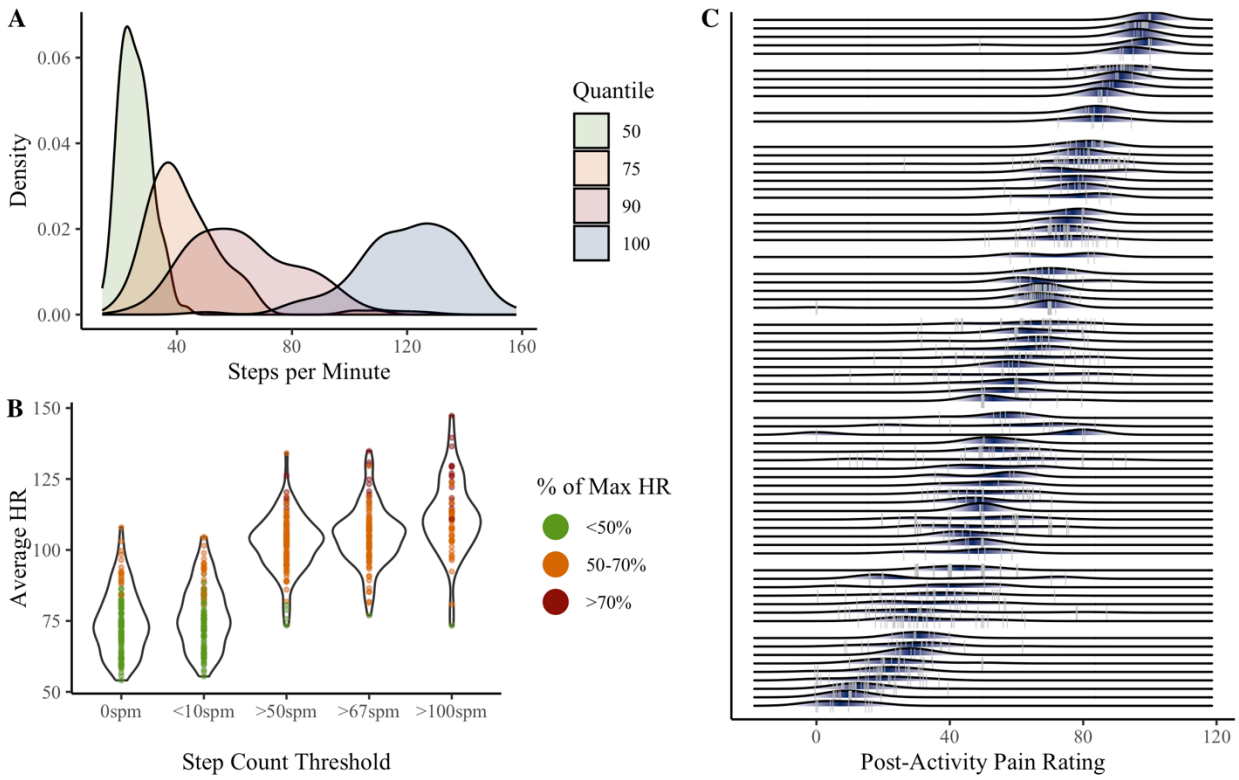

Figure 2. Variability in step quantiles (**A**), heart rate at step count thresholds (**B**), and movement-evoked pain ratings (**C**). Panel **A** shows between-person variability in step counts associated with the 50th, 75th, 90th, and 100th quantiles during non-sedentary minutes (step count ≥ 10 steps per minute). 50th quantile refers to the individual's average (non-sedentary) step count; 100th quantile refers to the individual's maximum steps per minute. Panel **B** shows person-level average heart rate associated with various step count thresholds (spm = steps per minute). Dots indicate person-level average heart rate for each threshold, across all available 6-minute windows. For each window, percent of maximum heart rate was calculated as observed average heart rate divided by the individual's maximum heart rate (208 - 0.7*Age). Panel **C** shows person-level variability in post-activity pain recorded within 3 hours of an activity bout (defined as at least 6 consecutive minutes walking at a speed ≥ 67spm). Grey vertical lines are observed post-activity pain ratings. Shaded areas are person-level probabilities (darker = higher probability).

3.2. Defining MEP

Of the full sample, 93 participants (82%) had any pain ratings following a 6-minute activity bout based on the 67spm threshold (T = 1060 total observations). Among individuals with any post-activity pain ratings, the median number of observations was 7 (*SD* = 11), corresponding to an average of 14% (*SD* = 13%) of the individual's EMA pain ratings.

To define MEP, we first removed sequential observations (t = 284, 27%) so that lag-1 pain more closely mimicked PAR. The remaining 776 post-activity pain observations across 93 individuals were used for the first operationalization of MEP (absolute post-activity pain). Calculating MEP change scores requires a lagged pain observation. We thus retained 537 MEP change scores across 87 individuals after removing observations that occurred at the first survey of the day when lag-1 pain was not available (t = 107, 38% of missing lag-1 pain occasions) or later in the day when lag-1 EMA was not completed (t = 132).

