



Systematic Review of the Effect of a One-Day Versus Seven-Day Recall Duration on Patient Reported Outcome Measures (PROMs)

Tessa Peasgood¹ · Julia M. Caruana¹ · Clara Mukuria²

Accepted: 9 November 2022 / Published online: 14 February 2023
© The Author(s) 2023

Abstract

Background There is ongoing uncertainty around the most suitable recall period for patient-reported outcome measures (PROMs).

Method This systematic review integrates quantitative and qualitative literature across health, economics, and psychology to explore the effect of a one-day (or ‘24-h’) versus seven-day (or ‘one week’) recall period. The following databases were searched from database inception to 30 November 2021: MEDLINE, EMBASE, PsycINFO, Web of Science, EconLit, CINAHL Complete, Cochrane Library, and Sociological Abstracts. Studies were included that compared a one-day (or ‘24-h’) versus seven-day (or weekly) recall period condition on patient-reported scores for PROM and Health-Related Quality-of-Life (HRQoL) instrument scores in adult populations (aged 18 and above) or combined paediatric and adult populations with a majority of respondents aged over 18 years. Studies were excluded if they assessed health behaviours only, used ecological momentary assessment to derive an index of daily recall, or incorporated clinician reports of patient symptoms. We extracted results relevant to six domains with generic health relevance: physical functioning, pain, cognition, psychosocial wellbeing, sleep-related symptoms and aggregated disease-specific signs and symptoms. Quantitative studies compared weekly recall scores with the mean or maximum score over the last seven days or with the same-day recall score.

Results Overall, across the 24 quantitative studies identified, 158 unique results were identified. Symptoms tended to be reported as more severe and HRQoL lower when assessed with a weekly recall than a one-day recall. A narrative synthesis of 33 qualitative studies integrated patient perspectives on the suitability of a one-day versus seven-day recall period for assessing health state or quality of life. Participants had mixed preferences, some noted the accuracy of recall for the one-day period but others preferred the seven-day recall for conditions characterised by high symptom variability, or where PROMs concepts required integration of infrequent experiences or functioning over time.

Conclusion This review identified a clear trend toward higher symptom scores and worse quality of life being reported for a seven-day compared to a one-day recall. The review also identified anomalies in this pattern for some wellbeing items and a need for further research on positively framed items. A better understanding of the impact of using different recall periods within PROMs and HRQoL instruments will help contextualise future comparisons between instruments.

Plain English Summary Questionnaires ask patients about their health over different time periods (e.g., “what were your symptoms like over the last week?” versus “what were your symptoms like today?”). Studies find that people may report their symptoms as more severe when they are asked to think about their symptoms over the last week compared to the last day. Understanding how different time periods influence patient responses will allow researchers to compare and develop new questionnaires and may help clinicians to choose the best questionnaire to understand their patient’s condition. We conducted a systematic literature review on studies which had looked at the impact of using different recall periods on patient responses. We found 24 studies that compared patient scores from questionnaires asking their health “over the last day” compared to “over the last week”. Overall, symptoms tended to be reported as more severe and health as poorer when they were reported over the last week compared to the last day on average. We also found 33 studies that asked patients to describe which recall period they preferred. Patients had mixed preferences with more preferring a seven-day recall where symptoms and health impacts varied a lot.

Key Points for Decision Makers

The findings of 24 included quantitative studies suggest that symptoms tend to be reported as more severe and health as poorer when reported over the last seven days compared to the last day.

The 33 included qualitative studies found that respondents had mixed preferences towards the different recall periods with a slight preference for seven-day recall where symptoms and health impacts varied a lot.

There are research gaps in understanding the impact of a one-day versus seven-day recall period for patients with mental health conditions and when asking positively framed questions.

1 Introduction

Patient-reported outcome measures (PROMs) are validated instruments or questionnaires used to collect information on a patient's health condition directly from the patient. One class of PROM instrument is that designed to assess the multi-dimensional construct health-related quality of life (HRQoL). Patient-reported outcome measures frame questions to patients within a particular recall period, such as asking about the severity of a symptom experienced or the presence of a symptom within, for example, 'today' or 'the last four weeks'. The choice of recall period may impact upon the answer. Short recall periods may not pick up symptoms or problems if they have not been experienced in that specified short period whereas long recall periods may suffer from recall bias and introduce uncertainty regarding what information respondents draw upon to answer them for example, they may use an assessment of their average symptoms over the time period, their worst symptoms or their recent symptoms [1, 2].

There is ongoing uncertainty around the most suitable recall period for assessing HRQoL [1–3]. The optimal recall period is driven by a number of concerns including: the objective of collecting PROM data, the nature and stability of the condition being assessed [4, 5] and the domain of assessment [2].

Patient-reported outcome measures may be collected in order to (i) gain knowledge about a disease trajectory; (ii) monitor and assess individual patients to support clinical decision making; (iii) evaluate care quality; and (iv) assess the effectiveness or cost effectiveness of treatments. The purpose of collecting the PROM data and the information needs of the decision at hand may influence the appropriate recall period [6–9].

The recall period selected for a PROM may influence the way in which respondents interpret questionnaire items and select relevant information to formulate a response. Poor memory may influence responses when individuals are asked to respond using longer recall periods, and this may differ by health domain (e.g., pain versus fatigue) [3]. For domains influenced by events (e.g., episodes or activities), recall may be impacted by the tendency to remember events as happening more recently than they actually did (referred to as "forward telescoping"), which can influence whether events are considered relevant to the recall period [10].

Longer recall periods may also lead participants to pay increased attention to salient events that are not representative of their general health state throughout the period, which may increase symptom severity reports (see: Kahneman et al. 1993; Stone et al. 2008) [9, 11]. Alternately, longer recall periods may result in reliance upon overall symptom or domain evaluations, rather than drawing upon specific episodes [12]. Reporting of mood-related symptoms may be influenced by longer recall periods that change the interpretation of emotion frequency questions [9]. For example, when referring to anger symptoms, more serious and intense episodes have been reported over a longer time frame [13].

Characteristics of the questionnaire item format may also interact with the influence of recall period on symptom reports. Participants may be influenced by positive or negative framing of questionnaire items (e.g., feeling energetic vs feeling tired) [14] and framing of outcomes in response options (e.g., symptom severity vs frequency). Repeated questionnaire administration may have carry-over effects where current responses influence future responses [15], which is relevant where the use of a short recall period requires repeated administration.

This review updates and refines the scope of previous reviews by adopting a targeted approach to the comparison of a one-day versus seven-day recall period on PROMs. While previous reviews (Schmier et al. 2004 [6]; Stull et al. 2009 [9]) suggest the presence of recall duration effects, they included little evidence specifically on the one-day versus seven-day recall comparison. A particular motivation of this review was to understand the potential impact of recall period on differences between the EQ-5D [17], which adopts the recall period of 'today', and the EQ-HWB (EQ Health and Wellbeing [18]) which adopts a recall period of 'the last 7 days'. Both measures are generic measures used to estimate utility scores for input into economic evaluations of healthcare [19], with EQ-5D focused on health and EQ-HWB on health and wellbeing or broader quality of life. Thus, the primary aim of this review was to determine recall period effects of a one-day versus seven-day recall period for domains included within the EQ-5D or the EQ-HWB.

2 Methods

This systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [20] and was prospectively registered with the PROSPERO database (ID: CRD42021251857).

2.1 Data Sources and Searches

The following sources were searched from database inception to 30 November 2021: MEDLINE, EMBASE, PsycINFO, Web of Science, EconLit, CINAHL Complete, Cochrane Library, and Sociological Abstracts. Search keywords were developed in consultation with an Academic Librarian and included ‘Patient Report’ OR ‘PROMs’ AND ‘Recall Duration’ and related terms, in addition to the recall duration comparator condition (see Supplementary Information S1Search terms). Findings were limited to the English language. Manual searches were conducted across the reference lists of recovered articles and relevant systematic reviews. Unpublished studies were sought from researchers affiliated with the EuroQol Research Group who own the Intellectual Property Rights to the EQ-5D and the EQ-HWB.

2.2 Study Selection

2.2.1 Inclusion Criteria

This review included studies that compared a one-day (or ‘24-h’) versus seven-day (or weekly) recall period condition on patient-reported scores for PROM and HRQoL instrument scores in adult populations (aged ≥ 18 years) or combined paediatric and adult populations with a majority of respondents aged over 18. The one-day recall condition included recall instructions of “over the last (or past) day”, “over the last (or past) 24 hours”, or “today”. The seven-day recall condition included recall instructions of “over the last (or past) seven days”, “over the last (or past) week”, or “this week”.

Four categories of studies were included:

- a) Studies which made comparisons between single or multiple items on the domains covered in the EQ-HWB or EQ-5D. These included: physical functioning (mobility, self-care [or personal care], daily activities, meaningful activities, hearing, vision), pain (pain and discomfort), cognition (memory and concentration), psychosocial wellbeing (loneliness, belonging, support, coping, self-worth, anxiety, depression, hope, safety/fear, anger/frustration), sleep-related symptoms (sleep disturbance, fatigue).

- b) Studies which made comparisons based on overall summary scores of HRQoL.
- c) Studies which used multi-item instruments to measure disease specific signs and symptoms where measures had an aggregate score (e.g., respiratory symptom severity in chronic obstructive pulmonary disease) [21]. Although less relevant to the central interest of comparing generic domains from the EQ-5D and EQ-HWB instruments summary symptom scores were included as a means of confirming findings.
- d) Qualitative studies exploring patient-reported perspectives of the suitability of a one-day versus seven-day recall duration period in PROM and HRQoL instruments.

2.2.2 Exclusion Criteria

Studies were excluded if the sample of participants was aged ≤ 18 years exclusively. Studies were excluded if they did not compare both a one-day and seven-day recall period, or if they assessed health behaviours only (e.g., tobacco smoking, physical activity). Studies were excluded if ecological momentary assessment (EMA [22]) was used to derive an index of daily recall, or if studies incorporated clinician reports of patient symptoms. Single-item condition-specific symptom reports (i.e., vomiting in the context of cancer treatment) [23] were considered less relevant for recall period comparisons of generic domains of health and wellbeing that are included in the EQ-HWB and EQ-5D instruments, which is the focus of interest for this review, and were therefore excluded from the scope of this review. Studies not available in the English language were excluded.

2.3 Data Extraction

After removal of duplicates, two researchers (JC and TP) independently screened titles and abstracts. Both authors applied eligibility criteria, and a final list of included articles was developed through consensus. Data were extracted from the included articles using a predetermined data extraction form by JC and cross-checked for accuracy by TP. Data extracted from quantitative studies included: participant and study characteristics (diagnosis, symptoms assessed, method of recall condition comparison, analytical approach), questionnaire characteristics (domains assessed, recall instruction, number of items, item framing, response options, score range, administration mode, time of responding, response rates) and recall condition effects on outcome scores as per instrument scales (score means and standard deviations for both recall conditions, and score differences between recall conditions). Where studies included the correlation of scores between recall conditions, this was only extracted where

score differences were not reported. In studies where data were reported for more than one time period and results were not averaged, only baseline data were extracted.

Data extracted from the qualitative studies included participant and study characteristics, questionnaire characteristics (including recall instructions), study methods (whether the study was nested in a qualitative study), the study context (instrument development, instrument validation or clinical trial), the methodology (focus group, interviews), the interview technique (cognitive debriefing, concept elicitation, think aloud), the analysis approach and summary results as reported by the author. In addition, any participant quotes relating to the recall period were also extracted.

2.4 Study Quality

As there was no suitable single quality check list that could be applied to studies comparing recall period, study quality was assessed using a subset of the relevant criteria extracted from the COSMIN checklist for assessing the risk of bias of PROMs (see Supplementary Information 'SI5. Quality Assessment' for a full list of criteria used) [24]. To evaluate risk of bias in quantitative studies, we assessed aspects of structural validity (e.g., sample size adequacy and statistical methods) and reliability (e.g., test conditions and study design validity). Some study designs use a standard instrument with a non-standard recall period. While recall period adjustment may have interfered with instrument validity, this was considered preferable to using a different instrument with a different recall period.

For qualitative studies, we assessed the quality of the study design and analysis (e.g., sample size adequacy, probing techniques, and data analyses). There were no exclusion criteria based on quality indicators.

