OPEN Research Article

The Oswestry Disability Index and 12-Item Short Form Health Survey Physical Component Scores Are Not Affected by Recall Bias in Posterior Lumbar Spine Surgery Patients: A Prospective Study Using Data From Fitness Trackers

Justin Tiao, BS
Ashley M. Rosenberg, BS
Dennis M. Bienstock, MD
Brittany Sacks, BS
Charles Laurore, BA, MS
Michael Herrera, MD
Dhruv S. Shankar, MD
Wesley H. Bronson, MD
Saad B. Chaudhary, MD
Jashvant Poeran, MD, PhD
James C. Iatridis, PhD
Andrew C. Hecht, MD

From the Department of Orthopaedic Surgery, Icahn School of Medicine at Mount Sinai, New York, NY (Mr. Tiao, Ms. Rosenberg, Dr. Bienstock, Ms. Sacks, Dr. Laurore, Dr. Herrera, Dr. Bronson, Dr. Chaudhary, Dr. Iatridis, and Dr. Hecht); the Department of Orthopaedic Surgery, University of Washington, Seattle, WA (Dr. Shankar); and the Department of Anesthesiology, Hospital for Special Surgery, New York, NY (Dr. Poeran).

Correspondence to Dr. Hecht: andrew.hecht@mountsinai.org

Data availability: All data was available to all co-authors for review, analysis, and manuscript preparation.

Trial registration numbers: N/A, IRB-18-00023.

JAAOS Glob Res Rev 2025;9: e24.00185

DOI: 10.5435/JAAOSGlobal-D-24-00185

ABSTRACT

Objective: To assess the effects of recall bias on prospectively collected patient-reported outcome (PRO) measures after lumbar laminectomy by analyzing correlations between PROs and step counts in measurement windows preceding PRO measurement.
Methods: Responses to the 12-item Short Form Health Survey (SF-12) and the Oswestry Disability Index (ODI) were collected postoperatively from 22 patients. Accelerometers recorded daily step counts. Median and maximum step counts were calculated for windows (1 day, 3 days, 1 week, and 2 weeks) preceding PRO measurement. Spearman rank correlation coefficients between PROs and step counts were calculated.
Results: Median and maximum step counts from 1- and 2-week windows more consistently correlated with SF-12 Physical Component Scores scores than 1- and 3-day windows over the postoperative period. Median steps from 1-, 2-week, and 3-day windows correlated more with ODI scores than the 1-day window. Maximum steps from 1- and 2-week windows correlated more than 1- and 3-day windows.
Discussion: PROs had higher concordance with step counts from the 1 week and 2 weeks before PRO measurement than the 1 day and 3 days prior. We therefore conclude that SF-12 and ODI are not markedly affected by recall bias because scores are not skewed by the events of the days immediately preceding measurement.

Patient-reported outcome (PRO) measures provide clinicians with a patient-centered estimate of postoperative recovery and surgical success. In a review of PROs used in spine surgery, McCormick et al[1] determined that a comprehensive questionnaire should include a measure addressing general health, such as the 12-item Short Form Health Survey (SF-12), a pain scale, and a disease-specific outcome measure, such as the Oswestry Disability Index (ODI), which is the most validated outcome measure for lumbar pathologies. As PROs are only obtained at a few discrete points in time, they may be influenced by recall bias, which we define in this study as whether the events of the days immediately preceding measurement affect the PRO score. The time point at which PRO measurements best represent the patient experience has yet to be elucidated.

PROs can be collected prospectively and retrospectively, each of which present unique challenges related to recall bias. In prospective PRO collection, an assessment of preoperative condition is collected before surgery and postoperative condition is assessed at set time points after surgery. In retrospective PRO collection, patients are asked to recall their perceived preoperative status at a time point after surgery to serve as a comparison to their self-reported status before surgery.[2] Both the retrospective and prospective collection methods are susceptible to recall bias: in the retrospective model, this includes errors in patients' recollection of their preoperative status during the postoperative period, whereas in the prospective model, this includes being overly influenced by the events just preceding PRO collection.[3]