3.3. Reliability of MEP

Variability in post-activity pain is visualized in Figure 2C. The average within-person standard deviation of post-activity pain was 10.64 (*SD* = 7.87, *Min* = 0, *Max* = 46.19). The ICC of .75 suggests approximately 25% of the variance in post-activity pain occurred within individuals. The average within-person standard deviation of MEP change score was 11.87 (*SD* = 8.61, *Min* = 0, *Max* = 47.96). The ICC of .08 suggests that approximately 92% of the variability in MEP change score occurred within individuals.

3.4. Multilevel predictors of absolute post-activity pain

Results of the univariate mixed-effects models predicting post-activity pain are presented in Figure 3. Individuals with greater pain on average tended to report higher post-activity pain (β = 22.51, $p < .001$). Average pain accounted for 70.5% of the variance in post-activity pain. When examined over time, post-activity pain tended to be higher when lag-1 pain was above average for the individual. There was also a quadratic effect of time, such that post-activity pain appeared highest at the mid-afternoon survey and lowest in the evening, though this accounted for <1% of variance in post-activity pain ratings. Post-activity pain ratings were not influenced by the amount of MVPA within the 3-hour window (Fig 3D) or time lag from activity to EMA (Fig 3F). There was a trend towards higher post-activity pain ratings when individuals had been more active so far that day (p = .064, Fig 3E).

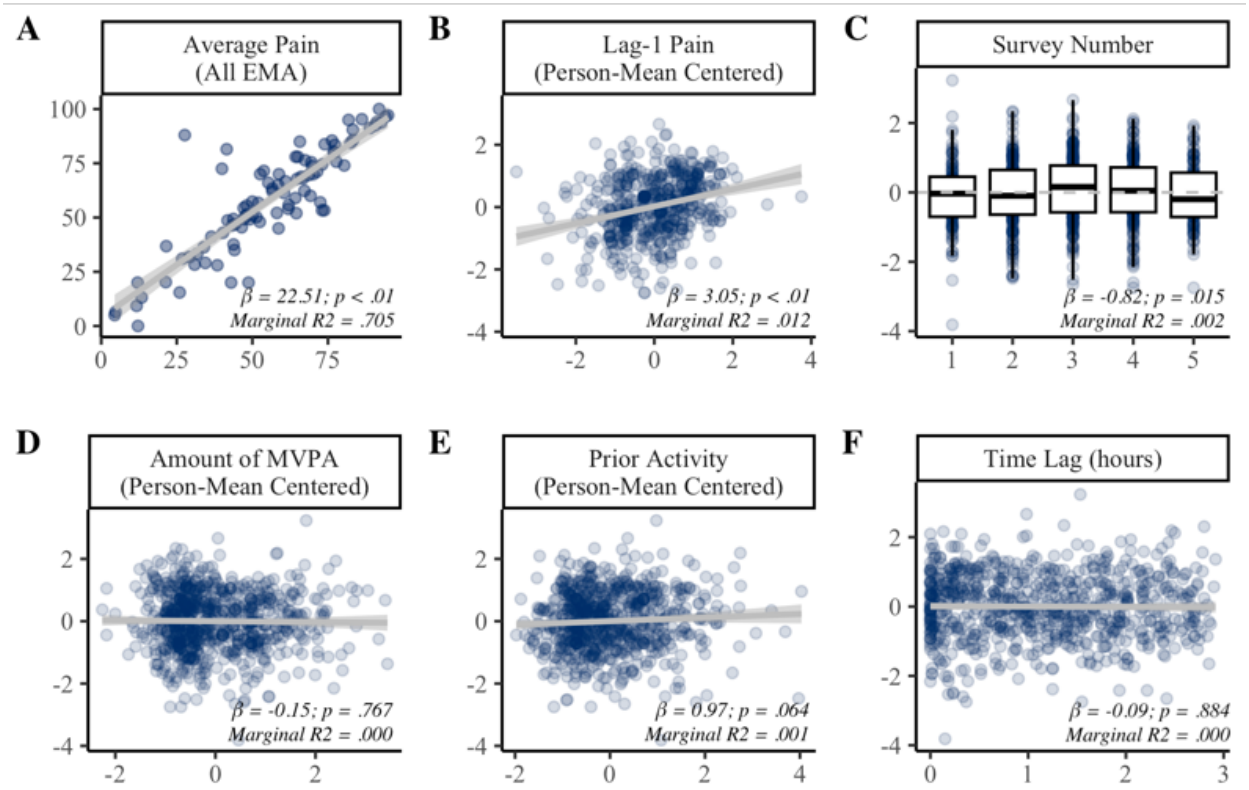

Figure 3. Predictors of post-activity pain ratings. Panel **A** shows the between-persons association of average pain across all EMAs (x-axis) with average post-activity pain (y-axis). Panels **B-F** show the effect of time-varying covariates (x-axis) on within-person variability in post-activity pain ratings (y-axis). In Panels B-F, post-activity pain is person mean centered such that 0 = average for the individual. MVPA = moderate-to-vigorous physical activity.

Results of the multivariate model are presented in Table 2. Average pain continued to be the strongest predictor of post-activity pain. Consistent with univariate models, lag-1 pain, time of day, and cumulative activity prior to the lag-1 pain observation accounted for unique within-person variance in post-activity pain ratings over time. The multivariate model accounted for 97.3% of the between-person variance and 7.4% of the within-person variance in post-activity pain.