2.5 Data Analysis and Synthesis

Extraction and analysis for the quantitative and qualitative data were undertaken by JC and TP. Characteristics of the quantitative and qualitative studies were summarised from the extracted information including the clinical group, the outcome assessed, sample size and the main findings relating to recall period comparisons. For the quantitative studies, findings were summarised within domains (physical function, pain, cognition, psychosocial wellbeing, fatigue & sleep), overall scores and aggregate measures of disease-specific signs. Within each, extracted scores from the seven-day versus one-day recall (mean of daily scores, maximum of daily scores and the same-day score) were assessed to identify if there were differences and whether these were statistically significant. For the qualitative studies, the extracted information on preferences or views related to the seven-day

versus one-day recall period were summarised descriptively within similar domains to the quantitative studies. The extracted quotes were coded thematically.

Comparison of recall period effects on instrument scores were not meta-analysed due to high levels of variability in patient groups, instruments, and methods of data collection and analysis. Instead, the statistical results of recall condition comparisons were synthesised into summary tables to gain insight into the presence of trends and systematic differences within assessment domains.

To help present a broad visual overview of any trends in the direction of the differences between scores differences were flagged where one-day scores (for mean of the daily score or the same-day score) were 10% lower for symptoms than weekly scores (10% higher for quality of life or functioning, re-scaling where necessary to start the scoring from zero). A convenient level of 10% was chosen to communicate a difference between scores, given the many different instruments and scales included.

To facilitate comparisons, on the summary table flagged results are coded green. Studies that found the opposite direction of difference, with daily scores higher for symptoms (lower for quality of life) than weekly scores are coded in red. Studies which found daily scores to be lower for symptoms (higher for quality of life) but with a difference of under 10% in the score or for which a percentage change from the weekly score is not possible to calculate (e.g., scores represented as T scores) are coded in amber. No differences between the maximum daily score and the seven-day scores were flagged as this represents a different type of comparison.

Conclusions drawn from qualitative studies assessing patient perspectives on the suitability of a one-day versus seven-day recall period for PROMs and HRQoL instruments were integrated into a narrative synthesis. Qualitative findings were summarised on a table showing each study's conclusion of their respondents most preferred recall period (one day, seven days or more than seven days) and whether this was drawn from close-ended questions (which asked for endorsement of a given recall period) or open-ended questions in which multiple given recall periods were discussed or questions were asked about the ideal recall period in that context.

3 Results

3.1 Search Results

In total, 945 records (excluding duplicates) were identified, and the titles and abstracts were screened. Full text versions were retrieved for 82 articles, of which 57 were eligible for

inclusion. Of these, 24 reported quantitative comparisons of one-day versus seven-day recall scores. The remaining 33 studies reported patient perspectives of optimal recall duration and were included in the narrative synthesis of qualitative studies. Figure 1 shows the flow of studies through the review and reasons for exclusion.

3.2 Characteristics of Included Studies

The quantitative and qualitative studies included in this review assessed adults with a diverse range of clinical conditions and from the general population (see Supplementary Information S12 and SI3).

3.2.1 Characteristics of Quantitative Studies Included

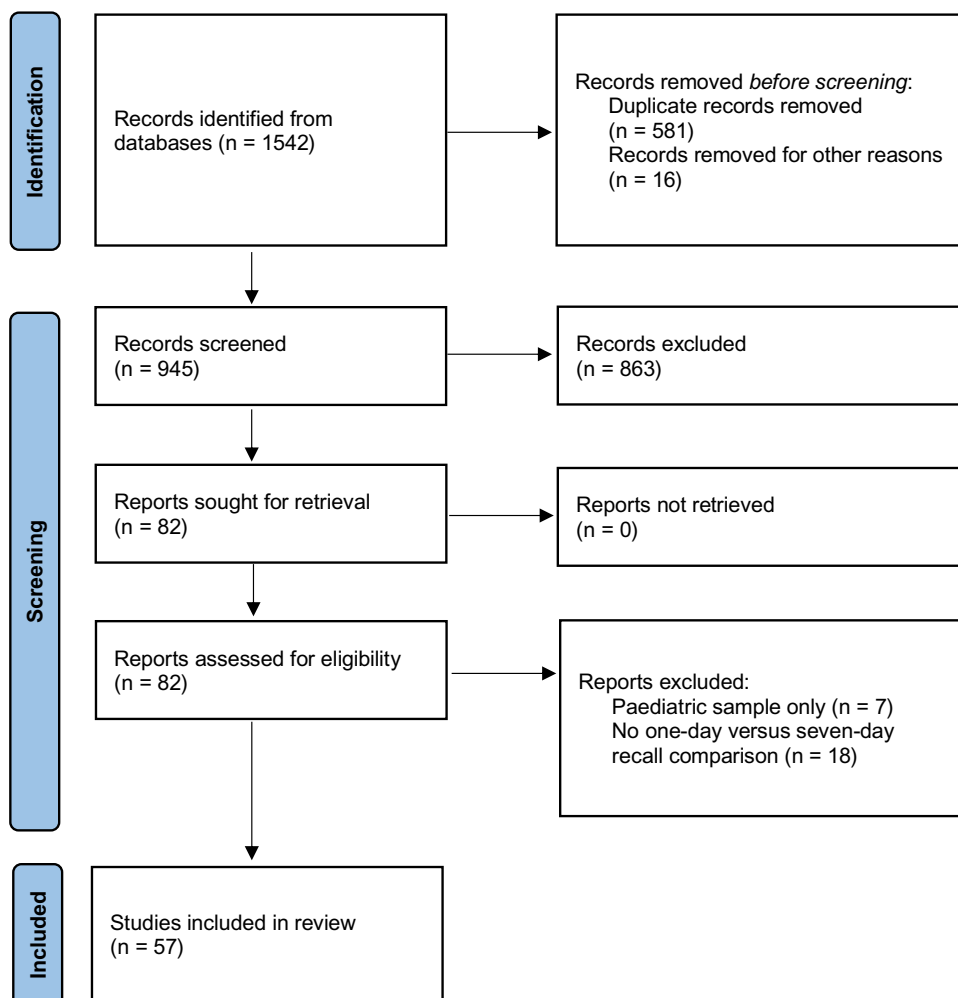
A total of 4701 participants were included across the 24 quantitative studies assessed. Sample sizes of individual studies ranged from 32 to 800 participants (median = 113; mean = 196 [SD = 206]), with 57.9% of the total sample

being women. Most (23 of 24) studies included only adults aged ≥ 18 years, while one study [25] included a blended sample comprising 34% paediatric participants aged between 12 and 18 years. Most (22 of 24) studies included participants diagnosed with a clinical condition, while four studies included individuals selected from the general population [26–29].

The instruments used to evaluate the effect of recall duration assessed either the signs, symptoms and impacts of a disease and its treatment, or quality of life generally. Instruments were mostly disease-specific (21 of 24, e.g., Psoriasis Signs and Symptoms Diary), but also included generic HRQoL instruments (3 [30–32] of 24, e.g., EQ-5D). Outcomes assessed in participants selected from the general population included pain; [27–29, 33] fatigue [27, 28, 33]; emotional states [27–29]; and physical functioning [26].

Recall instructions for the one-day recall condition included “today” or “during the day” (3 of 24 studies), “over the last (or past) 24 hours” (17 of 24), and “over the last (or past) day” (4 of 24). Recall instructions for the seven-day

Fig. 1 Identification and selection of quantitative and qualitative studies for this review. Adapted from Page et al. [20]



recall condition included “over the last (or past) seven days” (15 of 24) and “over the last (or past) week” (9 of 24). Most (17 of 24) studies used the same instrument adjusted only for recall period instruction.

The data collection period and number of assessments conducted differed between studies. Study data collection periods ranged from 1 [26, 29, 31, 34–37] to 100 [38] days with the number of daily recall assessments also ranging from 1 to 100, and weekly recall assessments from 1 to 14. Questionnaire administration formats included paper forms (10 of 24) [30, 31, 36–43], online (9 of 24) [23, 25–29, 32, 33, 38], electronic tablet or palm pilot (4 of 24) [21, 34, 35, 44], or by telephone (2 of 24) [45, 46]. Overall, response rates for weekly recall questionnaires ranged from 52% [38] to 100% [26, 29, 31, 34] (based on the lowest reported rate within each study: median = 94%, mean = 90% [SD = 13]), and response rates for daily recall questionnaires ranged 52%

[38] to 100% [26, 29, 31, 34] (median = 95%, mean = 89% [SD = 14]). Nine of 24 studies did not report weekly questionnaire response rates [23, 30, 35–37, 40, 42, 43, 45] 6 of 24 studies did not report daily questionnaire response rates [23, 30, 35–37, 45].

Three methods were used to index one-day recall scores for comparison with seven-day recall scores. These different approaches are shown in Figure 2b. First, in 11 of 24 studies [21, 23, 28, 32, 33, 39, 41, 42, 44–46], daily recall scores were averaged over seven consecutive days and compared to the seven-day recall score reported on the final assessment day (i.e., “mean” index). Second, for 6 of 24 studies [21, 23, 25, 38, 42, 46], the single highest daily recall score reported over seven consecutive days was compared to the seven-day recall of maximum (i.e., most severe, or worst) symptoms across the week (i.e., “maximum” index). Third, in 9 of 24 studies [21, 30, 31, 34–36, 40, 43, 78] one-day recall scores

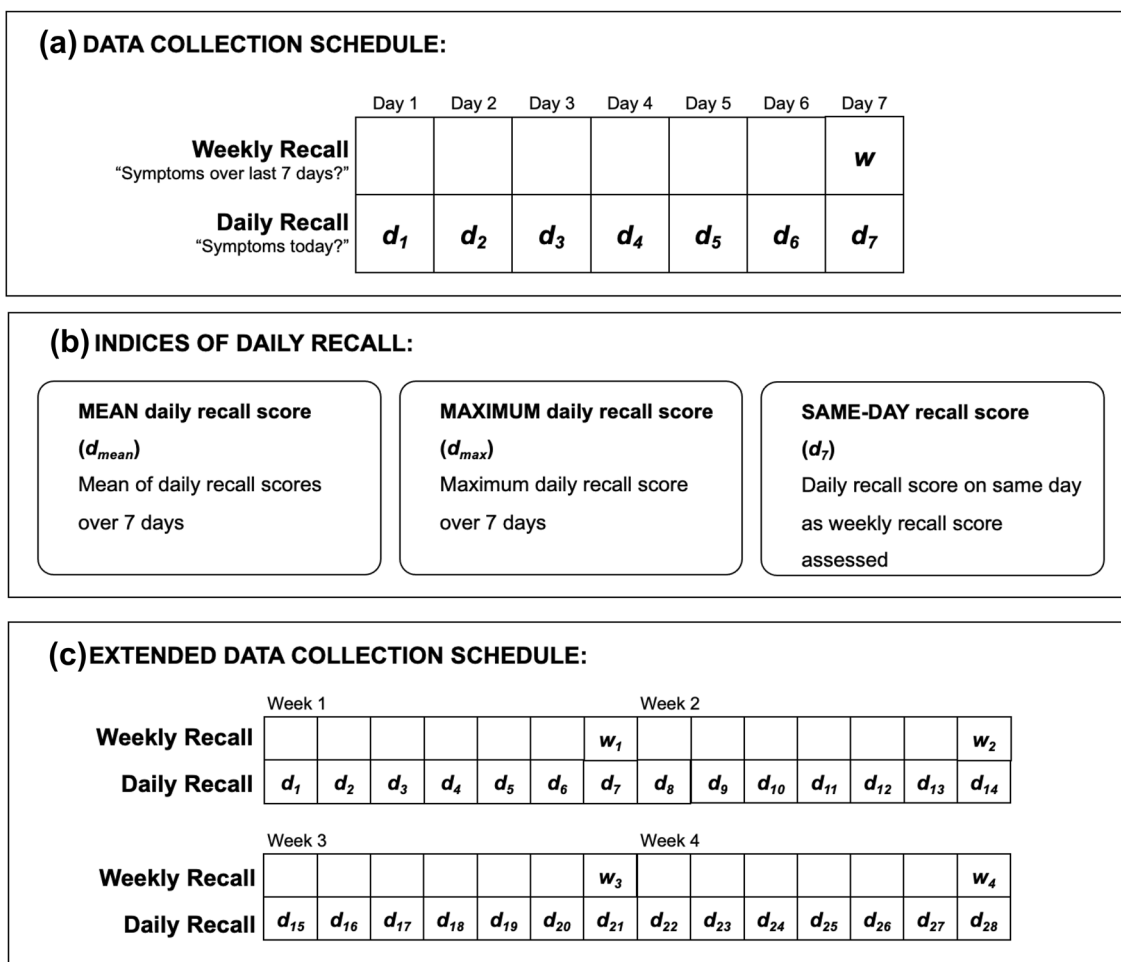


Fig. 2 Quantitative comparison of daily and weekly recall scores. In most studies, weekly and daily symptoms were assessed over the same consecutive seven-day period (part a in the figure). Three methods of daily symptom indexation were used in the quantitative comparison of daily versus weekly recall scores (i) the mean of daily

recall scores over the seven days (d_{mean}), (ii) the maximum of the daily recall scores (d_{max}), and (iii) the score reported on the same day as the weekly recall score (d_7) (part b in the figure). In some studies, the data collection period was extended beyond seven days to calculate an average of the chosen indexation method (part c in the figure)

were compared to scores for the seven-day recall instrument issued on the same day (i.e., “same day” index). One study [27] compared two separate days in which seven-day and one-day recall were asked in a random order to half the sample and compared; this was classified as ‘same-day’ index.