Many factors can influence a patient's ability to accurately recall an event, ranging from the recency of the event to the patient's mood at the time of PRO collection.[4] In cervical spine surgery, recall bias was identified as a contributing factor to notable deviations between preoperative function measured preoperatively and preoperative function estimated from recall at 4.7 and 22.1 months postoperatively.[5] Notably, patients described their preoperative neck pain, arm pain, and disability as being markedly worse than actual preoperative levels when the PRO was assessed preoperatively.[5] Similarly, in a sample of lumbar spine surgery patients, Aleem et al[6] also found that patients recalled markedly worse preoperative back pain, leg pain, and disability at a mean follow-up of 43.3 months after surgery. Thus, it is evident that the proximity in time of PRO measurement to the period that is being queried is important. Surprisingly, little is known about the effect that recall bias has on the prospective model of PRO collection wherein the recency effect of events that occur just before PRO measurement may influence patients' scores.

In this study, we propose to use data from patient-worn accelerometers, which continuously measure physical function and provide detailed objective data on patient activity, to assess if PROs are most representative of recency of physical activity. PROs have been shown to be correlated with step counts in a prior study supporting the possibility that common underlying physical function is captured by both step counts and PROs; however, that study concluded that PROs may better capture the subjective elements of recovery over step counts.[7] In this study, by comparing qualitative PRO results evaluating patient disability to objective step count measurements in the days and weeks before PRO measurement, we can evaluate if PROs suffer from prospective recall bias and identify the time before measurement that PROs best represent. The aim of this study was to use step count data from wearable fitness trackers to determine which of the 1-day, 3-day, 1-week, or 2-week data before PRO measurement most consistently correlated with PRO scores and to assess the potential effect of recall bias on PROs measured postoperatively after lumbar laminectomy surgery. A PRO measure was considered susceptible to recall bias if daily step counts immediately preceding PRO measurement (i.e., 1 to 3 days before) markedly correlated with the PRO score, whereas step counts recorded further in time from measurement were not. We also aimed to determine whether median or maximum step counts from measurement windows best correlated with PRO measurements to ascertain whether patients tend to report the experience of an average day in the period queried or if they were affected by outlier experiences. As recall bias has been shown to affect PRO measurements in prior literature, we hypothesized that maximum step counts from the 1- and 3-day windows before PRO collection would be better correlated with PRO scores than step counts from the 1- and 2-week windows before collection over the course of the study period, thus identifying that recall bias affects PROs measured after lumbar laminectomy surgery.

Supported by the Mount Sinai Spine Hospital Research Fund and in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases [NIH RO1 AR057397].

None of the following authors or any immediate family member has received anything of value from or has stock or stock options held in a commercial company or institution related directly or indirectly to the subject of this article: Mr. Tiao, Ms. Rosenberg, Dr. Bienstock, Ms. Sacks, Dr. Laurore, Dr. Herrera, Dr. Shankar, Dr. Bronson, Dr. Chaudhary, Dr. Poeran, Dr. Iatridis, and Dr. Hecht.

J. Tiao and A. M. Rosenberg are co-first authors.

Methods

Study Design/Inclusion and Exclusion Criteria

Following institutional review board approval, we prospectively recruited patients with lumbar stenosis undergoing lumbar laminectomy (with no more than two level fusions) from a high-volume, urban, academic institution from July 2018 to January 2022. All surgeries were performed by one of four surgeons. Patients were enrolled in the study at least 2 weeks before their surgery date. Informed consent was obtained over the phone. After enrollment, patients were provided with a Fitbit Flex 2 (Fitbit) wrist-worn accelerometer and instructed on how to use the device. Patients were asked to wear the accelerometer every day while awake for 6 months following surgery and charge the device while sleeping when the battery was low. They were also enrolled in a secure web-based survey collection application, Research Electronic Data Capture (REDCap), for collection of PROs.