Table 2. *Multivariate predictors of post-activity pain (N = 93, T = 776)*

Predictor	Estimate	Posterior SD	P	95% CI
[Intercept]	60.460	0.861	**<.001**	[58.762, 62.150]
Average Pain	22.361	0.740	**<.001**	[20.844, 23.782]
Lag-1 Pain	3.380	0.629	**<.001**	[2.137, 4.603]
Survey^2	-0.853	0.325	**.005**	[-1.490, -0.218]
Amount of MVPA	-0.711	0.554	.100	[-1.789, 0.386]
Prior Activity	1.551	0.562	**.003**	[0.442, 2.642]
Time Lag	-0.297	0.616	.442	[-1.523, 0.906]
R^2 Between	0.973	0.017		[0.930, 0.994]
R^2 Within	0.074	0.022		[0.037, 0.122]

Note. SD = Standard Deviation; P = Bayesian one-tailed p-value based on the posterior distribution; CI = Credible Interval

3.4. Multilevel predictors of MEP change scores

Results of the univariate mixed-effects models predicting MEP change scores are presented in Figure 4. Average pain was uncorrelated with average MEP change score. When examined over time, MEP change score tended to be lower when lag-1 pain was elevated ($\beta = -8.49$, $p < .001$). Lag-1 pain accounted for 22% of the variance in MEP change score. There was also a linear effect of time, such that MEP change score tended to decrease across surveys 2 (e.g., 12pm) to 5 (e.g., 9am). MEP change scores were not influenced by the amount of MVPA within the 3-hour window (Fig 4D) or time lag from activity to EMA (Fig 4F). However, MEP change scores did tend to be elevated when individuals had been more active so far that day ($\beta = 2.54$, $p = .001$). Survey number and prior activity each accounted for approximately 2% of the variance in MEP change score.

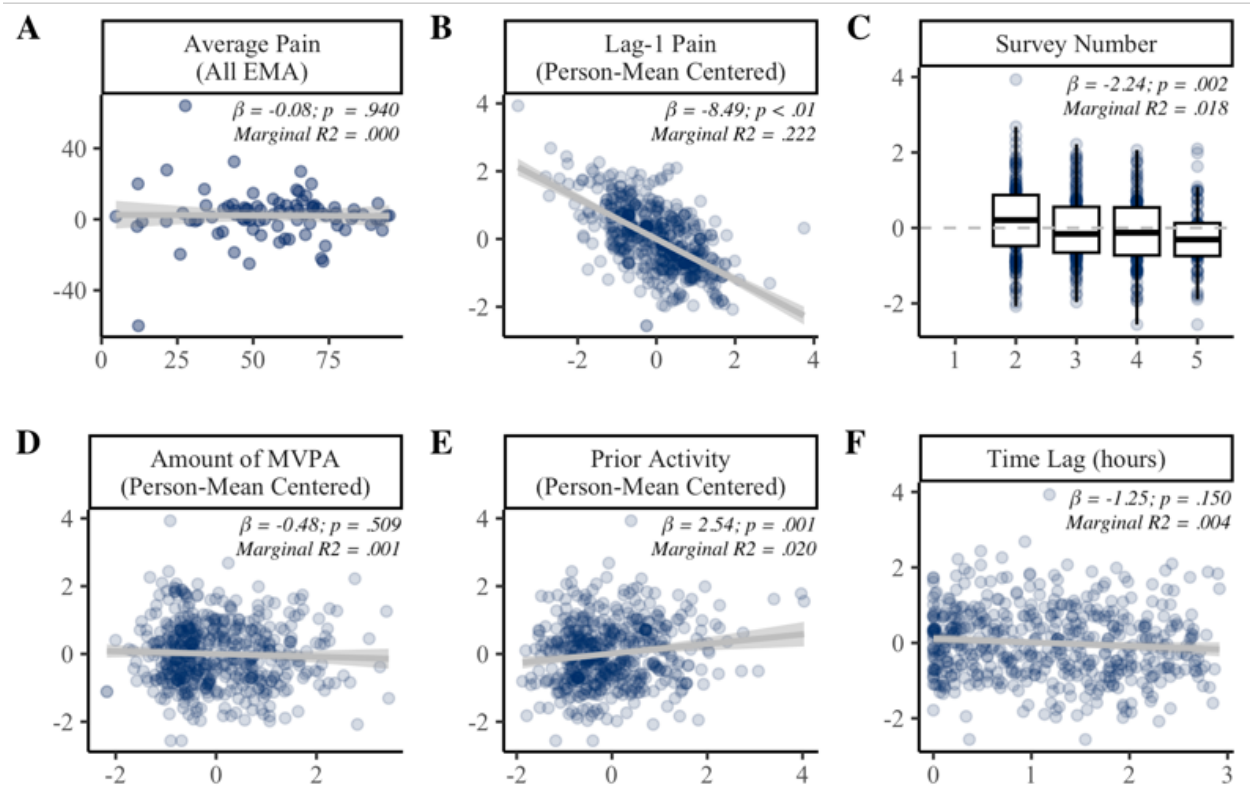

Figure 4. Predictors of MEP change scores. Panel **A** shows the between-persons association of average pain across all EMAs (x-axis) with average MEP change score (y-axis). Panels **B-F** show the effect of time-varying covariates (x-axis) on within-person variability in MEP change scores (y-axis). In Panels B-F, MEP change score is person mean centered such that 0 = average for the individual. MVPA = moderate-to-vigorous physical activity.