For these three methods of recall period comparison, if the one-day recall condition did not differ significantly from weekly recall scores across the sample on average, then the recall period was assumed to not have had a statistically significant effect on patient-reported outcomes. In some studies, the data collection period was extended beyond seven days to calculate an average of the chosen indexation method (see Extended Data Collection Schedule in Fig. 2). For example, in studies that assessed symptoms over 28 consecutive days, the weekly recall score was calculated by averaging across the four consecutive weeks of data collection (i.e., mean of W1, W2, W3, and W4). For the mean daily symptom index, the mean daily score was averaged from Day 1 to Day 28. For the maximum daily symptom index, the maximum daily score for each week was averaged over the four weeks. For the same-day symptom index, scores were averaged across Days 7, 14, 21, and 28.

Some studies report only the intraclass correlation (ICC) between scores; where this was the case, using guidelines from Koo and colleagues for ICCs [47] we judged 0.5–0.75 as moderate agreement, 0.75–0.9 as good, and above 0.9 as excellent.

In 2 [26, 27] of 24 studies, one-day and seven-day recall scores collected on different respondents were assessed for Differential Item Functioning (DIF) within an Item Response Theory (IRT) framework [27]. This method considers whether the responses to items using different recall periods are predicted equally well by knowledge of the underlying construct of interest (e.g., estimated level of pain or mobility).

Overall, across the 24 quantitative studies identified, the unique combinations of clinical condition (e.g., type 2 diabetes, psoriasis), symptom domain (e.g., physical functioning, psychosocial wellbeing), symptom descriptive (e.g., frequency, severity/intensity, impact/interference), and daily recall comparison method (e.g., mean, maximum, same-day, DIF) gave rise to 158 unique results for data extraction.

Most of the 24 quantitative studies reviewed were considered of reasonable quality with only minor methodological flaws (see Supplementary Information ‘SI5. Quality Assessment’). Three studies used different instruments or items to assess the recall condition. Most studies did not control for the effect of repeated questionnaire administration or recall period order. In the studies that did control for effects of repeated administration through study design, participants completed the daily questionnaires and weekly questionnaires across separate time periods, with participants randomly allocated to the order in which they receive

each recall period. Only four studies randomised participants to recall period order. For the nine studies comparing one-day recall scores with seven-day recall scores reported on the same day, 44% (4 of 9) assessed one-day recall scores after repeated administration, while the remaining assessed one-day recall scores from only a single questionnaire administration.

In half of the studies (12/24) the sample size was judged inadequate to support statistical analyses. Test conditions were similar between environments in most studies; one study did not have similar test conditions and nine had some uncertainty, mostly relating to the evidence provided on the time of day in which questionnaires were completed.

3.2.2 Characteristics of Qualitative Studies Included

In total, 1244 participants were included across the 33 qualitative studies reviewed. Sample sizes of individual studies ranged from 7 [48] to 207 [8] participants (median = 25; mean = 39 (SD = 41)). Of the 33 qualitative studies reviewed, five assessed fatigue and sleep-related symptom, three assessed pain-related symptoms, and one assessed physical functioning, eight assessed HRQoL and 17 assessed disease-specific signs and symptoms.

Qualitative methods included: one-on-one interviews (91%, 30 of 33), focus groups (21%, 7 of 33), and online survey (3%, 1 of 33). Data collection methods included: cognitive debriefing (76%, 25 of 33), concept elicitation (76%, 25 of 33), “think aloud” (15%, 5 of 33), and Delphi consensus (3%, 1 of 33).

Detailed responses to the COSMIN checklist criteria [24] used to assess study quality are provided in the Supplementary Information ‘SI5. Quality Assessment’. Most of the qualitative studies reviewed were considered as high quality. All 33 studies used appropriate qualitative study methods (e.g., individual interviews, focus groups, Delphi survey); 48% (16 of 33) of studies used open-ended probing techniques to elicit participant perspectives of recall duration. In contrast, 52% (17 of 33) studies used closed-ended probes to assess participant endorsement of a predetermined recall period, which may have been subject to framing effects. 97% (32 of 33) studies were conducted with an appropriate number of participants according to the COSMIN criteria (i.e., $N \geq 7$ [24]), while one study was not conducted in an adequate sample size ($N = 2$ [8]).

For the 32 studies that involved participant interviews or focus group, 41% (13 of 32) indicated the use of skilled moderators or interviewers; however, the majority (59%, 19 of 32) provided no indication of interviewer training or expertise. All 32 studies that involved participant interviews or focus groups indicated using an interview guide, and the majority (94%, 30 of 32) indicated audio recording and verbatim transcription of interviews. Most studies (31

of 33) used appropriate analysis techniques (e.g., thematic or content analyses), and 59% (19 of 33) clearly indicated involvement of at least two researchers in analyses.

3.3 Assessment of Recall Duration Effects

3.3.1 Physical Functioning

Eleven studies compared one-day and seven-day recall on instruments assessing physical functioning providing 20

unique results for data extraction (see Table 1). For the nine results using the mean daily recall indexation method, the majority (7 [21, 33, 34, 39, 44]) found weekly recall scores were lower than mean daily recall scores and 2 [27, 28] found no evidence of a significant difference. The single result using the maximum daily recall indexation method found that weekly scores were less than maximum daily recall scores [21]. Nine of the 10 results using the same-day recall indexation method found no significant difference between weekly and same-day recall scores [26, 31, 34, 35,

Table 1 Study results assessing the effect of a 7-day versus one-day recall period on patient-reported outcomes

1 st Author, Year	Outcome	Condition	N	seven-day Recall	one-day Recall Index			Study Conclusion
					Mean ^a	Maximum ^b	Same Day ^c	
Physical Functioning (N = 20).								
Armstrong, 2014 ³⁴	General Activity	Primary Brain Tumour	100	M(SD)= 2.28(2.52)	-	-	M(SD)= 1.98(2.39)	No sig. difference between week and same day.
Armstrong, 2014 ³⁴	Vision	Primary Brain Tumour	100	M(SD)= 1.13(2.08)	-	-	M(SD)= 0.98(2.00)	No sig. difference between week and same day.
Armstrong, 2014 ³⁴	Walking	Primary Brain Tumour	100	M(SD)= 1.74(2.45)	-	-	M(SD)= 1.65(2.54)	No sig. difference between week and same day.
Bennett, 2011 ³⁹	Difficulty Doing Daily Tasks	Type 2 Diabetes	136	M(SD)= 1.55(1.91)	M(SD)= 1.34(1.64)**	-	-	Mean daily < weekly.
Bennett, 2012 ²¹	Frequency of Activity Limitations due to breathing problems	COPD	98	M(SD)= 2.02(1.43)	M(SD)= 1.53(1.37)***	M(SD)= 2.42(1.57)**	1.45 (1.54)**	Mean daily and same day < weekly < maximum daily.
Broderick, 2010 ⁴⁴	Fatigue interference with normal work	Rheumatological Illness	87	No values reported ⁱⁱ	No values reported ⁱⁱ	-	-	Mean daily < weekly (p < 0.01).
Broderick, 2010 ⁴⁴	Fatigue interference with walking	Rheumatological Illness	87	No values reported ⁱⁱ	No values reported ⁱⁱ	-	-	Mean daily < weekly (p < 0.01).
Broderick, 2010 ⁴⁴	Pain interference with normal work	Rheumatological Illness	87	No values reported ⁱⁱ	No values reported ⁱⁱ	-	-	Mean daily < weekly (p < 0.01).
Broderick, 2010 ⁴⁴	Pain interference with walking	Rheumatological Illness	87	No values reported ⁱⁱ	No values reported ⁱⁱ	-	-	Mean daily < weekly (p < 0.01).
Broderick, 2013	Physical Functioning ^v	Osteoarthritis	98	M(SD)= 37.50(6.80)	M(SD)= 36.90(6.50)***	-	-	Mean daily < weekly.
Condon, 2020 ²⁶ (DIF)	Physical Function ^{iv, v}	Cancer	200	M(SD)= 39.10(8.50)	-	-	M(SD)= 39.30(8.00)	DIF: No sig. difference between weekly and mean daily.
Condon, 2020 ²⁶ (DIF)	Physical Functioning ^{iv, v}	General Population	200	M(SD)= 48.40(11.10)	-	-	M(SD)= 49.70(10.30)	DIF: No sig. difference between weekly and mean daily.
De Andre Ares, 2015 ³⁵	Pain interference with daily activities	Non-Cancer-related pain	698	M(SD) = 7.2 (2.5)	-	-	No values reported ⁱ	No sig. difference between weekly recall and same day 24 hour recall
De Andre Ares, 2015 ³⁵	Pain interference with walking	Non-Cancer-related pain	698	M(SD) = 6.4 (1.7)	-	-	No values reported ⁱ	No sig. difference between weekly recall and same day 24 hour recall
Schneider, 2013 ²⁷ (DIF)	Pain interference with daily activities	General Population	100	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	DIF: No sig. difference between weekly and mean daily.
Shi, 2010 ³⁷	Activity Interference	Chemotherapy for cancer	42	No values reported ⁱ	-	-	No values reported ⁱ	No sig. difference between weekly and mean daily.
Stone, 2016 ²⁸	Physical Functioning	Osteoarthritis	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	No sig. difference between weekly and mean daily.
Thavarajah, 2013 ³¹	Physical Wellbeing ^{iv}	Brain Metastases	40	M(SD)= 20.90(5.60)	-	-	M(SD)= 21.00(5.30)	No sig. difference between weekly and same day.
Pain-Related Symptoms (N=37)								
Armstrong, 2014 ³⁴	Pain Severity	Primary Brain Tumour	100	M(SD)= 1.61(2.33)	-	-	M(SD)= 0.93(0.70)**	Same day < weekly.
Bennett, 2011 ³⁹	Frequency of Aches & Pains	Type 2 Diabetes	139	M(SD)= 2.60(2.32)	M(SD)= 2.14(2.04)***	-	-	Mean daily < weekly.
Broderick, 2008 ⁴¹	Pain Interference & Severity	Rheumatological Illness	83	No values reported ⁱⁱ	No values reported ⁱⁱ	-	-	Mean daily < weekly (p-value not reported).
Broderick, 2010 ⁴⁴	Pain Interference	Rheumatological Illness	87	No values reported ⁱⁱ	No values reported ⁱⁱ	-	-	Mean daily < weekly (p < 0.01).

