



Recalling what we thought we knew about recall periods: a qualitative descriptive study of how adults diagnosed with cancer use recall periods for patient-reported outcome items about physical function

Theresa Coles¹ · Kate Plyler¹ · Alexy Hernandez¹ · Rebecca Fillipo¹ · Debra M. Henke¹ · Cara Arizmendi^{1,2} · C. Rory Goodwin³ · Thomas W. LeBlanc⁴ · Sandhya Lagoo-Deenadayalan⁵ · Bryce B. Reeve¹ · Kevin P. Weinfurt¹

Accepted: 13 November 2024 / Published online: 4 December 2024
© The Author(s) 2024

Abstract

Objectives In cancer studies, assessment of patients' physical function can provide insight into cancer-related symptoms and the side effects of treatment. Physical function can be assessed using patient-reported outcome measures (PROMs), which may or may not include a recall period—the amount of time the questionnaire asks the respondent to think back to answer the questions. More understanding is needed about how patients interpret and respond to items posed with different recall periods, and which recall period they actually use.

Study Design and Setting We conducted a qualitative study with 72 adults diagnosed with cancer to describe response processes when answering PROM items about physical function.

Results When describing their physical function during concept elicitation interviews, most participants recalled their functioning over 1 month or more. When presented with PROM items with no recall period, the most used period was more than 4 weeks. When presented with a 7-day recall, the most used period was 7 days. However, almost 30% of responses used recall periods greater than 1 week.

Conclusion Including a 7-day recall period improves recall consistency for patients when answering physical function PROM items, but there is still room for improvement. The inaccuracy of recall for PROMs in clinical trials influences the ability to calculate the change in treatment outcomes over time and may conceal or emphasize actual treatment effects.

Keywords Patient-reported outcome measures · Physical function · Cancer · Questionnaire · Qualitative research · Recall period

✉ Theresa Coles
Theresa.Coles@Duke.edu

¹ Center for Health Measurement, Department of Population Health Sciences, Duke University School of Medicine, 215 Morris St, Durham, NC 27701, USA

² AstraZeneca, Evinova, Durham, NC, USA

³ Department of Neurosurgery, Duke University School of Medicine, Durham, NC, USA

⁴ Division of Hematologic Malignancies and Cellular Therapy, Department of Medicine, Duke University School of Medicine, Durham, NC, USA

⁵ Department of Surgery, Duke University School of Medicine, Durham, NC, USA

Introduction

Physical function (PF) (see Box 1) is a recommended outcome used to evaluate treatment benefit and burden of symptoms and side effects in cancer studies [1].

Box 1 Definition of physical function [2].

Algorithm 1 Definition of physical function

Physical function is the level with which persons engage in activities that involve physical effort. This could range from fundamental actions (e.g., walking, bending, reaching, climbing stairs, changing or maintaining positions) to activities of daily living involving more than one fundamental action (e.g., bathing, eating, dressing, toileting), and to more complex activities such as participation in activities (e.g., hobbies, work, caregiving, volunteering, gardening) depending on the patient population or treatment being studied. The nature of engagement in activities could be described in multiple ways, including patients' ability or inability to conduct an activity, satisfaction with conducting an activity, level of completion of an activity, quality of movements needed to complete an activity, and others.

While PF can be evaluated using different types of outcome assessments, patient-reported outcome measures (PROMs) allow for measurement based on direct report by the patient. PROMs often use a recall period, or the amount of time the respondent is asked to think back in order to answer the questions (e.g., “last 24 h,” “past 7 days”), while other PROM items have no recall period specified. Depending on the context of use for the PROMs and endpoint specification, the choice of recall period should consider a number of factors, such as the recency of the outcome of interest, rate of change in the phenomenon, complexity of the information being retrieved, and how usual or unusual the phenomenon is (e.g., walking to a bus stop, running a marathon) [3]. Recall periods that do not align with the context of use can introduce measurement error that influences the ability to detect a treatment effect.

Previous research has identified at least 13 PF-specific PROMs that could be used in cancer populations [4]. Three widely used PROMs in cancer clinical trials are the Medical Outcomes Study Short Form 36 (SF-36) [5], European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) [6] and the Patient-Reported Outcomes Measurement Information System® (PROMIS®) PF items [7, 8]. The SF-36 PF asks patients about *current* physical health status; for example, how their “health now limits you.” The QLQ-C30 does not specify a recall period, asking patients, “Do you have any trouble...” The PROMIS® PF Item Bank version 2.0 [8] does not include a specific recall period, and ask patients

“Are you able ...” or “Does your health now limit you ...” However, an alternative PROMIS PF short form specifically for adults with cancer was recently developed with a 7-day recall period [9]. Other PF measures, such as the World Health Organization Disability Assessment Schedule 2.0 [10] and the Stanford Health Assessment Questionnaire-Disability Index (HAQ-DI) [11] have 30-day and 7-day recall

periods, respectively.