Results of the multivariate model are presented in Table 3. Consistent with univariate models, lag-1 pain, time of day, and cumulative activity prior to the lag-1 pain observation accounted for unique variance in MEP change scores. The multivariate model accounted for 1.4% of the between-person variance and 28.2% of the within-person variance in MEP change scores.

Table 3. *Multivariate predictors of MEP change scores (N = 87, T = 537)*

Predictor	Estimate	Posterior SD	P	95% CI
[Intercept]	6.424	2.327	**.004**	[1.823, 10.928]
Average Pain	-0.099	1.105	.468	[-2.309, 2.032]
Lag-1 Pain	-8.320	0.658	**<.001**	[-9.630, -7.055]
Survey	-1.298	0.634	**.021**	[-2.554, -0.052]
Amount of MVPA	-0.936	0.672	.081	[-2.268, 0.391]
Prior Activity	2.033	0.726	**.003**	[0.612, 3.462]
Time Lag	-0.049	0.792	.474	[-1.620, 1.516]
R^2 Between	0.014	0.041		[0.000, 0.148]
R^2 Within	0.282	0.045		[0.216, 0.349]

Note. SD = Standard Deviation; P = Bayesian one-tailed p-value based on the posterior distribution; CI = Credible Interval

3.5. Variability in PAR

As shown in Figure 5, we observed wide variability in PAR. When fewer than 1 steps per minute were recorded in the 30 minutes (T = 1667) and 3 hours prior to PAR observations (T = 829), median EMA pain ratings were 55 and 56 out of 100, and 21% of observations exceeded 80/100, indicating severe pain.

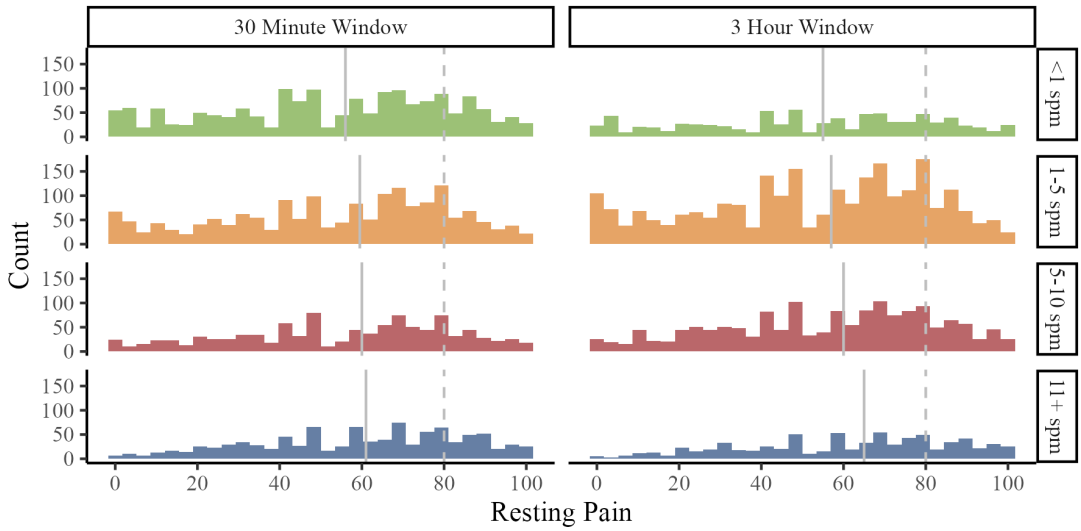

Figure 5. Variability in resting pain based on average steps per minute (spm) in the 30 minutes and 3 hours prior to EMA pain ratings. Solid grey lines = group median; dotted grey lines = severe pain cut-off (80/100).

4. Discussion

This study leveraged digital technology to assess movement-evoked pain (MEP) under free-living conditions among adults with moderate-to-severe chronic pain scheduled to receive elective lumbar/thoracolumbar fusion surgery. Numerous studies have assessed MEP as absolute post-activity pain ratings, such that individuals who report higher post-activity pain are considered to have greater MEP. We observed good between-person reliability of absolute post-activity pain scores. However, there was a high degree of overlap between post-activity pain ratings and average pain levels calculated from EMA pain ratings, which reduce risk of recall bias associated with retrospective pain reports[50]. Post-activity pain ratings also appeared elevated when pre-activity pain was elevated relative to the individual's average. Post-activity pain ratings therefore appear to capture overall disability and day-to-day fluctuations in pain, rather than the relationship between movement and pain.

MEP change scores are commonly used to isolate MEP from PAR by subtracting PAR from post-activity pain ratings. In the current study, MEP change scores had poor reliability when assessed naturalistically and over time. Factors related to the uncontrolled nature of digital assessment (e.g., amount of MVPA, time from activity to pain report) did not explain significant variance in post-activity pain ratings or MEP change scores. MEP change scores were highly influenced by lag-1 pain observations meant to mimic PAR, such that change scores were lower when lag-1 pain observations were elevated. This is suggestive of ceiling effects, wherein individuals experience lower MEP change scores because PAR is high. We acknowledge that some lag-1 pain observations may not truly capture PAR, as participants were not instructed to rest prior to the EMA pain rating. However, under these naturalistic conditions, we observed wide variability in PAR reports, including when no or very little activity was detected in the 30 minutes and 3 hours prior to PAR observations. Our findings suggest ceiling effects are common, especially when demand characteristics are absent, as over 20% of pain reports following true resting periods indicated severe pain.