Table 1 (continued)

Broderick, 2013 ³³	Pain Intensity	General Population	98	M(SD)= 2.48(2.40)	M(SD)= 2.07(2.20)***	-	-	Mean daily < weekly.
Broderick, 2013 ³³	Pain Intensity	Osteoarthritis	98	M(SD)= 5.51(1.90)	M(SD)= 5.23(1.90)***	-	-	Mean daily < weekly.
Broderick, 2013 ³³	Pain Interference ^v	General Population	98	M(SD)= 51.30(9.00)	M(SD)= 48.60(7.70)***	-	-	Mean daily < weekly.
Broderick, 2013 ³³	Pain Interference ^v	Osteoarthritis	98	M(SD)= 60.90(6.10)	M(SD)= 51.30(9.00)***	-	-	Mean daily < weekly.
De Andre Ares, 2015 ³⁵	Pain intensity and interference	Non-Cancer-related pain	698	No values reported ⁱ	-	-	No values reported ⁱ	Positive correlation between weekly and same day (ICC= 0.92 – 0.94).
Kamper, 2015 ³⁶	Pain Intensity	Whiplash Disorders	146	M(SD)= 57.6(19.8)	-	-	M(SD)= 51.50(20.40)	Same day < weekly.
Lackner, 2014 ⁴²	Average abdominal pain intensity	Irritable Bowel Syndrome	273	M(SD)= 4.61(2.00)	M(SD)= 3.50(1.80)*	-	-	Mean daily < weekly.
Lackner, 2014 ⁴²	Worst abdominal pain intensity	Irritable Bowel Syndrome	273	M(SD)= 6.66(2.30)	-	M(SD)= 6.69(2.20)	-	Weekly not sig. different from maximum daily.
Marty, 2009 ⁴⁵	Pain Intensity (0-10 scale)	Chronic Lower Back Pain	150	No values reported ⁱⁱⁱ	Difference (daily-weekly) = -0.45	-	-	Mean daily (current) < weekly.
Mathias, 2016 ⁴³	Pain	Plaque Psoriasis	106	No values reported ⁱ	No values reported ⁱ	-	-	Negative correlation between mean daily and weekly ($r = -0.67$). Instruments oppositely scored.
Mendoza, 2017 ²³	Pain Frequency	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	Weekly not sig. different from mean daily. Weekly not sig. different from maximum daily.
Mendoza, 2017 ²³	Pain Impact	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	Mean daily not sig. different from weekly. Maximum daily > weekly
Mendoza, 2017 ²³	Pain Severity	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	Mean daily not sig. different from weekly. Maximum daily > week
Schaffer, 2021 ⁴⁶	Pain on Average	Prostate Cancer	119	M(SD)= 3.60(1.82)	M(SD)= 3.40(1.71)	M(SD)= 4.24 (2.00)**	-	Mean daily < weekly but not sig. different. Maximum daily > weekly.
Schaffer, 2021 ⁴⁶	Pain at its worst	Prostate Cancer	119	M(SD)= 5.58(2.24)	M(SD)= 4.81(1.97)**	M(SD)= 5.96 (2.16)	-	Mean of daily < weekly. Maximum daily > weekly but not sig. different.
Schneider, 2013 ²⁷ (DIF)	Pain Interference	General Population	100	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	DIF: No sig. difference between mean daily and weekly.
Shi, 2010 ³⁷	Pain Severity	Chemotherapy for cancer	42	No values reported ⁱ	-	-	No values reported ⁱ	No sig. difference between same day and weekly (p-value not reported).
Stone, 2016 ²⁸	Pain Intensity	Breast Cancer	85	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$)
Stone, 2016 ²⁸	Pain Intensity	General Population	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$)
Stone, 2016 ²⁸	Pain Intensity	Hernia Repair Surgery	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$)
Stone, 2016 ²⁸	Pain Intensity	Osteoarthritis	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$)
Stone, 2016 ²⁸	Pain Intensity	Premenstrual Syndrome	93	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$)
Stone, 2016 ²⁸	Pain Interference	General Population	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$)
Stone, 2016 ²⁸	Pain Interference	Hernia Repair Surgery	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$)
Stone, 2016 ²⁸	Pain Interference	Osteoarthritis	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$)

37, 37] with one study finding that the same-day score was lower [21].

Where the daily scores are at least 10% lower than the weekly score (re-scaling where necessary to start the scoring from zero) for health problems or 10% higher for quality of life (excluding comparisons based on maximum problems)

results are colour coded as green, regardless of significance level. Coral indicates less than 10% difference in recall duration score in the same direction, or comparisons in the same direction but for which a percentage increase from the weekly score is not possible to calculate (e.g., scores

Table 1 (continued)

Stone, 2016 ²⁸	Pain Behaviour	Hernia Repair Surgery	98	<i>No values reportedⁱⁱⁱ</i>	<i>No values reportedⁱⁱⁱ</i>	-	-	Mean daily < weekly ($p < 0.001$)
Walentynowicz, 2018 ²⁹	Pain Frequency	General Population	469	M(SD)=1.743 (0.951)	-	-	M(SD)=1.717 (0.972)	Same day < weekly but not sig. different.
Wood, 2015 ³⁸	Pain Severity	Post-hematopoietic cell transplantation	32	M(SD)=0.86(1.01)	-	M(SD)=1.07(1.11), ES=0.20	-	Maximum daily > weekly.
Cognition (N=8).								
Armstrong, 2014 ³⁴	Difficulty remembering	Primary Brain Tumour	100	M(SD)=2.27(2.73)	-	-	M(SD)=2.42(2.71)	No sig. difference between weekly and same day.
Armstrong, 2014 ³⁴	Difficulty understanding	Primary Brain Tumour	100	M(SD)=1.61(2.17)	-	-	M(SD)=1.27(2.04)	No sig. difference between weekly and same day.
Bennett, 2011 ³⁹	Difficulty concentrating	Type 2 Diabetes	138	M(SD)=1.76(1.94)	M(SD)=1.42(1.57)***	-	-	Mean daily < weekly.
Shi, 2010 ³⁷	Difficulty remembering	Chemotherapy for cancer	42	<i>No values reportedⁱ</i>	-	-	<i>No values reportedⁱ</i>	No sig. difference between same day and weekly.
Schneider, 2013 ²⁷ (DIF)	Pain interference with concentration	General Population	100	<i>No values reportedⁱⁱⁱ</i>	<i>No values reportedⁱⁱⁱ</i>	-	-	DIF: No sig. difference between mean daily and weekly.
Schneider, 2013 ²⁷ (DIF)	Too tired to think clearly	General Population	100	<i>No values reportedⁱⁱⁱ</i>	<i>No values reportedⁱⁱⁱ</i>	-	-	DIF: No sig. difference between mean daily and weekly.
Wood, 2015 ³⁸	Concentration problems	Post-hematopoietic cell transplantation	32	M(SD)=0.50(0.59)	-	M(SD)=0.68(0.78), ES=0.25	-	Weekly < maximum daily.
Wood, 2015 ³⁸	Memory problems	Post-hematopoietic cell transplantation	32	M(SD)=0.47(0.65)	-	M(SD)=0.58(0.67), ES=(0.15)	-	No sig. difference between weekly and maximum daily.
Psychosocial wellbeing (N = 51)								
Armstrong, 2014 ³⁴	Severity of distress at its worst	Primary Brain Tumour	100	M(SD) = 2.49 (2.72)	-	-	M(SD) = 2.23 (2.65)	Weekly > Same day but not sig. different.
Armstrong, 2014 ³⁴	Frequency of interference with enjoyment of life	Primary Brain Tumour	100	M(SD) = 2.29 (2.84)	-	-	M(SD) = 1.98 (2.73)	Weekly > Same day but not sig. different.
Armstrong, 2014 ³⁴	Severity of irritability at its worst	Primary Brain Tumour	100	M(SD) = 1.75 (2.20)	-	-	M(SD) = 1.33 (1.83)	Weekly > Same day but not sig. different.
Armstrong, 2014 ³⁴	Frequency of interference with mood	Primary Brain Tumour	100	M(SD) = 2.33 (2.49)	-	-	M(SD) = 1.84 (2.35)**	Weekly > Same day
Armstrong, 2014 ³⁴	Frequency of interference with relations with others	Primary Brain Tumour	100	M(SD) = 1.67 (2.26)	-	-	M(SD) = 1.47 (2.25)	Weekly > Same day but not sig. different.
Armstrong, 2014 ³⁴	Severity of sadness at its worst	Primary Brain Tumour	100	M(SD) = 2.34 (2.86)	-	-	M(SD) = 1.92(2.77)	Weekly > Same day but not sig. different.
Bennett, 2010 ²⁵	Frequency of Poor Mood	Cystic Fibrosis	38	M(95%CI) = 0.29(0.15,0.43)	M(95%CI) = 0.2(0.09,0.31)**	M(95%CI) = 0.5(0.28,0.73)**	-	Mean daily < weekly < maximum of daily.
Bennett, 2011 ³⁹	Difficulty getting along with others	Type 2 Diabetes	138	M(SD) = 1.18(1.69)	M(SD) = 0.79 (1.25)**	-	-	Mean daily < weekly.
Bennett, 2011 ³⁹	Frequency of Irritability at its Worst	Type 2 Diabetes	138	M(SD) = 2.27(2.26)	M(SD) = 1.5 (1.55)***	-	-	Mean daily < weekly.
Bennett, 2012 ²¹	Frequency of Feeling Upset	COPD	98	M(SD) = 1.58(1.35)	M(SD) = 1.02 (1.11)***	M(SD) = 1.87(1.59)**	M(SD) = 0.95(1.26)**	Mean daily < weekly < maximum daily. Same day < weekly.
Broderick, 2010 ⁴⁴	Fatigue interference with relationships	Rheumatological Illness	104	<i>No values reportedⁱⁱ</i>	<i>No values reportedⁱⁱ</i>	-	-	Mean daily < weekly ($p < 0.01$).
Broderick, 2010 ⁴⁴	Pain interference with relationships	Rheumatological Illness	104	<i>No values reportedⁱⁱ</i>	<i>No values reportedⁱⁱ</i>	-	-	Mean daily < weekly ($p < 0.01$).
Mendoza, 2017 ²³	Anxiety Frequency	Adverse Events during Cancer Trials	126	<i>No values reportedⁱⁱⁱ</i>	Difference (weekly-daily) = 0.16 (Effect size 0.17)	Difference (weekly-daily) = -0.2 (Effect size -0.2)	-	Weekly < maximum daily. Mean daily not sig. different from weekly.

represented as T scores). Orange flags results showing the reverse relationship.

3.3.2 Pain-Related Symptoms

Sixteen studies compared one-day and seven-day recall on instruments assessing pain symptoms with 37 unique data

extraction points (see Table 1). For the 24 results using the mean daily recall indexation method, the majority (79%, 19 of 24 results) found weekly recalled scores were higher than mean daily recalled scores for pain-related symptoms; 21% (5 of 24 results) found no evidence of a significant difference. For the single study that assessed correlations between weekly and mean daily recall scores, a moderate

Table 1 (continued)

