

# BMJ Open Quality and reporting of patient-reported outcomes in elderly patients with hip fracture: a systematic review

Puck van der Vet ,<sup>1</sup> Sandra Wilson,<sup>2</sup> R Marijn Houwert,<sup>1</sup> Egbert-Jan Verleisdonk,<sup>3</sup> Marilyn Heng <sup>2</sup>

**To cite:** van der Vet P, Wilson S, Houwert RM, *et al*. Quality and reporting of patient-reported outcomes in elderly patients with hip fracture: a systematic review. *BMJ Open* 2022;**12**:e058197. doi:10.1136/bmjopen-2021-058197

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-058197>).

Received 12 October 2021  
Accepted 10 November 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Department of Trauma Surgery, University Medical Centre Utrecht, Utrecht, Netherlands

<sup>2</sup>Department of Orthopaedic Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA

<sup>3</sup>Department of Surgery, Diaconessenhuis, Zeist, Netherlands

## Correspondence to

Puck van der Vet;  
puck\_vandervet94@hotmail.com

## ABSTRACT

**Objective** To assess how patient-reported outcomes (PROs) are reported and to assess the quality of reporting PROs for elderly patients with a hip fracture in both randomised controlled trials (RCTs) and observational studies.

**Design** Systematic review.

**Data sources** Medline, Embase and CENTRAL were searched on 1 March 2013 to 25 May 2021.

**Eligibility criteria** RCTs and observational studies on geriatric ( $\geq 65$  years of age) patients, with one or more PRO as outcome were included.

**Data extraction and synthesis** Primary outcome was type of PRO; secondary outcome and quality assessment was measured by adherence to the Consolidated Standards of Reporting Trials (CONSORT) extension for patient-reported outcomes (CONSORT-PRO). Because of heterogeneity in study population and outcomes, data pooling was not possible.

**Results** 3659 studies were found in the initial search. Of those, 67 were included in the final analysis. 83.6% of studies did not adequately mention missing data, 52.3% did not correctly report how PROs were collected and 61.2% did not report adequate effect size. PRO limitations were adequately reported in 20.9% of studies and interpretation of PROs was adequately reported in 19.4% of studies. Most Quality of Life (QoL) outcomes were measured by the EuroQol 5-Dimension 3-Levels, and pain as well as patient satisfaction by Visual Analogue Scale.

**Conclusion** This study found that a high variety of PRO measures are used to evaluate geriatric hip fracture care. In addition, 47.8% of studies examining PROs in elderly patients with hip fracture do not satisfy at least 50% of the CONSORT-PRO criteria. This enables poorly conducted research to be published and used in evidence-based medicine and, consequently, shared decision-making. More efforts should be undertaken to improve adequate reporting. We believe extending the CONSORT-PRO extension to Strengthening the Reporting of Observational Studies in Epidemiology for observational studies would be a valuable addition to current guidelines.

## INTRODUCTION

Hip fractures are a major health problem, affecting approximately 18% of women and 6% of men.<sup>1</sup> Mortality and morbidity rates after hip fractures are high and often cause

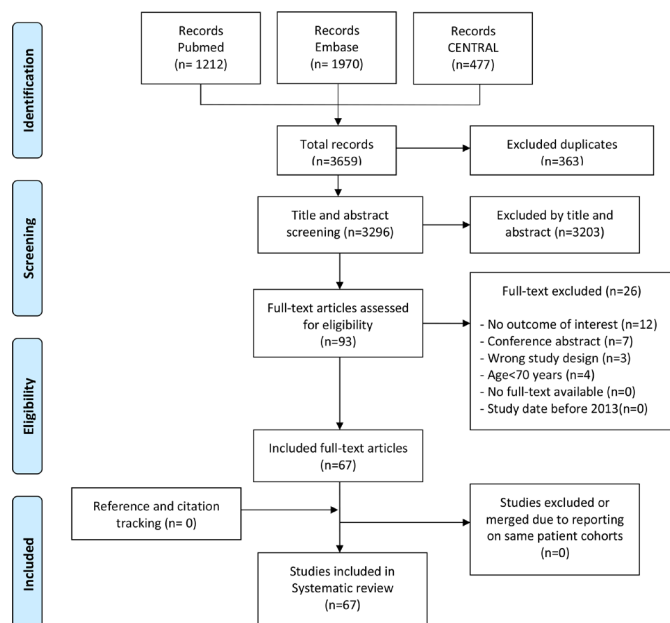
## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ To our knowledge, this is the first systematic review on the reporting quality of patient-reported outcomes in geriatric hip fracture care.
- ⇒ This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.
- ⇒ A limitation is that we did not perform a meta-analysis because we could not pool data.
- ⇒ Another limitation is that we did not include study protocols or non-English studies, which may have caused bias.