With repeated MEP observations, we were able to assess the impact of factors that could contaminate MEP tasks conducted in the lab or clinic. By capturing MEP change scores and post-activity pain ratings at various times of day, we observed time trends that suggest both metrics may be lowest in the evening, and post-activity pain ratings may be highest mid-afternoon. Because lag-1 pain observations were not available at the first survey of the day, MEP change scores are only available for surveys 2-5 (e.g., 12pm-9pm). However, we observed a linear effect of time, such that MEP change scores tended to decrease over this time period. We also observed an additive effect of prior activity, such that individuals experienced greater MEP when they had been more active so far that day. Although each covariate accounted for only a small proportion of variance in MEP, they remained statistically significant in multivariate models, which suggests additive effects of contaminating factors.

Interestingly, our analyses did not suggest systematic differences in post-activity pain ratings or MEP change scores collected less proximally to activity bouts. Relatively little research has investigated temporal aspects of MEP, including whether MEP is distinct from delayed-onset soreness that may occur hours to days after physical activity[7]. Our findings suggest MEP may not be temporally bound within a short window after pain-evoking movement. Post-activity pain

ratings and MEP change scores collected immediately after or during activity bouts appear largely similar to ratings collected up to 3 hours after activity. However, we cannot make any strong conclusions without repeated assessments tied to the same activity bout. Future studies can utilize event-contingent monitoring, wherein pain assessments are triggered by physical activity[13]. Additionally, because we only examined fixed effects, it remains possible that the temporality of MEP (e.g., rate of decay) varies across individuals, which may have meaningful implications for phenotyping and treatment personalization.

This study has several strengths. We evaluated variability in two common operationalizations of MEP across days and times of day among individuals with moderate-to-severe chronic pain. We additionally identified factors that may influence both absolute post-activity pain and MEP change scores when assessed only once, filling known gaps in the growing literature on MEP[7,8]. Participants were largely adherent to the assessment procedures, completing over 84% of EMAs on average and wearing the Fitbit for several weeks. We developed a novel framework to evaluate MEP ecologically by extracting pain ratings following bouts of MVPA lasting at least 6 minutes. By mimicking the 6MWT under free-living conditions, we reduced demand characteristics, as participants were not aware that their Fitbit and EMA data would be combined to evaluate MEP.

This study also has limitations. Our sample was predominantly White, middle-aged, and all participants had moderate-to-severe chronic pain warranting lumbar/thoracolumbar fusion surgery, which may limit generalizability of findings. It will be important to replicate results in larger, more diverse samples with chronic back pain, including individuals who are not surgical candidates and those with other types of musculoskeletal pain (e.g., hip or knee pain). We also relied on consumer-grade wearable devices to increase feasibility. Fitbits are the most commonly used wearable devices in research settings[15]. Although several studies find that Fitbits accurately measure steps, there is concern about overestimation under free-living conditions[12,15,37]. We analyzed heart rate data to ensure that the activity threshold was high enough to suggest at least moderate levels of physical activity. However, it will be important for future studies to further define MVPA among adults with moderate-to-severe musculoskeletal pain.

Given observed variability in MEP, especially MEP change scores, we recommend the continued development of digital methodologies for assessing MEP. Prior studies suggest that the within-person relationship between physical activity and pain is complex, with a number of factors including overall activity levels, genotype, and psychosocial factors moderating relationships across persons[11,32,56]. Thus, further within-person assessment is recommended, especially using methods that consider how relationships between physical activity and pain vary across persons[56]. To this end, we developed a dynamic index of MEP based on continuous Fitbit and EMA data for predicting surgical outcomes[19]. Using multilevel dynamic structural equation modeling[3], we estimated the degree to which engaging in more activity than typical for the individual is associated with subsequent increases in pain. This novel framework disentangles MEP from average pain, as someone with low average pain can still experience reliable increases in pain following activity. Additionally, the approach utilizes all available data, instead of only occasions when the individual completes a standardized task or engages in activity above a pre-specified threshold. All individuals who provide activity and EMA data have an MEP estimate indexing the degree to which pain increases with activity, after accounting for important

covariates (e.g., time of day). In our preliminary analyses, this dynamic index of MEP was a predictor of one-month spine surgery outcomes[19].

Digital tools can also be used to administer task-based assessment of MEP repeatedly and under naturalistic settings. A smartphone app version of the 6MWT is increasingly being used to evaluate functional impairment[31,40,43,45]. Future studies can use this approach to assess MEP by collecting pain ratings before and after the remotely delivered 6MWT. This procedure can be repeated across multiple days and at different times per day to further improve understanding of fluctuations in MEP. Additionally, remote protocols can be developed to capture pain in response to physical activities outside of walking. Although the 6MWT is a common MEP task, other movements (e.g., chair rises, trunk rotation, and standing forward reach) are also relevant to MEP in back pain patients[16]. Some more recent studies define MEP as the aggregate of post-activity pain ratings across several tasks, minus the pre-activity pain rating[23,46]. This procedure takes into account resting pain but will be less subject to ceiling effects given that post-activity pain ratings are weighted more heavily than pre-activity pain.