Mendoza, 2017 ²³	Anxiety Impact	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	Difference (weekly-daily) = 0 (Effect size 0.01)	Difference (weekly-daily) = -0.28 (Effect Size -0.3)	-	Weekly < maximum daily. Mean daily not sig. different from weekly.
Mendoza, 2017 ²³	Anxiety Severity	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	Difference (weekly-daily) = 0.08 (Effect Size 0.10)	Difference (weekly-daily) = -0.28 (Effect size -0.30)	-	Weekly < maximum daily. Mean daily not sig. different from weekly.
Mendoza, 2017 ²³	Sadness Frequency	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	Difference (weekly-daily) = -0.08 (Effect Size -0.12)	Difference (weekly-daily) = -0.14 (Effect Size -0.14)	-	Weekly < maximum daily. Mean daily not sig. different from weekly.
Mendoza, 2017 ²³	Sadness Impact	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	Difference (weekly-daily) = 0.01 (Effect Size 0.02)	Difference (weekly-daily) = -0.29 (Effect Size -0.34)	-	Weekly < maximum daily. Mean daily not sig. different from weekly.
Mendoza, 2017 ²³	Sadness Severity	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	Difference (weekly-daily) = 0.08 (Effect Size 0.13)	Difference (weekly-daily) = -0.11 (Effect Size -0.14)	-	Weekly < maximum daily. Mean daily not sig. different from weekly.
Schneider, 2013 ²⁷ (DIF)	Depression Frequency	General Population	100	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	DIF: No sig. difference between mean daily and weekly
Schneider, 2013 ²⁷ (DIF)	Frequency of pain interference with enjoyment of life	General Population	100	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	DIF: Weekly < Mean Daily
Schneider, 2013 ²⁷ (DIF)	Sadness frequency	General Population	100	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	DIF: Weekly < Mean Daily
Schneider, 2013 ²⁷ (DIF)	Worthlessness frequency	General Population	100	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	DIF: No sig. difference between mean daily and weekly
Shi, 2010 ³⁷	Distress Severity	Chemotherapy for cancer	42	No values reported ⁱ	-	-	No values reported ⁱ	No sig. difference between same day and weekly
Shi, 2010 ³⁷	Mood Interference	Chemotherapy for cancer	42	No values reported ⁱ	-	-	No values reported ⁱ	No sig. difference between same day and weekly
Shi, 2010 ³⁷	Sadness Severity	Chemotherapy for cancer	42	No values reported ⁱ	-	-	No values reported ⁱ	No sig. difference between same day and weekly
Stone, 2016 ²⁸	Anger	Premenstrual Syndrome	93	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$).
Stone, 2016 ²⁸	Anxiety	Breast Cancer	85	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$).
Stone, 2016 ²⁸	Anxiety	Hernia Repair Surgery	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$).
Stone, 2016 ²⁸	Depression	Breast Cancer	85	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$).
Stone, 2016 ²⁸	Depression	General Population	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$).
Stone, 2016 ²⁸	Depression	Premenstrual Syndrome	93	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly ($p < 0.001$).
Thavarajah, 2013 ³¹	Emotional Wellbeing ^{iv}	Brain Metastases	40	M(SD) = 16.8 (4.9)	-	-	M(SD) = 16.8 (4.9)	No sig. difference between same-day and weekly recall.
Thavarajah, 2013 ³¹	Social/ Family Wellbeing ^{iv}	Brain Metastases	40	M(SD) = 26.3 (3.6)	-	-	M(SD) = 26.3 (3.5)	No sig. difference between same-day and weekly recall.
Walentynowicz, 2018 ²⁹	Anger Frequency (1-5 scale)	General Population	469	M(SD) = 1.573(0.773)	-	-	M(SD) = 1.472(0.777)	Weekly > Same day but not sig. different
Walentynowicz, 2018 ²⁹	Anxiety Frequency (1-5 scale)	General Population	469	M(SD) = 2.269(1.120)	-	-	M(SD) = 2.057(1.117)*	Weekly > Same day
Walentynowicz, 2018 ²⁹	Excitedness Frequency ^{iv}	General Population	469	M(SD) = 2.473(1.115)	-	-	M(SD) = 2.372 (1.141)	Weekly > Same day but not sig. different.
Walentynowicz, 2018 ²⁹	Happiness Frequency ^{iv}	General Population	469	M(SD) = 3.107(1.087)	-	-	M(SD) = 3.057 (1.112)	Weekly > Same day but not sig. different.
Walentynowicz, 2018 ²⁹	Sadness Frequency	General Population	469	M(SD) = 1.816(1.010)	-	-	M(SD) = 1.649 (0.925)*	Weekly > Same day.

association was identified [43]. For the seven results using the maximum daily recall indexation method, majority (4 [23, 38, 46] of 7 results) found weekly recalled scores were lower than maximum daily recalled scores. The remaining 3 [23, 42, 46] results found no evidence of a significant difference between weekly and maximum daily recall scores. Of

the 5 results using the same-day recall indexation method, 2 [34, 36] found same-day recall scores to be lower than weekly recall scores, 2 found no significant difference [29, 37], while 1 [35] identified a positive (excellent) correlation between same-day and weekly recall scores.

Table 1 (continued)

Walentyńowicz, 2018 ²⁹	Stress Frequency	General Population	469	M(SD)=2.409(1.167)	-	-	M(SD)=2.158(1.128)*	Weekly > Same day.
Walentyńowicz, 2018 ²⁹	Calm Frequency ^{iv}	General Population	469	M(SD) = 3.134(1.03)	-	-	M(SD)=3.277(1.133)	Same day > weekly but not sig. different.
Wood, 2015 ³⁸	Anxiety Severity	Post-hematopoietic cell transplantation	32	M(SD) = 0.53(0.72)	-	M(SD) = 0.73(0.87)***	-	Weekly < maximum daily.
Wood, 2015 ³⁸	Sadness/ Unhappy Feelings Severity	Post-hematopoietic cell transplantation	32	M(SD) = 0.39(0.73)	-	M(SD) = 0.61(0.91)***	-	Weekly < maximum daily.
Fatigue & Sleep-related problems (N=25)								
Armstrong, 2014 ³⁴	Fatigue Severity	Primary Brain Tumour	100	M(SD)=3.54(2.86)	-	-	M(SD)=3.43(2.74)	No sig. difference between weekly and same day.
Armstrong, 2014 ³⁴	Sleep Disturbance	Primary Brain Tumour	100	M(SD)=2.61(2.93)	-	-	M(SD)=2.10(2.56)	No sig. difference between weekly and same day.
Bennett, 2010 ²⁵	Tiredness Frequency	Cystic Fibrosis	38	M(95%CI)=0.95(0.61,1.28)	M(95%CI)=0.78(0.52,1.05)	M(95%CI)=1.47(1.09,1.84)	-	Mean daily < weekly < maximum daily.
Bennett, 2011 ³⁹	Tiredness Frequency	Type 2 Diabetes	139	M(SD)=3.48(2.52)	M(SD)=2.84(1.88)	-	-	Mean daily < weekly.
Bennett, 2018 ⁴⁰	Sleep/Energy	Nocturia	193	No values reported ⁱ	-	-	No values reported ⁱ	Negative correlation between weekly and daily symptoms. (r = -0.776)
Broderick, 2008 ⁴¹	Fatigue Interference & Severity	Rheumatological Illness	83	No values reported ⁱⁱ	No values reported ⁱⁱ	-	-	Mean daily < weekly.
Broderick, 2010 ⁴⁴	Fatigue Interference	Rheumatological Illness	87	No values reported ⁱⁱ	No values reported ⁱⁱ	-	-	Mean daily < weekly.
Broderick, 2013 ³³	Fatigue Severity ^v	General Population	98	M(SD)=48.80(9.60)	M(SD)=44.20(9.70)***	-	-	Mean daily < weekly.
Broderick, 2013 ³³	Fatigue Severity ^v	Osteoarthritis	98	M(SD)=56.20(7.80)	M(SD)=52.20(9.10)***	-	-	Mean daily < weekly.
De Andre Ares, 2015 ³⁵	Pain interference with sleep	Non-Cancer-related pain	698	No values reported ⁱ	-	-	No values reported ⁱ	Positive correlation between weekly and same day (ICC =0.946).
Mendoza, 2017 ²³	Fatigue Impact	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	Mean daily < weekly < maximum daily.
Mendoza, 2017 ²³	Fatigue Severity	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	Mean daily < weekly < maximum daily.
Mendoza, 2017 ²³	Insomnia Severity	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	Mean daily < weekly < maximum daily.
Schneider, 2013 ²⁷ (DIF)	Fatigue Frequency	General Population	100	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	DIF: Some non-systematic difference between mean of daily and weekly.
Shi, 2010 ³⁷	Fatigue Severity	Chemotherapy for cancer	42	No values reported ⁱ	-	-	No values reported ⁱ	No sig. difference between same day and weekly.
Stone, 2016 ²⁸	Fatigue	Breast Cancer	85	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly
Stone, 2016 ²⁸	Fatigue	General Population	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly
Stone, 2016 ²⁸	Fatigue	Osteoarthritis	98	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly
Stone, 2016 ²⁸	Fatigue	Premenstrual Syndrome	93	No values reported ⁱⁱⁱ	No values reported ⁱⁱⁱ	-	-	Mean daily < weekly
Wood, 2015 ³⁸	Fatigue Severity	Post-hematopoietic cell transplantation	32	M(SD)=1.42(0.91)	-	M(SD)=1.77(1.05), ES = 0.35	-	Weekly < maximum daily.
Wood, 2015 ³⁸	Insomnia Severity	Post-hematopoietic cell transplantation	32	M(SD)=1.15(0.91)	-	M(SD)=1.42(0.98), ES = 0.28	-	Weekly < maximum daily.

3.3.3 Cognition-Related Symptoms

Five studies compared one-day and seven-day recall on instruments assessing cognition-related symptoms, providing eight unique results for data extraction (see Table 1). For the three results using the mean daily recall indexation method, one found weekly recalled scores were higher than mean daily scores for concentration difficulties [39]

but the remaining two results (drawn from one study [27]) found no evidence of a significant difference. The three results using the same-day daily recall indexation method [27, 38] found no evidence of a significant difference between weekly and same-day recall scores for difficulties in remembering and understanding. For the two results using the maximum daily recall indexation method (both drawn from the same study [38]), one found weekly

Table 1 (continued)

Overall HRQoL (N=5)								
Bansback, 2008 ³⁰	Utility measured by EQ-5D vs HUI2 ^{iv}	Advanced HIV-Recently resolved SAE	57	-	-	-	Shown as regressions	Difference in assessment of an event being unresolved, when controlling for difference in the instruments: 0.18 (95%CI: 0.07-0.31). Weekly recall < Same day recall HRQoL for patients with recently resolved events.
Bansback, 2008 ³⁰	Utility measured by EQ-5D vs HUI2 ^{iv}	Advanced HIV-Recently resolved SAE	57	-	-	-	Shown as regressions	Difference in assessment of an event being unresolved, when controlling for difference in the instruments: 0.15 (95%CI: 0.05-0.26). Weekly recall < Same day recall HRQoL for patients with recently resolved events.
Thavarajah, 2013 ³¹	HRQoL. FACT-Brain ^{iv}	Brain Metastases	40	M(SD)= 154.80(28.10)	-	-	M(SD)= 155.00(27.40)	No sig. difference between weekly and same day.
Thavarajah, 2013 ³¹	FACT-General ^{iv}	Brain Metastases (FACT-General; higher score the better HRQoL)	40	M(SD)= 83.00(16.20)	-	-	M(SD)= 83.00(15.80)	No sig. difference between weekly and same day.
Topp, 2019 ³²	Utility measured by SF-6D ^{iv}	Multiple Sclerosis or psoriasis	100	M(SD)= 0.70(0.13)	M(SD)= 0.74(0.13)***	-	-	Weekly < mean daily.
Aggregated Disease-Specific Signs, Symptoms, and Impacts (N=12)								
Armstrong, 2014 ³⁴	Brain Tumour Symptom Severity	Primary Brain Tumour	100	M(SD) = 1.59(1.32)	-	-	M(SD) = 1.41(1.15)	No sig. difference between weekly and same day.
Bennett, 2010 ²⁵	Respiratory Symptom Severity	Cystic Fibrosis	38	M(95%CI) = 0.86(0.66,1.06)	M(95%CI) = 0.76 (0.57,0.95)**	M(95%CI) = 1.24 (1.01,1.47)***	-	Mean daily < weekly < maximum daily.
Bennett, 2012 ²¹	Respiratory Symptom Severity	COPD	98	M(SD) = 1.71(1.38)	M(SD) = 1.13(0.97)**	M(SD) = 1.95(1.34)**	M(SD) = 1.08(1.17)***	Same day < mean daily < weekly < maximum daily.
Bennett, 2018 ⁴⁰	Nocturia Symptom Severity	Nocturia	193	No values reported ⁱ	-	-	No values reported ⁱ	Correlation between weekly and (oppositely scored) daily (r = -0.78).
Bennett, 2018 ⁴⁰	Nocturia Symptom Bother/Concern	Nocturia	193	No values reported ⁱ	-	-	No values reported ⁱ	Correlation between weekly and (oppositely scored) daily (r = -0.64).
Mathias, 2016 ⁴³	Psoriasis Symptom Severity	Plaque Psoriasis	106	No values reported ⁱ	No values reported ⁱ	-	-	Correlation between mean of daily and weekly scores (r = 0.95 - 0.96).
Mendoza, 2017 ²³	Cancer Treatment-Related Toxicity Symptom Severity	Adverse Events during Cancer Trials	126	No values reported ⁱⁱⁱ	Difference (weekly-daily) = 0.12 (ES 0.14)	Difference (weekly-daily) = 0.20 (ES 0.20)	-	No sig. difference between mean daily and weekly. Weekly < maximum daily.
Thavarajah, 2013 ³¹	Brain Cancer Symptoms ^{iv}	Brain Metastases	40	M(SD) = 71.80(13.40)	-	-	M(SD) = 71.90(13.30)	No sig. difference between same day and weekly.