Patients receiving anti-cancer treatment may have rapid declines (e.g., stem cell transplant), slow declines (e.g., sarcopenia, cumulative toxicity), increases (e.g., steroids), or cyclical changes (e.g., chemotherapy) in their PF. Changes in PF status within the recall window may influence patients' certainty about their responses. Accurately capturing patients' PF includes ensuring that the recall period is long enough to capture potential shifts but not too long that patients have difficulty remembering PF activities.

Some quantitative studies have explored recall periods for PF PROMs. One study tested PROMIS PF items with no specified recall period, a 24-h recall period, and a 7-day recall period [12]. The authors concluded that compared to no specified recall period, the use of a recall period has little or no effect on PROMIS PF responses or scores. In another study, researchers compared a 7-day recall with no specified recall on the PROMIS PF Short Form 8c (PF8c) [9] and concluded that no specified recall period was preferred because there was not a significant effect on PF8c responses based on the recall presented. Yet, we do not know *which* recall period patients *actually* use when describing their PF, irrespective of whether a recall period is specified. If, as the previous studies suggest, there is no interpretational difference between the results of a specific recall (e.g., 7-day) or no specified recall within a PROM, we do not know *why* patients may have used a different recall period than what was presented in the PROM. Therefore, knowing which

recall period patients are confident in reporting their PF is essential information to minimizing measurement error.

As part of the Patient Reports of Physical Functioning Study (PROPS) research program, we conducted qualitative interviews with adults diagnosed with cancer to understand their response processes when answering PROM items about PF. We had four sub-goals: (1) describe the timeframe adults diagnosed with cancer naturally used to report when asked about their current PF; (2) summarize how adults describe changes in PF over time; (3) examine the time window adults consider when answering questions posed with different recall periods and what recall period they actually use; and (4) describe the retrieval strategies adults employ when answering questions posed with different recall periods.

Methods

Study design

Within an interpretive/constructivist research paradigm, through which we aim to understand participants' recall period choices through their reported experience [13–15], we conducted one-on-one qualitative interviews with adults at least 22 years of age with clinician-diagnosed cancer who received an anticancer drug or biologic within 6 months of the interview, could read, speak, and understand English, and who received care at Duke University Hospital (Durham, NC). Sample size justification, exclusion criteria, and recruitment/enrollment procedures are described elsewhere [2].

We used purposive sampling [16] to ensure a diverse sample of adults with cancer in relation to their PF. We tracked and balanced patient characteristics for purposive sampling including age (22–60 vs > 60 years of age), race, ethnicity, education, cancer type, gender, and PF status per a patient-reported Eastern Cooperative Oncology Group (ECOG) performance status rating. Guest and colleagues found that little new information was learned after 12 interviews in a homogeneous sample [17]. The sample size was derived by prioritizing four relatively homogeneous groups. These groups were cancer status: (1) metastatic or (2) nonmetastatic, and location at the time of the interview: (3) inpatient or (4) outpatient. The sample within each group was intentionally diverse including a range of cancer types and ECOG statuses. Therefore we planned to recruit 18 participants to account for heterogeneity within each group. The final sample was determined to be 72 participants (18 * 4). Patients were screened for sample characteristics via medical chart review and a patient-reported screening survey for eligibility (Appendix 1).

1. Are you **able** to [enter activity]?

 - Completely able
 - Mostly able
 - Moderately able
 - Barely able
 - Unable to do

Fig. 1 Example cognitive interview item

Description of interview content

The study team developed a semi-structured interview guide (Appendix 2). Interviews included concept elicitation [14] and cognitive interviewing techniques [18] to learn how participants understand and respond to PF PROM items framed using different recall formats and were expected to last one hour. Interviews began with a brief study overview and concept elicitation questions asking participants to describe their PF at the time of the interview and how far back they naturally thought about their PF. Interviewers also explored changes in PF over time as described by participants within the week prior to the interview. Cognitive interviews focused on participants' response processes related to recall periods and retrieval [19]. Participants were presented with nine PROM items, separated into two sections, developed for the cognitive interviews, representing different facets of PF [2, 4] (Appendix 3). The first section contained four items that were presented with no recall period, and assessed ability, limitation, satisfaction, and difficulty. The second section contained five items that were presented with a 7-day recall period. Four of the items in this section assessed the same facets of PF (ability, limitation, satisfaction, and difficulty) as those without a recall period. This section also included an additional item, "In the past 7 days, did you [enter activity]" to assess whether the participant had completed the activity. Unlike the other four questions, this question was included only in the section containing questions with recall periods as it would not have made sense without a recall period. For consistency, each item presented the same physical activity. To ensure that the physical activity was relevant for each participant, all participants were asked four standard questions to help decide on an activity for the cognitive interviewing section. The interviewer and participant chose an activity that the participant could do but may have some trouble with and inserted those activities into the item stem. The four standard activities and corresponding forms offered were:

- "sit on the edge of the bed?" (Form D).
- "walk for 5 yards?" (Form C).
- "walk for 15 min?" (Form B).
- "walk briskly for more than a mile?" (Form A).