Overall, our study highlights the need for further research to evaluate the reliability and construct validity of commonly used methods for assessing MEP. The two approaches described above are likely most effective when used together to capture individual differences in the dynamic relationship between activity and pain, and to compare these patterns with task-based MEP assessments delivered remotely and repeated across multiple days and times of day. Phenotyping chronic pain remains a significant challenge, but improving the precision of MEP measurement may enhance efforts to identify meaningful subgroups within heterogeneous chronic pain populations.

Acknowledgments

This study was funded by Barnes Foundation, Big Ideas, Scoliosis Research Society, AO Spine North America, and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (1K23AR082986-01A1). Madelyn Frumkin was supported by the National Institute of Mental Health (F31MH124291). The sponsors of this study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript. All authors had full access to all the data in the study, agreed to submit for publication, and approved the final manuscript. The authors have no conflicts of interest to report.

References

1. Ader DN: Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care 45:S1, 2007.

2. Amtmann D, Cook KF, Jensen MP, Chen W-H, Choi S, Revicki D, Cella D, Rothrock N, Keefe F, Callahan L, Lai J-S: Development of a PROMIS item bank to measure pain interference. PAIN 150:173–82, 2010.

3. Asparouhov T, Hamaker EL, Muthén B: Dynamic structural equation models. Structural Equation Modeling 25:359–88, 2018.

4. Barnett A, van den Hoek D, Barnett D, Cerin E: Measuring moderate-intensity walking in older adults using the ActiGraph accelerometer. BMC Geriatr 16:211, 2016.

5. Boonstra AM, Stewart RE, Köke AJA, Oosterwijk RFA, Swaan JL, Schreurs KMG, Schiphorst Preuper HR: Cut-Off Points for Mild, Moderate, and Severe Pain on the Numeric Rating Scale for Pain in Patients with Chronic Musculoskeletal Pain: Variability and Influence of Sex and Catastrophizing. Front Psychol 7:1466, 2016.

6. Breivik H, Borchgrevink PC, Allen SM, Rosseland LA, Romundstad L, Breivik Hals EK, Kvarstein G, Stubhaug A: Assessment of pain. British Journal of Anaesthesia Narnia; 101:17–24, 2008.

7. Butera KA, Chimenti RL, Alsouhibani AM, Berardi G, Booker SQ, Knox PJ, Post AA, Merriwether EN, Wilson AT, Simon CB: Through the Lens of Movement-Evoked Pain: A Theoretical Framework of the “Pain-Movement Interface” to Guide Research and Clinical Care for Musculoskeletal Pain Conditions. The Journal of Pain 25:104486, 2024.

8. Corbett DB, Simon CB, Manini TM, George SZ, Riley JL, Fillingim RB: Movement-evoked pain: Transforming the way we understand and measure pain. PAIN 160:757, 2019.

9. Crow JA, Joseph V, Miao G, Goodin BR, Sibille KT, Cardoso J, Bartley EJ, Staud R, Fillingim RB, Booker SQ: A domain-oriented approach to characterizing movement-evoked pain. PAIN Reports 9:e1158, 2024.

10. Daniels MJ, Hogan JW: Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis. New York: Chapman and Hall/CRC; 2008.

11. Davis TJ, Hevel DJ, Dunton GF, Maher JP: Bidirectional Associations Between Physical Activity and Pain Among Older Adults: An Ecological Momentary Assessment Study. , 2022 [cited 2025 July 27]. Available from: https://journals.humankinetics.com/view/journals/japa/31/2/article-p240.xml

12. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, Hamilton CB, Li LC: Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. JMIR mHealth and uHealth 6:e10527, 2018.

13. Fritz J, Piccirillo ML, Cohen ZD, Frumkin M, Kirtley O, Moeller J, Neubauer AB, Norris LA, Schuurman NK, Snippe E, Bringmann LF: So You Want to Do ESM? 10 Essential Topics for Implementing the Experience-Sampling Method. Advances in Methods and Practices in Psychological Science SAGE Publications Inc; 7:25152459241267912, 2024.

14. Frumkin MR, Rodebaugh TL: The role of affect in chronic pain: A systematic review of within-person symptom dynamics. Journal of Psychosomatic Research 147:110527, 2021.

15. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, Buote R, Van Heerden D, Luan H, Cullen K, Slade L, Taylor NGA: Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. JMIR mHealth and uHealth JMIR Publications; 8:e18694, 2020.

16. Fullwood D, Means S, Merriwether EN, Chimenti RL, Ahluwalia S, Booker SQ: Toward understanding movement-evoked pain (MEP) and its measurement: A scoping review. Clin J Pain 37:61–78, 2021.

17. Gilron I, Lao N, Carley M, Camiré D, Kehlet H, Brennan TJ, Erb J: Movement-evoked Pain versus Pain at Rest in Postsurgical Clinical Trials and in Meta-analyses: An Updated Systematic Review. Anesthesiology 140:442–9, 2024.