COPD chronic obstructive pulmonary disease, DIF Differential Item Functioning, ES Effect Size, FACT Functional Assessment of Cancer Therapy, HIV Human Immunodeficiency Virus, SAE serious adverse events, sig. significant

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$ for two-tailed t-tests of between-group differences

^aMean of one-day recall over same seven consecutive days as seven-day recall

^bMaximum one-day recall over same seven consecutive days as seven-day recall

^cOne-day recall on same day as seven-day recall

ⁱOnly correlations assessed

ⁱⁱResults presented in figure, not numerically

ⁱⁱⁱDifferences between recall conditions reported, no individual values reported for each condition

^{iv}Higher score is better HRQoL

^{iv}Means shown as a T-score in which general population mean = 50 (SD = 10). See Supplementary Information 'SII.Extraction Quantitative' for a full tabulation of instruments and methods

recalled scores were lower than maximum daily recalled concentration problems while the other found no evidence of a significant difference between weekly and maximum daily recalled memory problems.

3.3.4 Psychosocial Wellbeing

Thirteen studies provided 51 unique results comparing one-day and seven-day recall on instruments assessing aspects of psychosocial wellbeing (see Table 1). For the 22 results using the mean daily recall indexation method, the majority

(14) found weekly recalled scores were lower than mean daily recalled scores and eight found no evidence of a significant difference between weekly and mean daily recalled psychosocial symptom scores. All 10 results using the maximum daily recall indexation method, found weekly recall scores were lower than maximum recall scores for psychosocial symptoms. Majority (14 of 19) of results using the same-day daily recall indexation method found no evidence of a significant difference between weekly and same-day recalled psychosocial symptom scores, while five found weekly recalled scores were higher than same-day recall scores [21, 29, 34]. Three of the same-day to weekly comparisons involved items which were framed positively, two (happy, excited) followed the pattern of weekly scores being higher than the daily recall score, but the item asking about feeling ‘calm’ showed daily recall as greater than weekly, although all three differences were not significant.

3.3.5 Fatigue and Sleep-Related Symptoms

Thirteen studies provided 25 unique results comparing one-day and seven-day recall on instruments assessing sleep-related symptoms (see Table 1). For the 14 results comparing daily recall scores averaged over seven consecutive days with seven-day recall scores, majority (13) found weekly recall scores to be higher than mean daily recall scores. The single study using DIF to assess recall period effects identified non-systematic item-level differences between weekly and daily recalled fatigue frequency scores [27]. All six results comparing the maximum daily recall with weekly recall scores found maximum daily scores to be higher than weekly recall scores. No significant effect of recall period was found for the three results comparing the daily recall score with seven-day recall scores reported on the same day. Two studies assessed correlations between same-day and weekly recall scores: one identified a negative (good) correlation between same-day and (oppositely scored) weekly recalled sleep adequacy scores [40], while the other identified a positive (excellent) correlation between same-day and weekly recalled pain interference with sleep [35].

3.3.6 HRQoL Scores

Three studies provided five unique results comparing one-day and seven-day recall on instruments assessing HRQoL (see Table 7) [30–32]. The one study comparing mean daily recall scores (using the Short Form 6 Dimensions [SF-6D [49]] measure of utility) averaged over seven consecutive days with seven-day recall scores found that weekly recall HRQoL was significantly lower than mean daily recall scores [32]. Two studies comparing daily HRQoL scores assessed on the same day as seven-day HRQoL scores. In one study, controlling for non-recall instrument differences

(EQ-5D with a recall of ‘today’ vs Health Utilities Index 2 and 3 [HUI-2 and HUI-3 [50]] with a recall of last week), weekly recall score was less than daily recall in participants with advanced HIV where patients had an unresolved event during the week [30]. In the other study, no significant difference was identified in participants with brain metastases using Functional Assessment of Cancer Therapy Brain (FACT-Br), or the FACT-General with different recall periods.

3.3.7 Aggregate Measures of Disease-Specific Signs and Symptoms

Seven studies compared one-day and seven-day recall on instruments assessing aggregated disease-specific sign and symptom scores, providing 12 unique results for data extraction (see Table 1) [21, 23, 25, 31, 34, 40, 43]. For the four results using the mean daily recall indexation method [21, 23, 25, 43], two [21, 25] found that weekly recall scores were lower than mean daily scores, while one [23] found no evidence for a significant difference between weekly and mean daily recall scores. One result using a correlational approach identified an excellent positive association between weekly and mean of daily recall scores [43]. All three results using the maximum daily recall indexation method found that weekly scores were lower than maximum daily scores [21, 25]. For the five results using the same-day daily recall indexation method, two [23, 34] found no significant difference between mean and same-day recall scores, while one [21] found that same-day scores were less than weekly recall scores. Two results using a correlational approach identified a negative (moderate and good) association between an instrument using weekly recall and a different instrument, oppositely scored, using same-day recall scores [40].

3.4 Participant Recall Period Preferences

Of the 33 qualitative studies reviewed (see Table 2), 18 assessed disease-specific signs and symptoms, 9 assessed HRQoL, 5 assessed fatigue and sleep-related symptom, 3 assessed pain-related symptoms, and 1 assessed physical functioning. Most studies (55%, 18 of 33) used closed-ended probes to assess participant perceptions of the suitability of a designated recall period, while 45% (15 of 33) of studies used open-ended probes to elicit participant recall period preferences.

Of the 18 studies assessing questions on disease specific signs and symptoms 3 found that respondents expressed different preferences depending on context, with a preference for seven-day recall for symptom impact but one-day recall for symptom severity. The remaining 15 reported broadly equal preference for seven-day recall (8/15) as one-day recall (7/15).

Table 2 Recall period preferred by majority of participants in qualitative studies

	Clinical group	N	PROM: outcome assessed	Participant recall period preference (days) ^a		
				1	7	> 7
<i>Pain-related symptoms</i>						
Becker, 2021 [70]	Rheumatoid arthritis	32	Pain interference		†	
Chiarotto, 2018 [8]	Non-specific low back pain	207	Pain intensity		†	
White, 2021 [71]	Sickle cell disease	18	Sickle cell disease-related pain frequency and severity	†		
<i>Physical functioning</i>						
Leggett, 2016 [72]	Rheumatic diseases	70	Work productivity			•
<i>Fatigue and sleep-related symptoms</i>						
Becker, 2021 [70]	Rheumatoid arthritis	32	Sleep disturbance		†	
English, 2021 [62]	Menopause	32	Impact of menopause-associated vasomotor symptoms on sleep		†	
Matza, 2015 [69]	Major depression	98	Fatigue associated with Major Depression			•
Naegeli, 2013 [55]	Ankylosing spondylitis	13	Worst Fatigue	†		
Raymond, 2021 [73]	Systemic lupus erythematosus	15	Fatigue in systemic lupus erythematosus		†	
HRQoL						
Aronson, 2021 [74]	Hypersensitivity pneumonitis	10	Disease-specific HRQoL			•
Chassany, 2015 [52]	Irritable bowel syndrome	30	Disease-specific impact on HRQoL		†	
Ernstsson, 2020 [67]	Type 1 diabetes	20	General HRQoL	NA	NA	
Goswami, 2020 ⁵³	Haematological malignancy	34	Disease-specific HRQoL	†		
Hyland, 2018 [56]	Severe asthma	16	Disease-specific HRQoL			•
Miedany, 2014 [75]	Inflammatory arthritis	94	Disease-specific HRQoL		†	
Speck, 2019 [63]	Migraine	11	Disease-specific HRQoL			†
Trudeau, 2020 [58]P==	Myelodysplastic Syndromes	16	Disease-specific HRQoL			•
White, 2017 [76]	Amyloid light chain amyloidosis	10	General HRQoL			†
<i>Disease-specific signs, symptoms, and impacts</i>						
Abrams, 2018 [53]	Nocturia	28	Impact of night-time urination			•
Banderas, 2021 [66]	Rheumatoid arthritis	20	Rheumatoid arthritis symptom severity		†	
Chassany, 2015 [52]	Irritable bowel syndrome	30	Intestinal gas symptom severity		†	
Daly, 2021 [77]	Paroxysmal nocturnal hemoglobinuria	15	Paroxysmal nocturnal hemoglobinuria symptom severity		†	
Feldman, 2016 [54]	Psoriasis	19	Psoriasis symptom severity			•
Gabes, 2021 [65]	Atopic eczema	7	Atopic eczema symptom severity		†	
Goswami, 2020 [60]	Haematological malignancy	34	Signs and symptoms		†	
Hayes, 2015 [78]	Hypogonadism	125	Symptom severity (sex drive)		†	
			Symptom severity (energy)		†	
Lebwohl, 2014 [79]	Psoriasis	29	Psoriasis symptom severity and impacts		†	
Martin, 2013 [51]	Psoriasis	59	Psoriasis symptom severity (unstable)			•
			Psoriasis symptom severity (stable)			•
Martin, 2019 [80]	Primary biliary cholangitis	20	Itch symptom severity			•
Mathias, 2017 [81]	Systemic lupus erythematosus	41	Symptom severity and impacts			•
McCollister, 2016 [57]	Pulmonary arterial hypertension	20	Symptom severity			•
			Symptom impacts			•
Naegeli, 2015 [61]	Psoriasis, Psoriatic arthritis	34	Itch symptom severity		†	
Paty, 2017 [64]	Varicose veins	31	Varicose vein symptom frequency			•
Revicki, 2018 [82]	Gastroparesis	25	Gastroparesis symptom severity			•
Schildmann, 2015 [83]	Palliative Care	25	Palliative care symptom severity			•
Shi, 2010 [37]	Cancer	20	Chemotherapy symptom severity			•

^a•Participants nominated preferred recall period in response to open-ended probing

†Participant endorsed recall period suggested in closed-ended probing. See supplementary information ‘SI3.Extraction Qualitative’ for a full tabulation of instruments and methods. In Goswami 2020, most preferred recall for symptoms was 3-days

Two of the three studies assessing pain-related symptoms found a preference for a seven-day recall period. The single study assessing physical functioning via work productivity found a preference for a seven-day recall period. A majority of studies (80%, 4 of 5) assessing fatigue and sleep-related symptoms found a preference for a seven-day recall period. Of those included studies considering measurement of the impact on HRQoL, a longer time period was preferred, with more studies (3 out of 9) preferring seven-day recall than one-day (1 out of 9) and others preferring period greater than seven days (4 out of 9) or having no clear preference (1 out of 9).

A number of themes were identified in these studies, i) duration should capture important effects, ii) accuracy of recall, iii) preference for unambiguous language and iv) adherence to the stated recall period.

i) duration should capture important effects

The seven-day recall was considered more appropriate for measuring symptoms in subjects with relatively stable symptoms, while those with variable symptoms or undergoing treatment and expecting rapid change may need the shorter one-day recall period to accurately reflect changes in symptoms [51, 52]. Discussions indicated an assumption that one-day recall instruments would be repeatedly administered, with respondents raising the issue of burden of completing the questionnaire on consecutive days [53, 54].

Where single administration was implied, some participants favoured the longer time period, which could be more representative of their overall experience, "I just think you'll get a bigger picture by looking at it over a course of a week" [55]. In reference to varying asthma symptoms one participant said, "You have a chance at remembering how you felt on average, because you can have bad days and you can have good days" [56]. The seven-day recall was preferred by some participants for quality-of-life measurement because not all impacted activities occur on every day of the week [57]. Some participants also expressed concern that a seven-day recall might be too short, and not adequate to reflect their symptoms where impactful events occurred at intervals greater than one week [58, 59].

ii) accuracy of recall

Some participants acknowledged the ease of recalling over one-day "24 hours I can really, really remember how bad my itching was and you get more of a bam, to the point, to a real good timeframe" [61]. Others did not find the seven-day recall problematic. "I did not find any great difficulty [recalling the past 7 days]. At first, you have to put yourself back into the situation and look back at the 7 past days. It simply requires a few seconds to remember" [62]. Participants indicated recall accuracy as a concern only for recall periods greater than one week (e.g., 4 weeks [63]).