Table 1 Forms used by participants

Form	N	%
A-walk briskly for more than a mile	13	18.31
B-walk for 15 min	15	21.13
C-walk for 5 yards	5	7.04
D-sit on the edge of the bed	0	0.00
X-other	38	53.52

Participants and/or interviewers attempted to choose the activity that was manageable yet slightly challenging for the participant. Participants used a nonstandard activity when the above tasks were either too easy or could identify a more relevant task. To choose the activity for Form X, the participant identified a typical activity or an activity they were trying to return to. Alternatively, if the participant had trouble identifying an activity, the interviewer chose one based on the participants description of typical activities. Figure 1 shows an example item with an agreed-upon activity inserted into the section, “[enter activity].”

We conducted cognitive debriefing using a concurrent verbal probing process [20]. Participants were asked how they chose a response, how far back they were thinking, when the last time was that they tried the activity in question (if at all), and how easy or difficult it was to answer the question (Appendices 2 and 3). Special attention was paid to participants’ response processes on retrieval. Interviews were offered by web conference or in person, giving participants their choice of mode. Interviews were conducted from February 2022 to November 2022, throughout the COVID-19 pandemic.

Analysis

PROPS data quality and analysis procedures were described in detail previously [2]. Qualitative content analysis was used to analyze participant narratives from transcripts [21]. Debriefing forms were used in lieu of transcripts in cases where participants did not consent to recording or if technical audio issues prevented transcription. Team members developed a codebook collaboratively and used NVivo

qualitative data analysis software (Version 12 for Windows) to apply codes. Theme coding was used as an inductive process, and pattern coding was used to aggregate response processes from the cognitive interviews [18].

After coding, analysts identified themes by noting when multiple participants reported similar experiences or interpretations of an item. Next, themes were compared descriptively, primarily by inpatient versus outpatient. Other characteristics (e.g., age, gender, cancer stage, treatment type) were associated with each transcript, providing context to salient themes; illustrative quotes were provided to support assertions.

The study was approved by the Duke University institutional review board (Pro00108611). Participants received remuneration of \$50.

Results

Participant characteristics

Interviewers attempted to contact 392 patients, of which 282 were reached. Of these 282 patients, 72 agreed to participate, while 161 declined, 35 were lost to follow up, and 14 were not eligible. We conducted interviews with 72 adults with cancer. One participant was not able to complete the entire interview, resulting in a sample size of 71 for the cognitive interview portion. A sample description is provided elsewhere [2] (Appendix 4) and briefly described here. Half the sample was female ($n = 36$). The study sample included 12 inpatients and 60 outpatients. It was split evenly between participants with nonmetastatic ($n = 34$, 47.2%) and advanced/metastatic ($n = 38$, 52.8%) disease. Participants’ primary cancer types were diverse, including 22 different cancer types [2] (Appendix 5). Almost half of participants ($n = 32$, 46.5%) chose one of the standard activities (A through D). No participants used “sit on the edge of the bed.” Several patients who could walk a mile but who chose to use “the ability to walk 15 min” were primarily concerned with the word “briskly.” Chosen activities for Form X included walking for a distance not included in Form A-C ($n = 7$), walking briskly

Table 2 Illustrative quotes describing natural recall periods used

Reason for choosing natural recall period	Illustrative quote	Participant description
Since diagnosis	<i>“Just since I received cancer, so just since February [xx], 2021. I’m not to where I was before.”</i>	Black, female, non-Hispanic, < age 60, outpatient, metastatic
Since starting treatment	<i>“I was just thinking about the last six months while I was on Xeloda.”</i>	Other race, female, non-Hispanic, < age 60, outpatient, nonmetastatic
Since surgery	<i>“That’s [physical function description] only been since Tuesday [2 days] since I had the operation.”</i>	White, female, non-Hispanic, < age 60, inpatient, metastatic

Table 3 Illustrative quotes describing change in PF over the past week

Frequency of PF change	Illustrative quote	Participant description
Day-by-day or multiple times during 1 week	<i>“It’s very up and down for the most part. One day I’ll feel fine and the next day I’ll be exhausted.”</i>	White, female, Hispanic, < age 60, outpatient, non-metastatic
Weekly with treatment cycles	<i>“But if we were on a regular chemo cycle, then it just comes and goes with it every two weeks.”</i>	White, male, non-Hispanic, < age 60, outpatient, metastatic
Unpredictable change	<i>“My physical function can change within a heartbeat. That’s the stressful thing about it.”</i>	White, male, non-Hispanic, > age 60, outpatient, nonmetastatic

Table 4 Recall period used for items with no recall period

Recall period used	Responses	% of Responses
Today	84	29.7
< 1 week	30	10.6
1 week	31	11.0
1–4 weeks	20	7.1
> 4 weeks	118	41.7
Total	283	100.0

for distances both greater and less than 1 mile ($n = 7$), yardwork ($n = 4$), such as mowing the lawn or gardening, other forms of high-intensity exercise such as running, exercise videos, biking ($n = 8$), bed mobility ($n = 3$), going up steps ($n = 2$), walking without a walker, bending over to pick something up, backpacking, driving, getting dressed, lifting heavy objects, and organizing classroom. Table 1 presents the proportion of participants using each form.