18. Goossens Z, Van Stallen A, Vermuyten J, De deyne M, Rice D, Runge N, Huysmans E, Vantilborgh T, Nijs J, Mairesse O, De Baets L: Day-to-day associations between pain intensity and sleep outcomes in an adult chronic musculoskeletal pain population: A systematic review. Sleep Medicine Reviews 79:102013, 2025.

19. Greenberg JK, Frumkin M, Xu Z, Zhang J, Javeed S, Zhang JK, Benedict B, Botterbush K, Yakdan S, Molina CA, Pennicooke BH, Hafez D, Ogunlade JI, Pallotta N, Gupta MC, Buchowski JM, Neuman B, Steinmetz M, Ghogawala Z, Kelly MP, Goodin BR, Piccirillo JF, Rodebaugh TL, Lu C, Ray WZ: Preoperative mobile health data improve predictions of recovery from lumbar spine surgery. Neurosurgery 95:617, 2024.

20. Greenberg JK, Frumkin MR, Javeed S, Zhang JK, Dai R, Molina CA, Pennicooke BH, Agarwal N, Santiago P, Goodwin ML, Jain D, Pallotta N, Gupta MC, Buchowski JM, Leuthardt EC, Ghogawala Z, Kelly MP, Hall BL, Piccirillo JF, Lu C, Rodebaugh TL, Ray WZ: Feasibility and acceptability of a preoperative multimodal mobile health assessment in spine surgery candidates. Neurosurgery 92:538–46, 2023.

21. Hall KS, Howe CA, Rana SR, Martin CL, Morey MC: METs and Accelerometry of Walking in Older Adults: Standard versus Measured Energy Cost. Med Sci Sports Exerc 45:574–82, 2013.

22. Ham SA, Reis JP, Strath SJ, Dubose KD, Ainsworth BE: Discrepancies between methods of identifying objectively determined physical activity. Med Sci Sports Exerc 39:52–8, 2007.

23. Kahan R, Woznowski-Vu A, Huebner JL, Pieper CF, Goode AP, George SZ, Wideman TH, Kraus VB, Colón-Emeric C, Simon CB: Psychological and immunological associations with movement-evoked low back pain among older adults. PAIN Reports 10:e1262, 2025.

24. Knezevic NN, Nader A, Pirvulescu I, Pynadath A, Rahavard BB, Candido KD: Circadian pain patterns in human pain conditions – A systematic review. Pain Practice 23:94–109, 2023.

25. Knox PJ, Simon CB, Pohlig RT, Pugliese JM, Coyle PC, Sions JM, Hicks GE: A Standardized Assessment of Movement-evoked Pain Ratings Is Associated With Functional Outcomes in Older Adults With Chronic Low Back Pain. The Clinical Journal of Pain 38:241, 2022.

26. Knox PJ, Simon CB, Pohlig RT, Pugliese JM, Coyle PC, Sions JM, Hicks GE: Construct validity of movement-evoked pain operational definitions in older adults with chronic low back pain. Pain Medicine 24:985–92, 2023.

27. Knox PJ, Simon CB, Pohlig RT, Pugliese JM, Coyle PC, Sions JM, Hicks GE: Movement-Evoked Pain Versus Widespread Pain: A Longitudinal Comparison in Older Adults With Chronic Low Back Pain From the Delaware Spine Studies. The Journal of Pain 24:980–90, 2023.

28. Kuznetsova A, Brockhoff PB, Christensen RH: lmerTest package: tests in linear mixed effects models. Journal of statistical software 82:1–26, 2017.

29. Lambin DI, Thibault P, Simmonds M, Lariviere C, Sullivan MJL: Repetition-induced activity-related summation of pain in patients with fibromyalgia. PAIN 152:1424–30, 2011.

30. Lozano-Meca JA, Gacto-Sánchez M, Montilla-Herrador J: Movement-evoked pain is not associated with pain at rest or physical function in knee osteoarthritis. European Journal of Pain 28:987–96, 2024.

31. Maldaner N, Sosnova M, Zeitlberger AM, Ziga M, Gautschi OP, Regli L, Weyerbrock A, Stienen MN: Evaluation of the 6-minute walking test as a smartphone app-based self-measurement of objective functional impairment in patients with lumbar degenerative disc disease. Journal of Neurosurgery 33:779–88, 2020.

32. Martire LM, Wilson SJ, Small BJ, Conley YP, Janicki PK, Sliwinski MJ: COMT and OPRM1 genotype associations with daily knee pain variability and activity induced pain. Scandinavian Journal of Pain De Gruyter; 10:6–12, 2016.

33. Medicine AC of S: ACSM's guidelines for exercise testing and prescription [Internet]. Lippincott williams & wilkins; 2013 [cited 2025 Jan 21]. Available from: https://books.google.com/books?hl=en&lr=&id=hhosAwAAQBAJ&oi=fnd&pg=PP1&dq=guidelines+for+exercise+testing&ots=llD92I-TVu&sig=7XqfUjXYXby4Nw-IeIXb2onWjHs

34. Mun CJ, Burgess HJ, Sears DD, Parthasarathy S, James D, Altamirano U, Sajith S, Lakhotia A, Fillingim RB, Youngstedt SD: Circadian Rhythm and Pain: a Review of Current Research and Future Implications. Curr Sleep Medicine Rep 8:114–23, 2022.