One participant expressed a preference for using one-day recall to measure quality of life due to daily activities and stressors potentially interfering with accurate memory – "I think using "today" is better, I had a hectic week last week, I went to a funeral, I had other things, I was a bit anxious" [60].

iii) preference for unambiguous language

Some participants indicated a preference to revise the 24-hour recall instruction to "since waking" to disqualify consideration of time while sleeping [64]. Weekly recall instructions were sometimes misinterpreted as the last previous full week (e.g., from Monday to Sunday) [65], or the 5-day working week [66]. Therefore, an explicit seven-day recall instruction was considered preferable to mitigate potential recall period misinterpretations [67].

iv) adherence to the stated recall period

Participants described processes that underpinned their interpretation of recall period instruction, including interpreting health "today" as meaning health generally [67]. Thus, participants reported overlooking temporary issues experienced on the day of reporting to provide a representative picture of their health state (not over the last 24 h per se) [8].

4 Discussion

This systematic review examined the effect of a one-day versus seven-day recall duration on PROM and HRQoL instrument scores in adults with a range of clinical conditions. Across the 24 quantitative studies identified, 158 unique results were identified. Overall, compared to the average symptoms reported with a 24-h recall over seven days, a seven-day recall mostly predicted worse symptoms and worse HRQoL for a range of clinical conditions.

Symptoms tended to be reported as more severe when assessed with a weekly recall than with a one-day recall averaged over the same period (76%, 58 of 76 results [two were only reported as correlation and not included in this total]); however, this difference was not statistically significant for 24% (18 of 76 results). This pattern was similar for comparisons based on the same-day reporting although a smaller percentage of results showed a significant difference 26% (12 of 46 [five were only reported as correlation and are not included in this total]). The weekly recall period tended to report lower symptom severity (i.e., better health) than the maximum of the daily score over the seven-day period 86% (25 out of 29 results), with the remaining 4 not finding a statistically significant difference.

The three findings on HRQoL instruments used to estimate utility scores [30, 32] suggest weekly recall period

leads to lower utility values than daily recall, particularly if negative events occurred during the previous seven days, which had been resolved.

The results reporting symptoms and HRQoL comparing mean of one-day recall across 7 days or the same day with the weekly recall 53% (35 of 66) find a one-day recall score that is at least 10% lower for symptoms or 10% higher for HRQoL (the green shading on Table 1) than the weekly recall score, and 89% (59 of 66) find one-day recall reporting lower problems or higher quality of life and only 6% (4 of 66) finding the opposite.

Within qualitative studies, participants identified four themes. First, ‘duration should accurately capture effects’ and preferred recall period varied depending upon the symptom and impact variability and the frequency of measurement. This aligns with findings in the review by Stull and colleagues [68] that there is no “one size fits all” ideal recall period. Second, ‘accuracy of recall’—although participants acknowledged the ease of the one-day recall they also had minimal concerns with accuracy of the seven-day recall. Third, participants expressed a ‘preference for unambiguous language’ when describing both recall periods. Finally, some participants noted a failure to ‘adhere to the recall period’ particularly for the framing of ‘today’, which they interpreted as health generally.

This review was intentionally limited in scope to a targeted comparison of a one-day versus seven-day recall period. Therefore, it does not consider longer recall periods that may be more suitable for chronic or variable conditions [56]. Information relevant to the understanding of recall duration effects may have been omitted through the exclusion of studies comparing other recall periods or symptoms reported using EMA. The PROSPERO-registered protocol was deviated from during the full-text screening to exclude studies using EMA to derive an index of daily recall scores, which was considered to not directly reflect one-day recall processes.

The review drew on different methods of exploring the impact of recall period, synthesising findings across many clinical conditions, different outcomes assessed, and different data collection and analysis techniques. The consistency of the findings amid this variability supports triangulation of our main findings.

4.1 Limitations of this Review

The search terms used did not exhaust all possible terms. For example, we did not include terms relating to ‘diaries’ which may have identified more one versus seven-day recall comparisons but would have reduced the precision of the search.

Other limitations of this review relate to the methodological flaws of included studies, such as inadequate control for the effect of repeated questionnaire assessments and the limited statistical power of between-group comparisons

made within small samples. Similarly, the few studies using a comparison of two different instruments for the one-day and seven-day recall periods is likely to have introduced measurement artifacts that may have confounded inferences regarding recall duration effects specifically. The qualitative studies reviewed were limited by closed-ended probing techniques, which may have restricted participant considerations of preferred recall duration.

Assessing the content validity of PROM and HRQoL instruments is inherently limited by the absence of a gold standard marker of patient experience against which recall period effects can be reliably distinguished. More broadly, the quantitative studies assessed in this review do not provide insight into the cognitive mechanisms and recall period actually utilised by participants when considering their health. Additionally, some studies reviewed suggest that people may reinterpret recall period instructions when responding, for example, interpreting ‘today’ as meaning health generally [67].

The potential for differences between seven-day versus one-day responses to arise due to selection effects based on when respondents are willing or able to complete questionnaires has not been well explored. If the last seven-day period includes days in which the respondent would not have engaged in questionnaire completion due to high level of symptoms (e.g., feeling depressed) this would generate the pattern found here for the same-day index comparison in which the seven-day recall reports poorer health levels. Similarly, if missing daily reports during the past seven days occur on days with relatively higher level of symptoms and comparisons are made on incomplete data, this would also generate the pattern found here for the mean of one-day recall versus seven-day recall comparison. Such selection effects may be particularly problematic for conditions effecting motivation such as mental health conditions.

4.2 Future Research

High-quality, sufficiently powered studies that account for repeated questionnaire administration are required to measure the effect of a one-day versus seven-day recall period in PROM and HRQoL instruments. Mixed methods study designs incorporating both quantitative comparison of scores and qualitative exploration of participant recall processing may confer insight into the cognitive mechanisms underpinning potential recall period effects. Of the 57 studies included in this review, only one study assessed recall duration effects in participants with a mental health condition (i.e., Major Depressive Disorder [69]). The absence of psychometric studies assessing the effect of recall duration for psychiatric symptoms and conditions could be addressed in future research.

This review identified few results which compared the recall period for positively framed items. The only

inclusions being from one study based on three items: happy, excited, and calm. Although the HRQoL instruments are scored positively (higher score shows better quality of life) they rely upon items reporting health problems using mostly negatively framed items. The results for the recall period on positive items, although not significantly different between recall period, are interesting in that items on feeling happy and excited suggest a higher score for weekly report, but not for calm. The interaction between item framing, arousal and recall period could usefully be explored in future research.

The variability in samples and instruments used in this review meant that results could not be pooled, and the magnitude of the impact of recall period remains uncertain. Of the 66 results reporting symptoms and HRQoL comparing mean of one-day recall across seven days or the same day with the weekly recall, the majority (89%) found that one-day recall showed fewer problems or a higher quality of life, although not all these individual findings showed a statistically significant difference. Whilst the direction of difference in recall period is clear, further research could usefully estimate the size of this recall effect more accurately.

5 Conclusion

This review identified a pattern of higher symptom scores and worse quality of life being reported for a seven-day compared to a one-day recall period on PROMs and HRQoL instruments. The review also identified anomalies in this pattern for two positively framed wellbeing items and a need for further research on recall effects in positively framed items. A better understanding of the impact of using different recall periods within PROMs and HRQoL instruments will help contextualise future comparisons between instruments which adopt different recall periods.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40271-022-00611-w>.

Declarations

Funding EuroQol Research Foundation Grant Number: EQ Project 224-2020RA.

Conflict of interest Tessa Peasgood and Clara Mukuria are members of EuroQol Group.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and materials The extraction of data from this review is available within the Supplementary Information.

Code availability Not applicable.

Author contribution statement Tessa Peasgood: Conceptualisation, Methodology, Investigation, Formal analysis, Writing—original draft, Writing—review and editing, Funding acquisition. Julia Caruana: Methodology, Investigation, Formal analysis, Writing—original draft Writing—review and editing. Clara Mukura: Conceptualization, Methodology, Investigation, Formal analysis, Writing—review and editing, Funding acquisition.

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


References

1. Sanghera S, Coast J. Measuring quality-adjusted life-years when health fluctuates. *Value Health*. 2020;23(3):343–50.
2. Norquist Josephine M, Girman C, Fehnel S, DeMuro-Mercon C, Santanello N. Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. *Qual Life Res*. 2012;21(6):1013–20.
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(4):929–42.
4. Cicely K, Emily JL, Charlotte EK, et al. Health-related quality of life in Parkinson's: impact of "off" time and stated treatment preferences. *Qual Life Res*. 2016;25(6):1505–15.
5. Reissmann DR. Methodological considerations when measuring oral health-related quality of life. *J Oral Rehabil*. 2020;23:23.
6. Easton RM, Bendinelli C, Sisak K, et al. Recalled pain scores are not reliable after acute trauma. *Injury*. 2012;43(7):1029–32.
7. Mendoza TR. New developments in the use of patient-reported outcomes in cancer patients undergoing immunotherapies. *Adv Exp Med Biol*. 2020;1244:335–9.
8. Chiarotto A, Boers M, Deyo RA, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain*. 2018;159(3):481–95.
9. Stone Arthur A, Broderick Joan E, Schwartz Joseph E, Schwarz N. Context effects in survey ratings of health, symptoms, and satisfaction. *Med Care*. 2008;46(7):662–7.
10. Bradburn NM. Temporal representation and event dating. The science of self-report. Psychology Press; 1999. pp. 61–74.
11. Kahneman D, Fredrickson BL, Schreiber CA, Redelmeier DA. When more pain is preferred to less: adding a better end. *Psychol Sci*. 1993;4(6):401–5.
12. Robinson MD, Clore GL. Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes. *J Pers Soc Psychol*. 2002;83(1):198.