loss of independence in elderly patients. The estimated cost of hip fractures in the USA was 17 billion in 2002.<sup>2</sup> In addition, due to the ageing population, the incidence of hip fractures continues to increase and is estimated to be 4.5 million in 2050.<sup>3</sup> Consequently, hip fractures are considered one of the biggest challenges of future healthcare.

Traditionally, research on patients with geriatric hip fracture tended to focus on objective, clinician-reported outcomes measures, such as mortality, complication rate and hospital length of stay. However, recent literature stresses the importance of patient-reported outcomes (PROs) to evaluate treatment, estimate cost-effectiveness and improve clinical decision-making and patient-centred care.<sup>4</sup>

A PRO is generally defined as an outcome of a patient's health condition or health behaviour that is reported directly by the patient, without interpretation of the patient's response by a clinician or anyone else.<sup>5 6</sup> Patient-reported outcomes measures (PROMs) are instruments to measure these outcomes. Earlier literature on PROs in elderly patients with a hip fracture shows that a hip fracture negatively affects PROs such as Quality of Life (QoL), and that Fear of Falling is increased after a hip fracture, thereby increasing the risk of a secondary fracture.<sup>7</sup> PROs are frequently used in recent hip fracture studies researching different types of treatment,



**Figure 1** Preferred Reporting Items for Systematic Reviews and Meta-Analysis flow diagram presenting the search and selection of studies evaluating PRO(M) in elderly patients with hip fracture.

or studies investigating the potential benefit of comanaged care pathways (ie, orthogeriatric care pathways).<sup>8,9</sup>

While there are already many instruments and questionnaires to measure PROs, and though more measures are currently being developed, there is little to no guidance as to which measure should be used and how PROs should be analysed and interpreted.<sup>10</sup> Earlier studies showed that the quality of reporting PROs in randomised controlled trials (RCTs) was low and that PROs are prone to risk of bias.<sup>11</sup> Therefore, in 2013, the Consolidated Standards of Reporting Trials (CONSORT) extension for patient-reported outcomes (CONSORT-PRO) was published for RCTs to improve quality of reporting.<sup>12</sup> So far, no such measurement exists for harbouring the quality of reporting PROs in observational studies. The objective of this study was to assess how PROs are reported and to assess the quality of reporting PROs for elderly patients with a hip fracture in RCTs and observational studies. We hypothesised that most RCTs would not adhere to the CONSORT-PRO extension and observational studies would have low quality of reporting PROs for elderly patients with a hip fracture.

## METHODS

This systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement.<sup>13</sup> There is no published protocol for this review.

### Search strategy and study selection

We conducted a literature search in PubMed/Medline, CENTRAL and Embase for PRO(M) studies on elderly

patients (above 65 years of age) with a hip fracture, published after 27 February 2013 (date of the publication of the CONSORT-PRO extension). The last search was conducted on 25 May 2021. Online supplemental table A shows the search syntax. After removing duplicates, two independent reviewers (SW and PV) screened titles and abstracts for eligible studies. Eligible studies were RCTs or observational studies on elderly patients with a hip fracture, with at least one PRO as an outcome measure. Outcome measures such as the Charnley score, which can be measured by both clinician and patient, were included only if reported by the patient and if it was described in the article as PRO.