Natural recall period

At the beginning of the interview, participants were asked, “How would you describe your physical functioning currently?” and then, “When you were thinking about your physical function, how far back in time were you thinking?” The purpose was to describe participants’ “natural recall period” when considering their current PF status. The majority of participants used a recall period of more than 1 month, with 27.8% (20/72) using a natural recall period of 1–6 months, and 44.4% (32/72) using a recall period of more than 6 months. Nineteen participants (26.4%) used a recall period of less than a month, with six (8.3%) using a recall period of 1–4 weeks, six (8.3%) using a recall period of less than 1 week, and seven (9.7%) using a recall period of today (at the time of the interview) to describe their current PF. (One participant did not specify a natural recall period.)

The three most common reasons for the natural recall periods chosen by participants were related to changes since diagnosis or getting sick (15/72, 20.8%), since

Table 5 Recall period used for items with 7-day recall period

Recall period used	Responses	% of Responses
Today	31	8.8
< 1 week	32	9.1
1 week	186	52.7
1–4 weeks	18	5.1
> 4 weeks	86	24.4
Total	353	100.0

starting treatment (16/72, 22.2%), or since most recent surgery (15/72, 20.8%) (Table 2).

The majority of inpatients used a recall period of less than 1 week (8/12, 66.7%). Yet the majority of outpatients used a recall period of more than 1 month, with 19 of 60 (31.7%) using a recall period of 1–6 months and 31 (51.7%) using a recall period of more than 6 months. There were no apparent differences in natural recall periods for participants with advanced/metastatic cancer and nonmetastatic cancer or by treatment type.

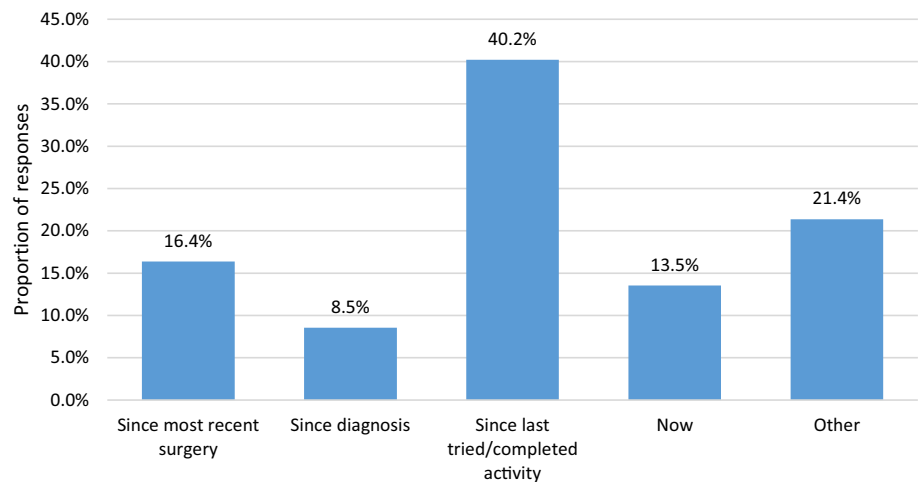
Physical function change over the past week

To examine the appropriateness of a 7-day recall period, participants were asked “Is your physical functioning now different than it was a week ago?” and were prompted to expand on their answer. As a follow up, participants were asked how often their PF changed within a week. Overall, they found it easy to think back 7 days and described changes in their functioning. Over half ($n = 37$, 51.4%) indicated no change in PF in the past week. Nine indicated that their PF fluctuated day-by-day or multiple times per week with treatment. Seven participants indicated changes that occurred week-by-week with cycles of treatment. For seven participants, pain was the primary indicator and driver of PF over the past week. In terms of frequency of change, most indicated that their changes in PF were unpredictable due to pain. Table 3 presents illustrative quotes on how participants experience changes in their PF.

During concept elicitation, participants did not have trouble recalling their PF or describing changes in their PF over time within the 7 days prior to the interview.

Table 6 Recall period used by treatment location

Recall period used by participants	No recall period in item wording				7-day recall period in item wording			
	Inpatient		Outpatient		Inpatient		Outpatient	
	Response	%	Response	%	Response	%	Response	%
Today	29	60	55	23.4	11	18.3	20	6.8
< 1 week	7	14	23	9.8	17	28.3	15	5.1
1 week	7	14.6	24	10.2	24	40.0	162	55.3
1–4 weeks	0	0.0	20	8.5	0	0.0	18	6.1
> 4 weeks	5	10	113	48.1	8	13.3	78	26.6
Total	48	100	235	100.0	60	100.0	293	100.0

Fig. 2 Common retrieval strategies for items with no recall period

Recall period

The unit of analysis for this section is the number of item responses per person, with up to 9 responses per person for 71 participants (639 responses total). When participants were offered items with no recall period (items 1–4 in Appendix 3), the most frequently used recall period was more than 4 weeks (41.7%). Table 4 shows the recall period used for items with no recall period.