35. Palada V, Gilron I, Canlon B, Svensson CI, Kalso E: The circadian clock at the intercept of sleep and pain. PAIN 161:894, 2020.

36. Quartana PJ, Campbell CM, Edwards RR: Pain catastrophizing: A critical review. Expert Review of Neurotherapeutics 9:745–58, 2009.

37. Redenius N, Kim Y, Byun W: Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Medical Research Methodology 19:29, 2019.

38. Riddle DL, Dumenci L: Preoperative measures of pain at rest and movement-evoked pain in knee arthroplasty: Associations with pain and function outcome trajectories from a prospective multicentre longitudinal cohort study. European Journal of Pain 29:e4723.

39. Roberts-Lewis SF, White CM, Ashworth M, Rose MR: Validity of Fitbit activity monitoring for adults with progressive muscle diseases. Disability and Rehabilitation 44:7543–53, 2022.

40. Salvi D, Poffley E, Orchard E, Tarassenko L: The Mobile-Based 6-Minute Walk Test: Usability Study and Algorithm Development and Validation. JMIR mHealth and uHealth 8:e13756, 2020.

41. Sayers A, Wylde V, Lenguerrand E, Beswick AD, Gooberman-Hill R, Pyke M, Dieppe P, Blom AW: Rest Pain and Movement-Evoked Pain as Unique Constructs in Hip and Knee Replacements. Arthritis Care & Research 68:237–45, 2016.

42. Schneider S, Stone AA, Schwartz JE, Broderick JE: Peak and end effects in patients daily recall of pain and fatigue: A within-subjects analysis. Journal of Pain 12:228–35, 2011.

43. Schubert C, Archer G, Zelis JM, Nordmeyer S, Runte K, Hennemuth A, Berger F, Falk V, Tonino PAL, Hose R, Ter Horst H, Kuehne T, Kelm M: Wearable devices can predict the outcome of standardized 6-minute walk tests in heart disease. npj Digit Med 3:92, 2020.

44. Silva GS, Yang H, Collins JE, Losina E: Validating Fitbit for Evaluation of Physical Activity in Patients with Knee Osteoarthritis: Do Thresholds Matter? ACR Open Rheumatol 1:585–92, 2019.

45. Simmich J, Andrews NE, Claus A, Murdoch M, Russell TG: Assessing a GPS-Based 6-Minute Walk Test for People With Persistent Pain: Validation Study. JMIR Formative Research 8:e46820, 2024.

46. Simon CB, Hicks GE, Pieper CF, Byers Kraus V, Keefe FJ, Colón-Emeric C: A Novel Movement-Evoked Pain Provocation Test for Older Adults With Persistent Low Back Pain: Safety, Feasibility, and Associations With Self-reported Physical Function and Usual Gait Speed. The Clinical Journal of Pain 39:166, 2023.

47. Slipetz LR, Falk ,A., and Henry TR: Missing Data in Discrete Time State-Space Modeling of Ecological Momentary Assessment Data: A Monte-Carlo Study of Imputation Methods. Multivariate Behavioral Research Routledge; 0:1–16.

48. Srikandarajah S, Gilron I: Systematic review of movement-evoked pain versus pain at rest in postsurgical clinical trials and meta-analyses: A fundamental distinction requiring standardized measurement. PAIN 152:1734–9, 2011.

49. Stone AA, Broderick JE, Junghaenel DU, Schneider S, Schwartz JE: PROMIS fatigue, pain intensity, pain interference, pain behavior, physical function, depression, anxiety, and anger scales demonstrate ecological validity. Journal of Clinical Epidemiology Pergamon; 74:194–206, 2016.

50. Stone AA, Schwartz JE, Broderick JE, Shiffman SS: Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. Personality and Social Psychology Bulletin 31:1340–6, 2005.

51. Stone AA, Shiffman S: Ecological momentary assessment (EMA) in behavioral medicine. Annals of Behavioral Medicine Lawrence Erlbaum; 16:199–202, 1994.

52. Sullivan MJL, Larivière C, Simmonds M: Activity-related summation of pain and functional disability in patients with whiplash injuries. PAIN® 151:440–6, 2010.

53. Sullivan MJL, Thibault P, Andrikonyte J, Butler H, Catchlove R, Larivière C: Psychological influences on repetition-induced summation of activity-related pain in patients with chronic low back pain. PAIN® 141:70–8, 2009.

54. Tanaka H, Monahan KD, Seals DR: Age-predicted maximal heart rate revisited. J Am Coll Cardiol 37:153–6, 2001.

55. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De Bourdeaudhuij I, Ewald B, Gardner AW, Hatano Y, Lutes LD, Matsudo SM, Ramirez-Marrero FA, Rogers LQ, Rowe DA, Schmidt MD, Tully MA, Blair SN: How many steps/day are enough? For older adults and special populations. Int J Behav Nutr Phys Act 8:80, 2011.

56. Tynan M, Virzi N, Wooldridge JS, Morse JL, Herbert MS: Examining the Association Between Objective Physical Activity and Momentary Pain: A Systematic Review of Studies Using Ambulatory Assessment. The Journal of Pain 25:862–74, 2024.