Exclusion criteria included if patients had comorbidities that could limit mobility, had a history of spine surgery in the 5 years before recruitment, lacked fluency in English, or did not own a smartphone device with Bluetooth capabilities. Patients could withdraw themselves from the study at any point, and patients were withdrawn if they did not comply with the study protocol. Additional inclusion criteria included if patients completed a minimum of three of four of the PRO questionnaires. Compliance with daily accelerometer use has historically been approximately 90% in this cohort.[7]

Outcome Measures

PROs (SF-12 and ODI) were collected at 2 weeks, 1 month, 3 months, and 6 months after surgery. SF-12 was not collected at the 2-week postoperative time point because the survey requires a recall period of 4 weeks; the ODI questionnaire does not specify a time component but is assumed to reflect function in the preceding few weeks. Specific measures of interest were the SF-12 Physical Component Scores (PCS) and ODI scores at each time point. Subjects used a Fitbit Flex 2 activity tracker to continuously track daily step counts in the postoperative period through the 6 months postoperative time point. Descriptive measures (median and maximum) of daily step counts were calculated using the measurement windows of 1 day, 3 day, 1 week, and 2 weeks preceding each PRO measurement. Median step counts were used to represent the patient's average experience over the measurement window, whereas maximum step counts represented positive outlier days within the measurement window (Supplementary Figure 1, http://links.lww.com/JG9/A401). Days in which patients did not use the activity tracker were excluded from median calculations.

Statistical Analysis/Recall Bias Analysis

Spearman rank correlation coefficients were used to evaluate correlations between patients' SF-12 PCS and ODI scores at each PRO measurement time point and their associated median or maximum steps for each measurement window. Patients with missing data at a given time point were excluded from any analysis requiring said data. Correlations with P values of <0.05 were considered statistically significant. Measurement windows were compared by assessing the proportion of notable correlation coefficients ($P < 0.05$) across all PRO measurement time points; PRO scores were considered to be more representative of measurement windows that had a higher number of notable correlations. To detect a moderate correlation ($r = 0.55$ to 0.60) between the aforementioned variables, with an alpha of 0.05 and power of 80%, a sample size of between 19 and 24 would be needed.[8] Recall bias was deemed present if daily step counts in the 1- and/or 3-day measurement windows markedly correlated with the PRO score, whereas step counts in the 1- and/or 2-week measurement windows were not. Conversely, recall bias was deemed absent if daily step counts in the 1- and/or 2-week measurement windows markedly correlated with the PRO score, whereas step counts in the 1- and/or 3-day measurement windows were not.

Results

Study Population

This study enrolled 22 patients in total. We included a maximum of 21 patients in the 2-week postoperative period, 22 patients in the 1-month postoperative period, 20 patients in the 3-month postoperative period, and 18 patients in the 6-month postoperative period. The study population's physical activity levels and SF-12 PCS scores generally increased over the study period, and ODI scores generally decreased over the study period (Supplementary Figures 1, 2A, and 2B, http://links.lww.com/JG9/A402).

Spearman Rank-Order Correlation Analysis for Recall Bias

Median daily step counts reflecting the 1- and 2-week measurement windows preceding PRO measurement most consistently correlated with SF-12 PCS measurements across all three postoperative time points (for 1-week measurement window: P = 0.001 at 1 month postoperatively, P = 0.001 at 3 months, and P = 0.019 at 6 months; for 2-week measurement window: $P < 0.001$ at 1 month postoperatively, $P < 0.001$ at 3 months, and P = 0.014 at 6 months). By contrast, median step counts reflecting the 1- and 3-day measurement windows preceding postoperative PRO measurement less consistently correlated with SF-12 PCS scores because no correlation was observed at the 6-month postoperative time point for both windows (for 1 day: P = 0.003 at 1 month postoperatively and P = 0.011 at 3 months, for 3 day: P = 0.003 at 1 month postoperatively and P = 0.003 at 3 months; Table 1). Similar, yet somewhat weaker, patterns were observed when looking at maximum step counts.

Median step counts measured during the 3-day, 1-, and 2-week measurement windows before PRO completion exhibited notable correlations with ODI score at the 1- and 3-month postoperative time points (for 3-day measurement window: P = 0.007 at 1 month postoperatively, P = 0.047 at 3 months postoperatively; for 1-week window: P = 0.014 at 1 month postoperatively and P = 0.013 at 3 months postoperatively; for 2-week window: P = 0.012 at 1 month postoperatively and P = 0.020 at 3 months postoperatively). Median step counts for the 1-day measurement window preceding postoperative PRO measurement did not correlate with ODI score at any of the postoperative time points (Table 2). Similar patterns were seen for maximum steps.