Hereafter, full texts were screened by the same two reviewers. Inclusion criteria were studies on patients over 65 years of age, who were treated (non)-surgically for a hip fracture, and the study had to report one or more PRO. Exclusion criteria were non-English studies, no available full text, letters, case reports, reviews, conference papers and study protocols. Disagreements on eligibility and inclusion were resolved through discussion and, if needed, consultation of a third researcher. References of the included studies were screened for studies that were not found in the original literature search.

### Data extraction

Data were extracted by two independent reviewers (SW and PV). Methodological quality was assessed using the Methodological Index for Non-Randomized Studies (MINORS) criteria for observational studies and the CONSORT 2010 PRO extension tool for randomised trials. MINORS is a 12-item list that corresponds with a scale ranging from 0 to 24 for comparative studies and 0–16 for non-comparative studies (the last 4 criteria are not applicable to non-comparative studies).<sup>14</sup> Disagreements were resolved through discussion and consensus. Online supplemental table B shows details of the critical appraisal.

### Patient and public involvement

No patients and/or public were involved in the design and conduct of this research.

### Primary and secondary outcomes

Primary outcomes included the type of PROM reported and whether it was a primary or a secondary endpoint in the study. Secondary outcome was adherence to the CONSORT-PRO extension. The CONSORT-PRO extension adds five PRO extension items to the CONSORT-2010 Statement and provides PRO elaborations to nine CONSORT criteria, which results in 14 criteria. We based this scoring list on an earlier conducted study on the uptake of CONSORT-PRO in all types of studies.<sup>15</sup> Two independent reviewers (PV and SW) scored the adherence to the CONSORT-PRO criteria. For the purpose of this analysis, we scored the criteria ‘2’ if it was adequately described (by which was meant that a criterium (for instance missing data) was mentioned in the article and that the article involved an explanation of how the criterium was assessed), ‘1’ when it was inadequately

**Table 1** Characteristics of journals with the highest number of included studies and highest impact factors

Journal	Impact factor 2020	No. of included studies	CONSORT/STROBE endorsement	CONSORT-PRO score mean (range)
<i>Age and Ageing</i>	10.668	2	Yes	26
<i>The Journals of Gerontology: Series A: Biol Sci Med Sci</i>	6.053	2	Yes	22.5 (22–23)
<i>Journal of American Geriatric Society</i>	5.562	1	Yes	18
<i>The Journal of Bone and Joint Surgery</i>	5.284	2	Yes	23.5 (23–24)
<i>Anaesthesia and Analgesia</i>	5.108	1	Yes	10
<i>The Bone &amp; Joint Journal</i>	5.082	5	Yes	20 (19–21)
<i>Journal of the American Medical Directors Association</i>	4.669	2	No	16.5 (11–22)
<i>Osteoporosis International</i>	4.507	2	No	14.5 (11–18)
<i>CORR</i>	4.176	2	Yes	24
<i>Quality of Life Research</i>	4.147	1	Yes	22
<i>Archives of Physical Medicine and Rehabilitation</i>	3.966	1	Yes	25
<i>BMC Geriatrics</i>	3.921	2	Yes	22
<i>Acta Orthopaedica</i>	3.717	2	Yes	17 (16–18)
<i>European Journal of Trauma and Emergency Surgery</i>	3.693	1	Yes	13
<i>Clinical Rehabilitation</i>	3.477	3	Yes	23
<i>PLoS One</i>	3.240	1	Yes	26
<i>International Orthopaedics</i>	3.075	1	Yes	24
<i>Archives of Orthopaedic and Trauma Surgery</i>	3.067	1	Yes	18
<i>BMJ Open</i>	2.692	1	Yes	20
<i>Injury</i>	2.586	8	Yes	15 (9–21)
<i>Journal of Orthopaedic Trauma</i>	2.512	1	Yes	19
<i>BMC Musculoskeletal Disorders</i>	2.362	4	Yes	15.5 (12–19)