13. Winkelman P, Knauper B, Schwarz N. Looking back at anger: reference periods change the interpretation of emotion frequency questions. *J Pers Soc Psychol.* 1998;3:719.
14. Thomas DL, Diener E. Memory accuracy in the recall of emotions. *J Pers Soc Psychol.* 1990;59(2):291.
15. Downie W, Leatham P, Rhind V, Wright V, Branco J, Anderson J. Studies with pain rating scales. *Ann Rheum Dis.* 1978;37(4):378–81.
16. Schmier JK, Halpern MT. Patient recall and recall bias of health state and health status. *Expert Rev Pharmacoecon Outcomes Res.* 2004;4(2):159.
17. Devlin NJ, Brooks R. EQ-5D and the EuroQol group: past, present and future. *Appl Health Econ Health Policy.* 2017;15(2):127–37.
18. Brazier J, Peasgood T, Mukuria C, et al. The EQ-HWB: overview of the development of a measure of health and wellbeing and key results. *Value Health.* 2022;25(4):482–91.
19. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation: Oxford University Press; 2017.
20. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *J Clin Epidemiol.* 2021;134:178–89.
21. Bennett AV, Amtmann D, Diehr P, Patrick DL. Comparison of 7-day recall and daily diary reports of COPD symptoms and impacts. *Value Health.* 2012;15(3):466–74.
22. Shiffman S, Stone AA, Fau-Hufford MR, Hufford MR. Ecological momentary assessment (1548-5943 (Print)).
23. Mendoza TR, Dueck AC, Bennett AV, et al. Evaluation of different recall periods for the US National Cancer Institute's PRO-CTCAE. *Clin Trials.* 2017;14(3):255–63.
24. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res.* 2018;27(5):1159–70.
25. Bennett AV, Patrick DL, Lymp JF, Edwards TC, Goss CH. Comparison of 7-day and repeated 24-hour recall of symptoms of cystic fibrosis. *J Cyst Fibros.* 2010;9(6):419–24.
26. Condon DM, Chapman R, Shaunfield S, et al. Does recall period matter? Comparing PROMIS physical function with no recall, 24-hr recall, and 7-day recall. *Qual Life Res.* 2020;29(3):745–53.
27. Schneider S, Choi SW, Junghaenel DU, Schwartz JE, Stone AA. Psychometric characteristics of daily diaries for the Patient-Reported Outcomes Measurement Information System (PROMIS): a preliminary investigation. *Qual Life Res.* 2013;22(7):1859–69.
28. 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. *J Clin Epidemiol.* 2016;74:194–206.
29. Walentynowicz M, Schneider S, Stone AA. The effects of time frames on self-report. *PLoS ONE.* 2018;13(8):1–18.
30. Bansback N, Sun H, Guh DP, et al. Impact of the recall period on measuring health utilities for acute events. *Health Econ.* 2008;17(12):1413–9.
31. Thavarajah N, Bedard G, Zhang L, et al. The Functional Assessment of Cancer Therapy-Brain (FACT-Br) for assessing quality of life in patients with brain metastases: a comparison of recall periods. *J Pain Manag.* 2013;6(3):223–34.
32. Topp J, Andrees V, Heesen C, Augustin M, Blome C. Recall of health-related quality of life: How does memory affect the SF-6D in patients with psoriasis or multiple sclerosis? A prospective observational study in Germany. *BMJ Open.* 2019;9(11).
33. Broderick JE, Schneider S, Junghaenel DU, Schwartz JE, Stone AA. Validity and reliability of patient-reported outcomes measurement information system instruments in osteoarthritis. *Arthritis Care Res.* 2013;65(10):1625–33.
34. Armstrong TS, Vera-Bolanos E, Acquaye A, Gilbert MR, Mendoza TR. Impact of recall period on primary brain tumor patient's self-report of symptoms. *Neuro-Oncol Pract.* 2014;1(2):55–63.
35. de Andres AJ, Cruces Prado LM, CanosVerdecho MA, et al. Validation of the Short Form of the Brief Pain Inventory (BPI-SF) in Spanish patients with non-cancer-related pain. *Pain Pract.* 2015;15(7):643–53.
36. Kamper SJ, Grootjans SJM, Michaleff ZA, Maher CG, McAuley JH, Sterling M. Measuring pain intensity in patients with neck pain: does it matter how you do it? *Pain Pract.* 2015;15(2):159–67.
37. Shi Q, Trask PC, Wang XS, et al. Does recall period have an effect on cancer patients' ratings of the severity of multiple symptoms? *J Pain Symptom Manag.* 2010;40(2):191–9.
38. Wood WA, Deal AM, Bennett AV, et al. Comparison of seven-day and repeated 24-hour recall of symptoms in the first 100 days after hematopoietic cell transplantation. *J Pain Symptom Manag.* 2015;49(3):513–20.
39. Bennett A, Patrick D, Bushnell D, Chiou C, Diehr P. Comparison of 7-day and repeated 24-h recall of type 2 diabetes. *Qual Life Res.* 2011;20(5):769–77.
40. Bennett JB, Gillard KK, Banderas B, Abrams S, Cheng L, Fein S. Development of a new patient-reported outcome (PRO) measure on the Impact of Nighttime Urination (INTU) in patients with nocturia-Psychometric validation. *Neurourol Urodyn.* 2018;37(5):1678–85.
41. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. *Pain.* 2008;139(1):146–57.
42. Lackner JM, Jaccard J, Keefer L, et al. The accuracy of patient-reported measures for GI symptoms: a comparison of real time and retrospective reports. *Neurogastroenterol Motil.* 2014;26(12):1802–11.
43. Mathias SD, Feldman SR, Crosby RD, Colwell HH, McQuarrie K, Han C. Measurement properties of a patient-reported outcome measure assessing psoriasis severity: the psoriasis symptoms and signs diary. *J Dermatol Treat.* 2016;27(4):322–7.
44. Broderick JE, Schneider S, Schwartz JE, Stone AA. Interference with activities due to pain and fatigue: accuracy of ratings across different reporting periods. *Qual Life Res.* 2010;19(8):1163–70.
45. Marty M, Rozenberg S, Legout V, et al. Influence of time, activities, and memory on the assessment of chronic low back pain intensity. *Spine.* 2009;34(15):1604–9.
46. Schaffer EM, Basch EM, Schwab GM, Bennett AV. Comparison of weekly and daily recall of pain as an endpoint in a randomized phase 3 trial of cabozantinib for metastatic castration-resistant prostate cancer. *Clin Trials.* 2021;18(4):408–16.
47. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–63.
48. Gabes M, Tischer C, Herrmann A, Howells L, Apfelbacher C. The German RECAP questionnaire: linguistic validation and cognitive debriefing in German adults with self-reported atopic eczema and parents of affected children. *J Patient-Rep Outcomes.* 2021;5(1).
49. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ.* 2002;21(2):271–92.
50. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health Qual Life Outcomes.* 2003;1(1):1–13.
51. Martin ML, McCarrier KP, Chiou CF, et al. Early development and qualitative evidence of content validity for the Psoriasis

- Symptom Inventory (PSI), a patient-reported outcome measure of psoriasis symptom severity. *J Dermatol Treat.* 2013;24(4):255–60.
52. Chassany O, Tugaut B, Marrel A, et al. The Intestinal Gas Questionnaire: development of a new instrument for measuring gas-related symptoms and their impact on daily life. *Neurogastroenterol Motil.* 2015;27(6):885–98.
 53. Abrams S, Martin S, Gillard KK, Cheng L, Fein S. Development of the Impact of Nighttime Urination (INTU) questionnaire to assess the impact of nocturia on health and functioning. *NeuroUrol Urodyn.* 2018;37(5):1686–92.
 54. Feldman SR, Mathias SD, Schenkel B, et al. Development of a patient-reported outcome questionnaire for use in adults with moderate-to-severe plaque psoriasis: The Psoriasis Symptoms and Signs Diary. *J Dermatol Dermatol Surg.* 2016;20(1):19–26.
 55. Naegeli AN, Flood E, Tucker J, Devlen J, Edson-Heredia E. The patient experience with fatigue and content validity of a measure to assess fatigue severity: qualitative research in patients with ankylosing spondylitis (AS). *Health Qual Life Outcomes.* 2013;11(1).
 56. Hyland ME, Lanario JW, Pooler J, Masoli M, Jones RC. How patient participation was used to develop a questionnaire that is fit for purpose for assessing quality of life in severe asthma. *Health Qual Life Outcomes.* 2018;16(1).
 57. McCollister D, Shaffer S, Badesch DB, et al. Development of the Pulmonary Arterial Hypertension-Symptoms and Impact (PAH-SYMPACT) questionnaire: a new patient-reported outcome instrument for PAH. *Respir Res.* 2016;17(1).
 58. Trudeau JJ, He J, Rose E, Panter C, Randhawa S, Gater A. Content validity of patient-reported outcomes for use in lower-risk myelodysplastic syndromes. *J Patient Rep Outcomes.* 2020;4(1):69.
 59. Aronson KI, Ali M, Reshetynak E, et al. Establishing content validity of a disease-specific health-related quality of life instrument for patients with chronic hypersensitivity pneumonitis. *J Patient Rep Outcomes.* 2021;5(1).
 60. Goswami P, Oliva EN, Ionova T, et al. Development of a novel hematological malignancy specific patient-reported outcome measure (HM-PRO): content validity. *Front Pharmacol.* 2020;11 (no pagination).
 61. Naegeli AN, Flood E, Tucker J, Devlen J, Edson-Heredia E. The Worst Itch Numeric Rating Scale for patients with moderate to severe plaque psoriasis or psoriatic arthritis. *Int J Dermatol.* 2015;54(6):715–22.
 62. English M, Stoykova B, Slota C, et al. Qualitative study: burden of menopause-associated vasomotor symptoms (VMS) and validation of PROMIS Sleep Disturbance and Sleep-Related Impairment measures for assessment of VMS impact on sleep. *J Patient Rep Outcomes.* 2021;5(1).
 63. Speck RM, Shalhoub H, Ayer DW, Ford JH, Wyrwich KW, Bush EN. Content validity of the Migraine-Specific Quality of Life Questionnaire version 2.1 electronic patient-reported outcome. *J Patient Rep Outcomes.* 2019;3(1):39.
 64. Paty J, Elash CA, Turner-Bowker DM. Content validity for the VVSymQ instrument: a new patient-reported outcome measure for the assessment of varicose veins symptoms. *Patient.* 2017;10(1):51–63.
 65. Gabes M, Tischer C, Herrmann A, Howells L, Apfelbacher C. The German RECAP questionnaire: linguistic validation and cognitive debriefing in German adults with self-reported atopic eczema and parents of affected children. *J Patient Rep Outcomes.* 2021;5(1):13.
 66. Banderas B, Skup M, Shields AL, Mazar I, Ganguli A. Development of the Rheumatoid Arthritis Symptom Questionnaire (RASQ): a patient reported outcome scale for measuring symptoms of rheumatoid arthritis. *Curr Med Res Opin.* 2017;33(9):1643–51.
 67. Ernstsson O, Burstrom K, Heintz E, Molsted Alvesson H. Reporting and valuing one's own health: a think aloud study using EQ-5D-5L, EQ VAS and a time trade-off question among patients with a chronic condition. *Health Qual Life Outcomes.* 2020;18(1).
 68. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. 2009. pp. 929–42.
 69. Matza LS, Murray LT, Phillips GA, et al. Qualitative research on fatigue associated with depression: content validity of the Fatigue Associated with Depression Questionnaire (FASD-V2). *Patient.* 2015;8(5):433–43.
 70. Becker B, Raymond K, Hawkes C, et al. Qualitative and psychometric approaches to evaluate the PROMIS pain interference and sleep disturbance item banks for use in patients with rheumatoid arthritis. *J Patient Rep Outcomes.* 2021;5(1).
 71. White MK, Saucier C, Bailey M, et al. Content validation of a self-report daily diary in patients with sickle cell disease. *J Patient Rep Outcomes.* 2021;5(1).
 72. Leggett S, Van Der Zee-Neuen A, Boonen A, et al. Content validity of global measures for at-work productivity in patients with rheumatic diseases: an international qualitative study. *Rheumatology (United Kingdom).* 2016;55(8):1364–73.
 73. Raymond K, Park J, Joshi AV, White MK. Patient experience with fatigue and qualitative interview-based evidence of content validation of The FACIT-Fatigue in Systemic Lupus Erythematosus. *Rheumatol Ther.* 2021;8(1):541.
 74. Aronson KI, Ali M, Reshetynak E, et al. Establishing content validity of a disease-specific health-related quality of life instrument for patients with chronic hypersensitivity pneumonitis. *J Patient Rep Outcomes.* 2021;5(1):9.
 75. Miedany YE, Gaafary ME, Youssef S, Ahmed I, Palmer D. The arthritic patients' perspective of measuring treatment efficacy: Patient Reported Experience Measures (PREMs) as a quality tool. *Clin Exp Rheumatol.* 2014;32(4):547–52.
 76. White MK, Bayliss MS, Guthrie SD, Raymond KP, Rizio AA, McCausland KL. Content validation of the SF-36v2 R health survey with AL amyloidosis patients. *J Patient Rep Outcomes.* 2017;1(1):13.
 77. Daly RP, Jalbert JJ, Keith S, Symonds T, Shammo J. A novel patient-reported outcome instrument assessing the symptoms of paroxysmal nocturnal hemoglobinuria, the PNH-SQ. *J Patient Rep Outcomes.* 2021;5(1).
 78. Hayes RP, Henne J, Kinchen KS. Establishing the content validity of the Sexual Arousal, Interest, and Drive Scale and the Hypogonadism Energy Diary. *Int J Clin Pract.* 2015;69(4):454–65.
 79. Lebowohl M, Swensen AR, Nyirady J, Kim E, Gwaltney CJ, Strober BE. The Psoriasis Symptom Diary: Development and content validity of a novel patient-reported outcome instrument. *Int J Dermatol.* 2014;53(6):714–22.
 80. Martin ML, Stassek L, Blum SI, Joshi AV, Jones D. Development and adaptation of patient-reported outcome measures for patients who experience itch associated with primary biliary cholangitis. *J Patient Rep Outcomes.* 2019;3(1):2.
 81. Mathias SD, Berry P, De Vries J, et al. Patient experience in systemic lupus erythematosus: development of novel patient-reported symptom and patient-reported impact measures. *J Patient Rep Outcomes.* 2017;2(1):11.
 82. Revicki DA, Lavoie S, Speck RM, et al. The content validity of the ANMS GCSI-DD in patients with idiopathic or diabetic gastroparesis. *J Patient Rep Outcomes.* 2018;2(1):61.
 83. Schildmann EK, Groeneveld EI, Denzel J, et al. Discovering the hidden benefits of cognitive interviewing in two languages: The first phase of a validation study of the Integrated Palliative care Outcome Scale. *Palliat Med.* 2015;30(6):599–610.

Authors and Affiliations

Tessa Peasgood¹  · Julia M. Caruana¹ · Clara Mukuria²

✉ Tessa Peasgood
Tessa.Peasgood@unimelb.edu.au

² School of Health and Related Research, The University of Sheffield, Sheffield, UK

¹ Center for Health Policy, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC, Australia