Discussion

The inherent subjectivity of PRO measures makes them susceptible to be influenced by recent activities and events as a source of potential recall bias. To our knowledge, this is the first study to correlate prospectively collected SF-12 PCS and ODI measurements following lumbar laminectomy with objective step count data in varying measurement windows preceding PRO collection (1 day through 2 weeks) to determine the effect of prospective recall bias, which occurs when the events of the days immediately preceding PRO measurement affect the PRO score. Our study found that median and maximum step counts in the 1-week window before PRO measurement most consistently correlated with both SF-12 PCS and ODI scores over the postoperative recovery period when compared with other measurement windows; PROs also highly correlated with step counts in the 2-week window. PROs did not highly correlate with step counts in the measurement windows (1- and 3-day) immediately preceding PRO measurement, which was somewhat surprising, and suggests that PROs are minimally affected by recency bias of the events of the days just before measurement. Further supporting this point, maximal measures of physical activity did not better correlate with PRO scores than median measures, suggesting that patients' reporting of PROs were not markedly biased by positive outlier days immediately

Table 1. **Correlations Between 12-item Short Form Health Survey Physical Component Scores (PCS) and Median and Maximum Steps per Day Over Various Measurement Windows at 2 Weeks, 1, 3, and 6 Months After Lumbar Laminectomy Surgery**

Measurement Window	2 Weeks			1 Week			3 Days			1 Day		
Postoperative Time Point	1 mo (n = 21)	3 mo (n = 20)	6 mo (n = 18)	1 mo (n = 21)	3 mo (n = 20)	6 mo (n = 18)	1 mo (n = 19)	3 mo (n = 20)	6 mo (n = 15)	1 mo (n = 19)	3 mo (n = 16)	6 mo (n = 13)
Median steps vs. SF-12 PCS												
Rho	0.714	0.714	0.567	0.669	0.693	0.548	0.649	0.632	0.329	0.647	0.618	0.236
P	**<0.001**	**<0.001**	**0.014**	**0.001**	**0.001**	**0.019**	**0.003**	**0.003**	0.232	**0.003**	**0.011**	0.437
Maximum steps vs. SF-12 PCS												
Rho	0.690	0.660	0.470	0.684	0.674	0.575	0.768	0.630	0.304	0.647	0.618	0.236
P	**0.001**	**0.002**	**0.049**	**0.001**	**0.001**	**0.013**	**<0.001**	**0.003**	0.271	**0.003**	**0.011**	0.437

mo = month; SF-12 PCS = 12-item Short Form Health Survey Physical Component Scores
Bolded values indicate a statistically significant correlation between variables using Spearman rank correlation coefficient test at a 95% confidence level.

Table 2. Correlations Between Oswestry Disability Index Scores and Median and Maximum Steps per Day Over Various Measurement Windows at 2 Weeks, 1, 3, and 6 Months After Lumbar Laminectomy Surgery

Measurement Window	2 Weeks				1 Week				3 Days				1 Day			
Postoperative Time Point	2 wks (n = 21)	1 mo (n = 22)	3 mo (n = 20)	6 mo (n = 16)	2 wks (n = 20)	1 mo (n = 22)	3 mo (n = 20)	6 mo (n = 16)	2 wks (n = 20)	1 mo (n = 20)	3 mo (n = 20)	6 mo (n = 13)	2 wks (n = 17)	1 mo (n = 20)	3 mo (n = 16)	6 mo (n = 11)
Median steps vs. ODI scores																
Rho	−0.217	−0.525	−0.514	−0.342	−0.399	−0.515	−0.544	−0.197	−0.331	−0.587	−0.450	0.179	−0.171	−0.348	−0.432	0.319
P	0.346	**0.012**	**0.020**	0.195	0.081	**0.014**	**0.013**	0.464	0.155	**0.007**	**0.047**	0.559	0.513	0.133	0.095	0.339
Maximum steps vs. ODI scores																
Rho	0.055	−0.648	−0.453	−0.323	−0.175	−0.593	−0.569	−0.374	−0.332	−0.571	−0.374	−0.201	−0.171	−0.348	−0.432	0.319
P	0.814	**0.001**	**0.045**	0.223	0.459	**0.004**	**0.009**	0.153	0.153	**0.009**	0.104	0.511	0.513	0.133	0.095	0.339

mo = week month; ODI = Oswestry Disability Index
Bolded values indicate a statistically significant correlation between variables using Spearman rank correlation coefficient test at a 95% confidence level.