Impact factor retracted from *Journal of Citation Reports*.  
 CONSORT, Consolidated Standards of Reporting Trials; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; .

described (by which is meant that a criterium (for instance missing data) was mentioned, but an explanation of how missing data was handled was missing), and ‘0’ when it was not reported/missing, thus resulting in a score ranging from 0 to 28. Disagreements were resolved through discussion and consensus with a third reviewer. In accordance with the previously mentioned article, adherence was deemed ‘good’ when >80% of criteria were met, ‘moderate’ when 50%–79% of criteria were met, and ‘poor’ when <49% of criteria were met. Before scoring, the reviewers discussed the CONSORT-PRO extension to ensure uniformity. A preformed data sheet was used for data extraction, including basic study characteristics: first author, date of publication, journal of publication, country of study, intervention, sample size and follow-up period. Journals were screened on their websites and their ‘instruction for authors’ page on CONSORT/STROBE endorsement. Journals’ impact factor was recorded through the 2020 Web of Science Journal Citation Report.

### Statistical analysis

Statistical analyses were carried out using SPSS Statistics V.27 (IBM Corporation). Results were evaluated descriptively, using total numbers and percentages for categorical variables. Mean values (M) along with the standard deviation (SD) were used for numeric, normally distributed data. The Shapiro-Wilks test was used to assess normal distribution. We

used the  $\chi^2$ -test to assess differences in categorical values. Since this systematic review contained little missing data (<5%), and the data were missing completely at random, we performed complete case analysis. Multivariable logistic regression analysis was used to observe the relation between (in)adequate reporting of CONSORT-PRO criteria and the study characteristics. To correct for possible overfitting, we included only those variables that were notable at an alpha level of 0.10 in the bivariate analyses. The models standardly included study design and journal. P values with a significance level of 0.05 were considered statistically significant. A subgroup analysis was performed according to study design to assess differences between experimental and observational research.

### RESULTS Search

A flow chart of the literature search and study inclusion is shown in [figure 1](#). After title and abstract screening, 93 studies were assessed for eligibility, of which 67 studies were included in the final analysis. Forty-two were RCTs and 25 were observational cohort studies. A total of 44 journals were included, with impact factors ranging from 0.471 to 10.668 ([table 1](#)).

**Table 2** Characteristics of included studies

	Number of studies (n=67), n (%)
Type of study	
RCT	35 (52.2)
Follow-up RCT	5 (7.5)
Prospective observational cohort	7 (10.4)
Retrospective observational cohort	19 (28.4)
Follow-up observational	1 (1.5)
Mentioned PRO as	
Primary outcome	33 (50.0)
Secondary outcome	33 (50.0)
Mean follow-up	
1–6 months	15 (25.0)
7–12 months	23 (38.3)
1–2 years	12 (20.1)
2–12 years	10 (16.8)
Number of centres	
Single centre	40 (65.6)
Multicentre	21 (34.4)
Mentioned CONSORT/STROBE adherence in article	10 (15.4)
Funding	
No influence of funding on results	43 (68.3)
Possible influence of funding on results	20 (31.7)

Missing values were excluded from analysis.  
CONSORT, Consolidated Standards of Reporting Trials; RCT, randomised controlled trial; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

### Study characteristics

The 67 included studies analysed a total of 53 931 patients, of which 15 124 were men. Weighted mean age was  $83.90 \pm 1.98$  (range 71.0–99). Follow-up ranged from 0 to 144 months. Study characteristics are displayed in

table 2 and detailed study characteristics are shown in online supplemental table C. Most studies originated from Scandinavia, Europe and China. Figure 2 displays demographic distribution.