This table presents the recall periods participants reported across four items with no recall period. The total number of responses possible is 288 (72 participants \times 4 items with no recall = 288). One participant did not complete the cognitive interview (missing all four items with no recall period) and one participant did not provide a recall period for one item.

When participants were offered a 7-day recall (items 5–9 in Appendix 3), the most frequently used recall period was 7 days (52.7%). Table 5 shows the recall period used for items with a 7-day recall period.

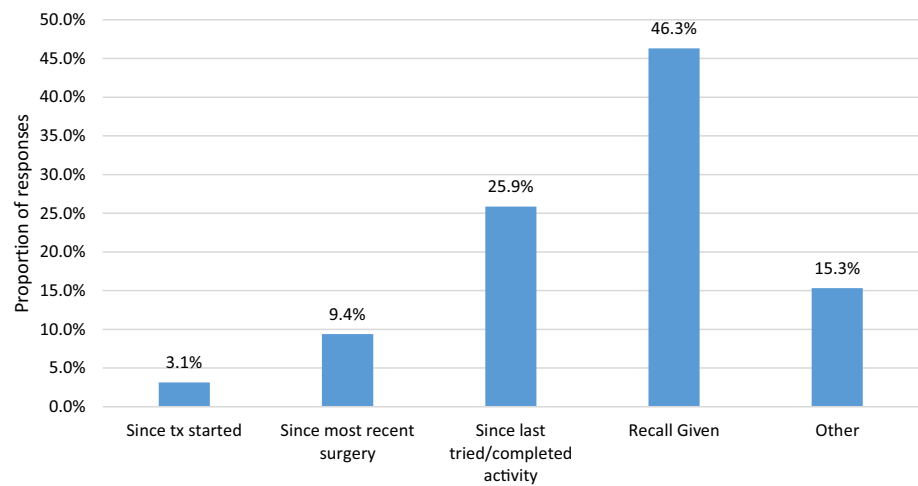
This table presents the recall periods participants reported using five items with 7-day recall. The total number of responses possible is 360 (72 participants \times 5 items with no recall = 360). One participant did not complete the cognitive

interview (missing all five items with no recall period) and two responses were missing.

Table 6 shows which recall periods were used by participants depending on treatment location. For inpatients, PF changes are precipitous due to surgery or treatment. Inpatients tended to use a recall period of 1 week or less. Outpatients used a recall period of more than 4 weeks more often than any other recall period length (48.1% of responses) when no recall period was offered. When a 7-day recall period was offered, outpatients most often used a 1-week recall period or less (67.2% of responses).

This table presents the recall periods participants reported using four items with no recall period versus five items with a 7-day recall period. The total number of possible inpatient responses for participants responding to items with no recall and 7-day recall are 48 and 60 respectively. The total number of possible outpatient responses for participants responding to items with no recall and 7-day recall period are 240 and 300 respectively. One outpatient did not complete the cognitive interview, accounting for most of the missing responses.

Fig. 3 Common retrieval strategies for items with 7-day recall period



Retrieval strategies

For items with no recall period, the most common retrieval process was thinking back to the last time an activity was tried or completed (113 responses; 40.2% responses). Figure 2 shows the most common retrieval strategies used for items with no recall period.

For inpatients, the most common retrieval strategies were PF since their most recent surgery (50%, $n = 24$), since they last tried or completed an activity (31.3%, $n = 15$), or now (i.e., at the time of the interview) (14.6%, $n = 7$). For outpatients, retrieval strategies focused on PF since they last tried or completed an activity (42.1%, $n = 98$), now (13.3%, $n = 31$), since their diagnosis (10.3%, $n = 24$), or since their most recent surgery (9.4%, $n = 22$). Figure 3 shows the most common retrieval strategies for items with recall periods; using the recall given was the strategy used most often (46.3%).

For inpatients, the most frequently used retrieval strategies for items with 7-day recall were thinking of their PF since their most recent surgery (45.0%, $n = 27$), the recall period offered (33.3%, $n = 20$), or since they last tried or completed an activity (15.0%, $n = 9$). For the 9 inpatients that used a strategy of when they last tried the activity, all had tried the activity within 7 days since the interview. Outpatients most often thought about the recall period offered (49.0%, $n = 143$) or since they last tried or completed an activity (28.1%, $n = 82$) when retrieving responses for items with a 7-day recall period. For the participants that used the retrieval strategy of when they last tried an activity, 18/82 responses were based on reports of having tried the activity outside of the 7-day recall window.