before PRO measurement when completing the surveys. Overall, our results suggest that patients reporting PROs tend to recall their average experience over a period rather than being influenced by outlier events, thus refuting our original hypothesis that PROs would be affected by recall bias. We provide objective, accelerometer-backed evidence that prospective PROs accurately represent patient function and are not affected by recall bias when filled out postoperatively. This suggests that surgeons can administer PROs every 2 weeks without recall bias affecting the results. However, as this cadence may induce patient response fatigue, future studies should investigate recall bias over longer periods such as at 1- or 2-month intervals, with the goal of reducing the number of discrete time points at which patients are asked to fill out PROs.

Past studies in the literature using the retrospective model of recall bias assessment have found notable recall bias among lumbar spine surgery patients.[5,6,9,10] This has been specifically confirmed for the ODI; patients undergoing lumbar decompression procedures, on average, reported 9.6% higher ($P < 0.05$) preoperative ODI disability scores when this was queried postoperatively.[6] Rodrigues et al[9] also found that lumbar spine surgery patients overestimate the effectiveness of surgery, recalling markedly worse preoperative functional status during the postoperative period than the preoperative functional status previously collected. Evidently, recall bias plays a known role in the retrospective collection of PROs.[2,5,6,9,10] Although this has not yet been elucidated in the prospective collection of PROs, it is possible that recall bias may also be affected by recency bias. We defined the presence of recency bias as the PRO being influenced by events of the days before filling out the PRO, specifically if the 1- and 3-day measurement windows for step count correlated more with the PRO than the 1- and 2-week measurement windows. This is similar, but distinct, from recall bias, which is when patients inaccurately recall their health status on their PROs, which we elucidated from the longer 1- and 2-week measurement windows' correlation with PROs. The utilization of these different measurement windows allows us to test the effect of recency bias on patients' overall recall.

Although PROs reference a patient's overall feeling, this study provides information on whether patients truly report their overall feeling or allow their PROs to be biased by singular "one-time" events just before PRO completion. Findings from this study suggest that lumbar spine surgery patients are not influenced by recall bias. We found that ODI and SF-12 scores appear to reflect limits in patient activity and their disability in the preceding 1 to 2 weeks to collection very well and were less representative of the events of the days fresher in patients' memories. The difference in the effect of recall bias in the prospective and retrospective models of PRO collection outlined points to the advantages in prospective models in limiting the effect of bias on estimates on patients' physical function.

Understanding the period that PRO scores represent is important for their accurate interpretation by providers. In a systematic review analyzing recall periods for PRO collection in clinical trials, Stull et al[3] found that the optimal recall period for a given PRO depends on the phenomenon in patient behavior that is being studied. For example, shorter periods of recall are more appropriate when rapid fluctuations in behavior are expected, and longer periods of recall are appropriate for observing more stable phenomena. Although a shorter recall period reduces the potential for error, there is a potential for a loss in information if changes outside of the recall window are not captured.[3] We found that patients' ODI and SF-12 scores were most reflective of the 1 to 2 weeks before survey completion, which was relatively shorter than recall periods studied in prior literature.[6]

Regarding PROs, it is important to note that each PRO is unique in its susceptibility to bias. For instance, Rodrigues et al[9] found that in a cohort of 46 patients who underwent cervical spine surgery, retrospective PRO collection 1 year after surgery regarding preoperative status was subject to recall bias in measures of disability, such as Neck Disability Index, but did not affect quality of life, preoperative status, and pain measures. Conversely, our study noted similar trends in recall bias for both ODI and SF-12 PCS, perhaps because they query similar domains of the patient experience (disability and physical activity). Multiple studies have examined the effect of both prospective and retrospective recall bias on more general quality of life PRO measures, such as the SF-12 and Short Form-36 Health Survey (SF-36). Howell et al[11] found that recall bias does not affect the SF-12 up to 3 months after total hip arthroplasty. This is consistent with our study, which found that recall bias does not affect the SF-12 up to 6 months after lumbar spine surgery. Conversely, Lingard et al[12] found that recall bias does affect the SF-36 score after total knee arthroplasty; only moderate agreement was found between recalled postoperative assessments of preoperative pain and functional status and what patients had reported preoperatively. In examining hip arthroscopy patients, researchers compared prospectively collected PRO scores to retrospectively collected PRO scores and found that retrospectively collected PROs for hip arthroscopy were subject to recall bias.[13] They found that retrospectively collected PROs reflected worse pain and function than their prospectively recorded counterparts, which again highlights the importance of prospectively collecting PRO data to limit the effect of recall bias. Thus, this study, in the context of the literature on recall bias following orthopaedic surgery, suggests that there is little influence of recall bias following surgery if PROs are collected prospectively.