### PRO reporting

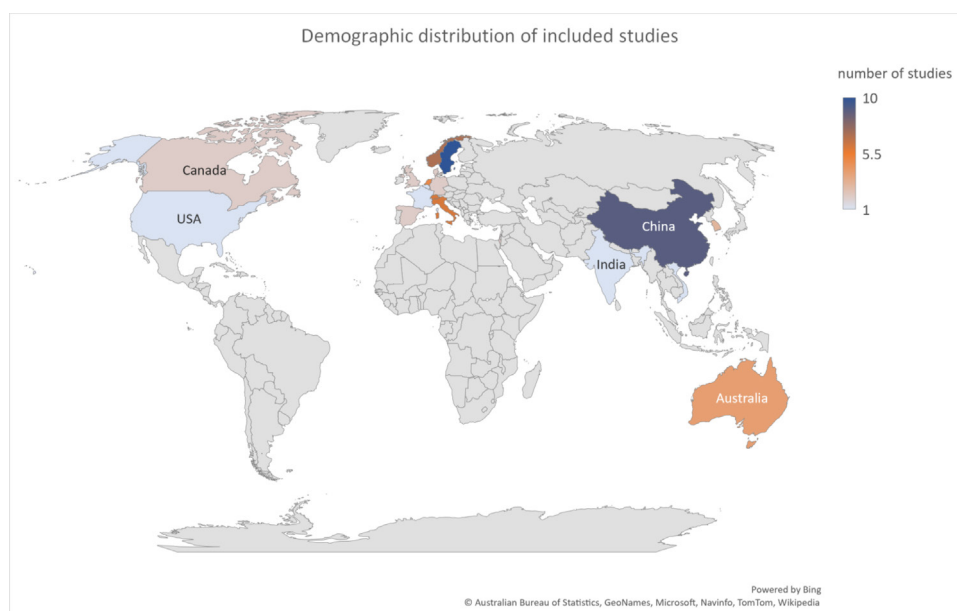
In total, 48 studies reported QoL. Of those, 32 (66.7%) used the EQ-5D-3L as measurement. Twenty-one studies included pain as an outcome measurement. Of those 21 studies, 13 (61.9%) used a VAS as measurement and only one study used a validated pain scale (the PAINAD). Patient-reported functional outcomes were measured using different outcome measurements, most of which were not PROs (ie, Harris Hip score). ADLs were most commonly measured using the Barthel index (45.8%). Details on PROM reporting are shown in table 3.

### Consolidated Standards of Reporting Trials extension for Patient-Reported Outcomes

A total of 10 articles (15.4%) mentioned adherence to the CONSORT/STROBE statement. There were no studies that mentioned the CONSORT-PRO, or any other PRO-reporting guideline. Figure 3, table 4 and online supplemental table D show the adherence to the CONSORT-PRO criteria. Overall reporting mean was  $16.6 \pm 5.8$  (range 1–26).

Thirty (44.8%) studies adequately described the PRO in their abstract and validity and reliability were cited in 76.1% of studies. Adequate reporting was also noted on criteria C15 (baseline outcomes, 76.1%) and C16 (denominator, 83.6%).

Inadequate reporting on PRO hypothesis was noted in 85.1% of studies (n=57). Statistical approaches to missing data were not adequately reported in 83.6% of studies (n=56). 61.2% of studies did not adequately report an effect size or a precision measure, 80.6% did



**Figure 2** Demographic distribution of the included studies. Count: number of studies from each country. Light blue=1, orange=4, 5, dark blue=8.

**Table 3** PRO(M) outcomes

Type of PRO(M)	No. of studies
Quality of life measurement (n=48)	
EuroQoL 5-Dimension 3-Level (EQ-5D-3L)	32
Short Form-36 (SF-36)	7
Short Form-12 (SF-12)	4
Barthel index	1
DEMqoL	1
Depression list	1
AQoL-8D	1
Qualeffo 41	1
Pain (n=21)	
Visual Analogue Scale (VAS)	13
Numeric Rating Scale (NRS)	6
Ordinal scale (4-point)	1
PAINAD	1
Functional outcomes (n=5)	
Oxford hip score	1
Charnley score	1
Falls Efficacy Scale-International (FES-I)	1
SF-36	1
WOMAC	1
Patient satisfaction (n=9)	
VAS	4
Binary	2
Ordinal scale (3-point, 4-point, 5-point, 10-point)	3
Activities of daily living, ADL (n=24)	
KATZ	6
Barthel index	11
Nottingham	2
Ordinal scale (3-point)	1
NHLSD	1
FIM	2
Instrumental ADL	4
Canadian Occupational Performance Measure	1