Discussion

This study explored patients' response processes when answering PROM items about PF with and without a specified recall period. When talking naturally about their PF status, most participants recalled their functioning over 1 month or longer. Findings were different between natural conversation and cognitive testing; when no recall period was offered for items tested in the cognitive interview, slightly more than half the time participants used a recall of 1 week or less. When asked open-ended questions about their PF over the past week, participants found it easy to describe changes (or lack of) in their PF over the prior week. Some participants indicated that their PF fluctuated due to treatment or pain. When answering PROM questions during the cognitive interview, they used a 1-month recall period less often than when describing PF naturally (during concept elicitation). When participants were not offered a recall period in PROM items, most responses focused on the last time they tried/completed the activity in question. For PROM items with a 7-day recall period, almost half of the retrieval strategies used the recall period offered, but the remainder focused on other issues such as the last time they completed an activity or treatment.

While a previous study showed that participants responded similarly for either no recall or 7-day recall [12], our study illuminates the variability in participant retrieval strategies and actual recall periods used for PROM items including no specified recall or 7-day recall. The primary implication of our findings is that offering a 7-day recall period helps to standardize how far back in time participants retrieve information (i.e., there is more consistency in the recall period used); yet even when prompted with a 7-day recall period, almost a third of responses relied on a longer recall period. One possible reason for using a longer recall period is that participants followed conventions of conversation by trying to guess what the interviewer truly wants to

know and then adjusting their response strategies to be as helpful as possible [22]. For example, the participant might know that the prior 7 days were unusual (e.g., took a multi-day road trip that involved mostly sitting) and so used a longer period to give a more accurate picture of their PF under usual circumstances. This explanation is also consistent with our findings concerning differences between inpatient and outpatient participants (i.e., inpatients tended to use shorter recall periods corresponding to their recent hospitalization) and the strategy they used in our initial open-ended queries about their physical PF (i.e., participants tended to report PF since a salient clinical event that might have led to a change in their PF).

Specifying a 7-day recall helps but does not ensure consistency among participants in their recall periods, which—as an anonymous reviewer pointed out—raises the question of how much consistency in the use of a recall period is sufficient for clinical research. At this time, there appears to be no standard or agreed-upon threshold for how consistent patients need to be in the use of recall periods. The issue might be similar to that of missing data, where the impact of missing data depends on the mechanism of missingness. Some deviations from the specified recall period could reflect random measurement error if respondents' deviations are due to unsystematic missingness, such as random lack of attention or discipline in responding. More of this type of inconsistency may result in attenuated associations with other variables and/or lower statistical power. Other, more systematic, inconsistencies in recall period used might reflect conscious decisions by the respondent to use a recall period that they believe is aligned with what the researcher wants to know; for example, choosing a shorter recall period that reflects their time as an inpatient on the assumption that the researcher wants to interpret the scores to assess how the patient is now. If the respondent's assumption about the researcher's intended interpretation of the scores is correct, then the deviation may not bias the researcher's conclusions from the data. If the respondent's assumption is wrong, it might contribute bias. More work is needed to understand the implications of both the extent and types of deviations from the specified recall period.

An assumption that investigators make with any recall period, is that study participants are using the recall period presented. Our qualitative study is unique in that we assessed if this assumption was true. Many quantitative studies have assessed the difference between PRO scores when patients are presented with multiple daily assessments versus a single assessment with a 7-day recall, for example [23]. In these cases, the shorter recall is widely considered to be the most accurate. The quantitative studies address the question: Are the responses to multiple assessments with a daily recall and a single assessment with n -day recall similar enough so that we can use the single assessment with longer recall? A

threshold could be chosen a priori to determine if there is an “acceptable” agreement between responses when different recall periods are used. However, the quantitative work published thus far does not provide a threshold for how *consistent* patients need to be in recall period use. This qualitative study, and previous quantitative studies, are asking different questions. Our qualitative study is focused on which recall period people report using, rather than how similar are the responses to the PROM items if we ask them with different recall periods. Our study opens the door for us to be curious about how often patients use the recall periods offered, and what risks to validity may be at play in different contexts of use given the potential uncertainty of recall periods being used. Multi-item PF PROMs ask participants about a range of activities. Many participants linked retrieval of their PF to the last time they tried an activity. Future research may be needed to explore whether patients use different recall periods for different activities, potentially influencing the recall period used by patients. Our study was limited in asking only about an activity that was relevant to the participants. Therefore, our data do not extend beyond items addressing activities they conducted less frequently. For activities that are not being completed often, these activities arguably should not be asked about anyway—qualitative studies assessing the content of the PROM items would help us avoid this issue.

Other limitations of this study include the inability to investigate phenomena that might influence recall, such as present-state bias and, potentially, response shift. The contiguous order in which PROM items were presented (first typical recall during concept elicitation, then no recall and finally a 7-day recall) during the cognitive interviews may have influenced which recall period participants used during the latter section of the interview. Another limitation is that participants were recruited from a single medical center. In addition, results are limited to sample-level conclusions, and individual-level patterns for recall strategy were not assessed. Cognitive interviews focused on one activity that was relevant to each participant. This was a strength in that it ensured relevance, but it was also a limitation because we were unable to examine recall periods for activities that participants did not do. This study also lacked representation of participants who were not English-speaking within and outside the United States.