Although we have demonstrated that postoperative steps in the week after PRO measurement correlated with PRO scores in the postoperative period, this is premised on the assumption that as the measurement window for physical activity measurement is changed, factors aside from recall bias do not affect the association between steps and PRO score. However, this assumption ignores factors extrinsic to physical disability and pain with physical activity during the recovery trajectory that may affect the association between steps and PRO score. More specifically, these factors may play a more influential role on scoring in the immediate postoperative period and at the end of the recovery trajectory. For example, immediately after surgery, some patients may be more cautious in performing physical activity regardless of experienced pain, which could decrease step counts and explain an overall lack of correlation between step counts and PRO scores for any measurement window at the 2-week postoperative time point. Similarly, the lack of correlation between step counts and ODI scores at 6 months after surgery could be explained by lifestyle adjustment fatigue. More specifically, certain patients could be more eager to return to normal life and increase step counts despite discomfort, which could also confound correlations between ODI score and steps. However, ODI assessments include questions not purely related to physical activity, such as questions about sleep, personal care, and social life; these questions could affect ODI scores and resulting correlations with step counts as well. Thus, it is important to keep in mind that the presence or lack thereof of recall bias in measuring the patient experience is not the only factor affecting how patients' physical activity profiles compare to their PRO scores. However, as these extrinsic factors may be extremely variable between individual patients and in the context of our finding that ODI score and steps correlated better at the 1- and 2-week measurement windows than the 1- and 3-day measurement windows, this study still provides important objective information on the nonimpact of recall bias when patients fill out PROs prospectively. Future studies comparing PROs or evaluating the differences in PROs across different time frames prospectively and at predetermined follow-up intervals may provide further evidence on the effect of recall bias and would be a beneficial addition to the literature.

The primary limitation of this study is the size of the cohort. Unfortunately, although building these data registry, we had to suspend study enrollment because of the Covid-19 pandemic. We also identified that the restrictions associated with lockdowns during the pandemic may have had an effect on activity levels after surgery.[7] Thus, we consider that this cohort must stand alone with its current numbers. Notwithstanding this, the study represents a sound pilot. It is a prospective, within-subject cohort study that analyzes continuous data over an extended period. It included strict inclusion and exclusion criteria, which ensured that participants would not be markedly affected by confounding factors (ie, different procedures, other health conditions).

Ultimately, the sample size was sufficient for us to detect notable correlations based upon our power analysis, allowing us to draw conclusions regarding recall bias in the general patient population. Nevertheless, a future study with larger sample size will obviously yield greater power and the ability to further stratify the study population. Some strategies that may reduce attrition bias for future studies may include (1) providing consistent patient reminders to wear the device regularly at postoperative appointments, (2) incentives for wearing the device above a certain percentage of the study duration, and (3) implementation of real-time monitoring systems that send automatic alerts if noncompliance is detected.

Further limitations in creating these data registry, including selection bias, compliance with device usage, and seasonal differences in activity levels, are well documented in our initial study.[7] Also, although the correlation of mean and median step counts to PROs in the spine literature is well documented,[7,14] we are the first study, to our knowledge, using maximum step counts to test recall bias. However, as maximum and median step counts showed consistent results in our study, this gives confidence that maximum step counts can be used in this context. Finally, it is important to recognize the difference between recall that is biased and recall that is merely inaccurate. Recall that is biased is more likely to have a systematic error in a single direction, which is consistent with our finding that median and maximum step counts from the 1- and 2-week windows more consistently correlated with SF-12 PCS scores than 1- and 3-day windows over the postoperative period. By contrast, recall that is inaccurate would lead to random error around an average, which we did not find in our results. Patient factors, such as variations in patients' normal weekly activity level, lifestyle changes, or their willingness to walk consistently, are more likely to lead to recall that is biased, especially later on in the recovery period when physical ability should be improved from immediately after surgery. Other patient factors, such as patient memory and mood at the time of PRO completion, are more likely to lead to recall that is merely inaccurate because these factors have a higher level of unpredictability. However, this study is limited in that factors such as willingness to walk, lifestyle changes, memory, and mood were not available through the PROs or the objective physical activity metrics that were collected.