Numbers are noted in total numbers of studies that reported on a certain PRO, some studies used multiple measurement for 1 PRO (ie, both KATZ and Nottingham for ADL).

AQoL 8D, Assessment of Quality of Life 8 Dimensions; DEMQOL, Dementia Quality of Life Measure Qualeffo 41: Quality of Life questionnaire; NHLSD, nursing home life-space diameter; PAINAD, pain assessment in advanced dementia; WOMAC, The Western Ontario and McMaster Universities Osteoarthritis Index.

not adequately report interpretation of the PRO results and 79.1% did not adequately report on any PRO-related limitations.

### Subgroup analyses

#### RCT versus observational

Overall, CONSORT-PRO criteria were reported more adequately in RCTs than in observational cohort studies

( $p=0.026$ ). A significant difference between RCTs and observational research was found in adequate reporting of sample size calculations ( $p=0.002$ ).

#### Predictors of higher CONSORT-PRO score

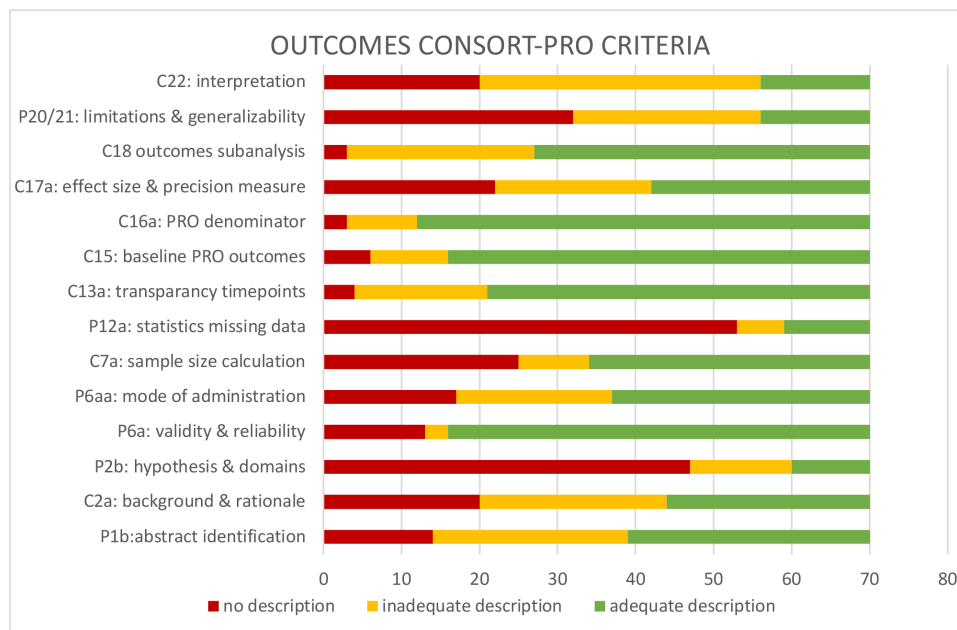
A significant predictor of higher adherence to the CONSORT-PRO score was the study design (coefficient: 1.240, 95% CI 0.165 to 2.333,  $p=0.026$ , adjusted  $R^2=0.279$ ). The second significant predictor for higher CONSORT-PRO scores was if CONSORT-adherence was stated in the methods of an article (coefficient: 1.925, 95% CI 0.422 to 3.429,  $p=0.012$ , adjusted  $R^2=0.279$ ). Journal, country and whether PRO was reported as a primary or secondary outcome did not significantly predict CONSORT-PRO adherence.