A strength of this study is that we captured participants' actual recall period and their strategies for retrieving answers. Such data reveal how PF question format can be improved in the future to ensure more accurate answers. Another strength is that we included a substantial proportion of inpatients, which may be an infrequent but highly important period in patients' lives, particularly related to physical functioning. Results differed slightly between inpatients and outpatients; we chose to include inpatients because research in this group has been limited.

Our results show that presenting a 7-day recall period leads to more consistent retrieval strategies across adults diagnosed with cancer, in comparison with PROM items that do not use a recall period. Investigators should also consider their target population by setting (inpatient versus outpatient), disease activity, and effects of treatment when selecting the appropriate PROM recall period. Some researchers have tried strategies to improve retrieval, including using video recordings of participants at baseline [24]. Future research could test different strategies for improving the accuracy of recall.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11136-024-03847-1>.

Acknowledgements The authors would like to acknowledge the patients who participated in this study, who took the time to provide important insight into their experiences with physical function. The authors would also like to acknowledge and thank Dr. Vishal Bhatnagar and Erica Horodniceanu from the Oncology Center of Excellence in U.S. Food & Drug Administration for their essential contributions to the study. The authors would also like to thank Liz Wing for medical editing support.

Author contributions The corresponding author (TC) confirms that she had full access to all the data in the study and had final responsibility for the decision to submit for publication. Each author has contributed significantly to, and is willing to take public responsibility for, the following aspects of the study. TC: conceptualization, funding acquisition, methodology, interpretation of results, writing original draft; KP: data collection, analysis, interpretation of results, writing–review and editing; AH: data collection, analysis, interpretation of results, writing–review and editing; RF: data collection, analysis, interpretation of results, writing–review and editing; DMH: data collection, analysis, interpretation of results, writing–review and editing; CA: data collection, writing–review and editing; SLD: facilitated data collection, interpretation of results, writing–review and editing; CRG: facilitated data collection, interpretation of results, writing–review and editing; TWL: facilitated data collection, interpretation of results, writing–review and editing; BBR: methodology, interpretation of results, writing–review and editing; KPW: conceptualization, methodology, interpretation of results, writing–review and editing.

Funding This work was supported by the U.S. Food & Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) Oncology Center of Excellence (Contract 75F40120C00069), and as part of a financial assistance award funded by FDA/HHS. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS or the U.S. Government.

Data availability Summary reports are available upon request to the corresponding author.

Declarations

Conflict of interest TC has research funding from Merck and Pfizer; a prior consulting agreement with Regxbio; and received a speaking honorarium from ISPOR. CA was a fellow at Duke (2021–2023) funded by AstraZeneca and is now an employee at AstraZeneca (2023–present). CRG is a consultant for Stryker and Medtronic. BBR had a consulting role with Novartis. TWL has received honoraria for consulting/advisory boards from AbbVie, Agilix, Agios/Servier, Apel-

lis, Astellas, AstraZeneca, BlueNote, BMS/Celgene, Genentech, GSK, Lilly, Meter Health, Novartis, and Pfizer; speaking related honoraria from AbbVie, Agios, Astellas, BMS/Celgene, Incyte, and Rigel; equity interest in DoseNrx (stock options in a privately-held company); royalties from UpToDate; research funding from the AbbVie, American Cancer Society, AstraZeneca, BMS, Deverra Therapeutics, Duke University, GSK, Jazz Pharmaceuticals, and Seattle Genetics. TWL is a Scholar in Clinical Research of the Leukemia & Lymphoma Society. BBR has received honoraria for participating in a meeting with Novartis. All other authors have no disclosures.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

References

1. Kluetz, P. G., Slagle, A., Papadopoulos, E. J., Johnson, L. L., Donoghue, M., Kwitkowski, V. E., Chen, W. H., Sridhara, R., Farrell, A. T., Keegan, P., Kim, G., & Pazdur, R. (2016). Focusing on core patient-reported outcomes in cancer clinical trials: Symptomatic adverse events, physical function, and disease-related symptoms. *Clinical Cancer Research*, 22(7), 1553–1558. <https://doi.org/10.1158/1078-0432.Ccr-15-2035>
2. Coles, T., Plyler, K., Hernandez, A., Fillipo, R., Henke, D. M., Arizmendi, C., Lagoo-Deenadayalan, S., Goodwin, C. R., LeBlanc, T. W., Horodniceanu, E. G., Bhatnagar, V., Reeve, B. B., & Weinfurt, K. P. (2024). What facets of physical function are most important to adults diagnosed with cancer? *Quality of Life Research*, 33(7), 1829–1839. <https://doi.org/10.1007/s11136-024-03654-8>
3. Stull, D. E., Leidy, N. K., Parasuraman, B., & Chassany, O. (2009). Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. *Current Medical Research and Opinion*, 25(4), 929–942. <https://doi.org/10.1185/03007990902774765>
4. Coles, T., Fillipo, R., Plyler, K., Hernandez, A., Henke, D. M., Arizmendi, C., Cantrell, S., Lagoo-Deenadayalan, S., Rory Goodwin, C., LeBlanc, T. W., Horodniceanu, E. G., Bhatnagar, V., Reeve, B. B., & Weinfurt, K. (2024). Facets of physical function assessed by patient-reported outcome measures in oncology research. *Quality of Life Research*, 33(7), 1819–1828. <https://doi.org/10.1007/s11136-024-03640-0>
5. Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. conceptual framework and item selection. *Medical Care*, 30(6), 473–483.
6. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., Haes, J. C. J., Jmd., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P. B., Schraub, S., Sneeuw, K., Sullivan, M., & Takeda, F. (1993). The European organization for research and treatment of cancer QLQ-C30: A quality-of-life instrument for use in international