Conclusion

The use of accelerometry data in conjunction with PROs collected over a 6-month recovery period for lumbar laminectomy patients demonstrated higher concordance between subjective PROs and objective step count measurements over 1- and 2-week measurement windows compared with 1- and 3-day measurement windows before PRO completion, indicating little evidence for recall bias. Less consistent correlations with the 1- and 3-day measurement windows suggest that SF-12 PCS and ODI do not seem to be skewed by outlier days just before measurement. The PROs correlated most strongly with the 1- and 2-week step measurement windows; thus, administering PROs at 2-week intervals may provide for the most accurate measurement. However, to reduce patient response fatigue and increase compliance, future studies should investigate recall bias over longer periods to minimize the frequency at which PROs need to be filled out while still providing an accurate representation of a patient's overall feeling.

References

1. McCormick JD, Werner BC, Shimer AL: Patient-reported outcome measures in spine surgery. *J Am Acad Orthop Surg* 2013;21:99-107.

2. Gotlin MJ, Kingery MT, Baron SL, McCafferty J, Jazrawi LM, Meislin RJ: Recall bias in retrospective assessment of preoperative patient-reported American shoulder and elbow surgeons scores in arthroscopic rotator cuff repair surgery. *Am J Sports Med* 2020;48:1471-1475.

3. Stull DE, Leidy NK, Parasuraman B, Chassany O: Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. *Curr Med Res Opin* 2009;25:929-942.

4. Chang EM, Gillespie EF, Shaverdian N: Truthfulness in patient-reported outcomes: Factors affecting patients' responses and impact on data quality. *Patient Relat Outcome Meas* 2019;10:171-186.

5. Aleem IS, Currier BL, Yaszemski MJ, et al: Do cervical spine surgery patients recall their preoperative status? A cohort study of recall bias in patient-reported outcomes. *Clin Spine Surg* 2018;31:E481-e487.

6. Aleem IS, Duncan J, Ahmed AM, et al: Do lumbar decompression and fusion patients recall their preoperative status? A cohort study of recall bias in patient-reported outcomes. *Spine (Phila Pa 1976)* 2017;42:128-134.

7. Bienstock DM, Shankar D, Kim J, et al: Accelerometry data delineate phases of recovery and supplement patient-reported outcome measures following lumbar laminectomy. *World Neurosurg* 2022;160: e608-e615.

8. Hulley SBCS, Browner WS, Grady D, Newman TB. *Designing Clinical Research: An Epidemiologic Approach*. ed 4. Philadelphia, PA: Lippincott Williams & Wilkins (LWW); 2013.

9. Rodrigues R, Silva PS, Cunha M, Vaz R, Pereira P: Can we assess the success of surgery for degenerative spinal diseases using patients' recall of their preoperative status? *World Neurosurg* 2018;115:e768-e773.

10. Zini MLL, Banfi G: A narrative literature review of bias in collecting patient reported outcomes measures (PROMs). *Int J Environ Res Public Health* 2021;18:12445.

11. Howell J, Xu M, Duncan CP, Masri BA, Garbuz DS: A comparison between patient recall and concurrent measurement of preoperative quality of life outcome in total hip arthroplasty. *J Arthroplasty* 2008;23: 843-849.

12. Lingard EA, Wright EA, Sledge CB; Kinemax Outcomes Group: Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. *J Bone Joint Surg Am* 2001;83: 1149-1156.

13. Crutchfield CR, Givens RR, O'Connor M, deMeireles AJ, Lynch TS: Recall bias in the retrospective collection of common patient-reported outcome scores in hip arthroscopy. *Am J Sports Med* 2022;50: 3190-3197.

14. Haddas R, Lawlor M, Moghadam E, Fields A, Wood A: Spine patient care with wearable medical technology: State-of-the-art, opportunities, and challenges: A systematic review. *Spine J* 2023;23:929-944.