- clinical trials in oncology. *Journal National Cancer Institute*, 85(5), 365–376. <https://doi.org/10.1093/jnci/85.5.365>
7. Atkinson, T. M., Stover, A. M., Storer, D. F., Saracino, R. M., D'Agostino, T. A., Pergolizzi, D., Matsoukas, K., Li, Y., & Basch, E. (2017). Patient-reported physical function measures in cancer clinical trials. *Epidemiologic Reviews*, 39(1), 59–70. <https://doi.org/10.1093/epirev/mxx008>
 8. Schalet, B. D., Hays, R. D., Jensen, S. E., Beaumont, J. L., Fries, J. F., & Cella, D. (2016). Validity of PROMIS physical function measured in diverse clinical samples. *Journal of Clinical Epidemiology*, 73, 112–118. <https://doi.org/10.1016/j.jclinepi.2015.08.039>
 9. Peipert, J. D., Chapman, R., Shaunfield, S., Kallen, M. A., Schalet, B. D., & Cella, D. (2022). Do you recall?: Results from a within-person recall study of the patient-reported outcomes measurement information system (PROMIS) short form v2.0 - physical function 8c. *Value Health*, 25(2), 161–166. <https://doi.org/10.1016/j.jval.2021.08.011>
 10. World Health Organization. World Health Organization disability assessment schedule: WHODAS II. Phase 2 field trials. Health services research. 2000.
 11. Bruce, B., & Fries, J. F. (2003). The stanford health assessment questionnaire: Dimensions and practical applications. *Health and Quality of Life Outcomes*, 1, 20. <https://doi.org/10.1186/1477-7525-1-20>
 12. Condon, D. M., Chapman, R., Shaunfield, S., Kallen, M. A., Beaumont, J. L., Eek, D., Mitra, D., Benjamin, K. L., McQuarrie, K., Liu, J., Shaw, J. W., Martin Nguyen, A., Keating, K., & Cella, D. (2020). Does recall period matter? Comparing PROMIS(R) physical function with no recall, 24-hr recall, and 7-day recall. *Quality of Life Research*, 29(3), 745–753. <https://doi.org/10.1007/s11136-019-02344-0>
 13. Kivunja C, Kuyini AB. Understanding and Applying Research Paradigms in Educational Contexts. Available at: <https://files.eric.ed.gov/fulltext/EJ1154775.pdf>. Accessed July 25, 2024. *International Journal of Higher Education*. 2017;6(5):26–41.
 14. Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1—eliciting concepts for a new PRO instrument. *Value Health*, 14(8), 967–977. <https://doi.org/10.1016/j.jval.2011.06.014>
 15. Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: Part 2—assessing respondent understanding. *Value Health*, 14(8), 978–988. <https://doi.org/10.1016/j.jval.2011.06.013>
 16. Patton MQ. Qualitative research & evaluation methods: integrating theory and practice. Fourth edition. ed. Thousand Oaks, California: SAGE Publications, Inc.; 2015.
 17. Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82.
 18. Willis G. Analysis of the Cognitive Interview in Questionnaire Design. Oxford University Press; 2015.
 19. Matthews J. Review: The Psychology of Survey Response. [Website]. 2011. Available at: <https://www.effortmark.co.uk/review-psychology-survey-response/>. Accessed September 7, 2023.
 20. Willis, G. B., & Miller, K. (2011). Cross-cultural cognitive interviewing: Seeking comparability and enhancing understanding. *Field Methods*, 23(4), 331–341.
 21. Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
 22. Weinfurt, K. P. (2019). Viewing assessments of patient-reported health status as conversations: Implications for developing and evaluating patient-reported outcome measures. *Quality of Life Research*, 28(12), 3395–3401. <https://doi.org/10.1007/s11136-019-02285-8>
 23. Arizmendi, C., Wang, S., Kaplan, S., & Weinfurt, K. (2024). Evaluating recall periods for patient-reported outcome measures: A systematic review of quantitative methods. *Value Health*, 27(4), 518–526. <https://doi.org/10.1016/j.jval.2024.01.016>
 24. Smith, S. M. S., Jan, S., Descallar, J., & Marks, G. B. (2019). An investigation of methods to improve recall for the patient-reported outcome measurement in COPD patients: A pilot randomised control trial and feasibility study protocol. *Pilot Feasibility Studies*, 5, 92. <https://doi.org/10.1186/s40814-019-0475-9>

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