### DISCUSSION

To our knowledge, this is the first systematic review that analyses (the quality of) reporting PROs in studies on geriatric patients with a hip fracture. Though PROs are often reported in hip fracture studies, the quality of reporting PROs is not monitored. We found high variability in outcome measures. EQ-5D-3L was most commonly used for measuring QoL, VAS for pain and patient satisfaction, KATZ for ADL. Second, we found that while the overall quality of reporting seems to improve compared with previous literature, the reporting of missing data, PRO limitations and PRO interpretation continues to be suboptimal.

Earlier studies have also reported high variation in outcome measures. A recent scoping review from the National Trauma Research Action Plan group also found high heterogeneity in outcome measures.<sup>16</sup> Standardisation of PROMs should be pursued to adequately analyse and fully comprehend health outcomes in the trauma population. A promising initiative to make PRO reporting more uniform is the PRO measurement information system (PROMIS), a reporting system that was developed in 2007 with the support of the National Institute of Health. This study, however, finds that PROMIS is still infrequently used in research on elderly patients with a hip fracture.

Compared with previous literature, this study generally found higher CONSORT-PRO adherence. In 2015, Bylicki *et al* investigated CONSORT-PRO adherence in randomised clinical trials evaluating cancer therapy. They used a modified version of the CONSORT-PRO questionnaire and found lower adherence scores in 8 out of 13 criteria.<sup>17</sup> Stevens *et al* conducted a systematic review on PRO reporting in RCTs of unplanned general surgeries and also found lower adherence scores in 7 of the 11 CONSORT-PRO criteria that they reported.<sup>18</sup> Mercieca-Bebber *et al* also studied uptake of the CONSORT-PRO in studies of high-impact journals in 2017 for the International Society for Quality of Life Research (ISOQOL) and they found similar reporting rates (67.6% in their study vs 67.3% in this study; in studies who did not report



**Figure 3** Number of studies adhering to the Consolidated Standards of Reporting Trials extension for Patient-Reported Outcomes (CONSORT-PRO) criteria. Red: no description, yellow: inadequate description, green: adequate description.

CONSORT adherence) and (77.7% vs 80.0% in studies who reported CONSORT adherence).<sup>15</sup>

A possible explanation for the difference in reporting rates could be that our study was conducted at least 4 years later than the earlier systematic reviews on PRO reporting. Knowledge of the CONSORT-PRO extension may be more prominent now, eight years after publication of the CONSORT-PRO extension, which may also improve reporting of PROs in general over time. It

could also mean that reporting guidelines (CONSORT, CONSORT-PRO, STROBE) are effective in improving reporting quality. Still, it remains notable since we also include observational studies, for which no PRO guideline exists.

Nonetheless, adherence to several important CONSORT-PRO criteria remained poor. Most included articles in this study did not adequately state their statistical approaches for dealing with missing data. Missing data

**Table 4** Modified Consolidated Standards of Reporting Trials extension for Patient-Reported Outcomes (CONSORT-PRO) extension

Item	Description	No description n (%)	Inadequate description n (%)	Adequate description n (%)
1	P1b: the PRO should be identified in the abstract as a primary or secondary outcome	13 (19.4)	24 (35.8)	30 (44.8)
2	C2a: Background and rationale for PRO assessment	19 (28.4)	24 (35.8)	24 (35.8)
3	P2b: hypothesis should be stated and relevant domains identified, if applicable	44 (65.7)	13 (19.4)	10 (14.9)
4	P6a: Evidence of PRO-instrument validity and reliability should be provided or cited if available	13 (19.4)	3 (4.5)	51 (76.1)
5	P6aa: Mode of administration, including the person completing the PRO and methods of data collection	16 (23.9)	19 (28.4)	32 (47.8)
6	C7a: How sample size was determined: only if PRO was a primary study outcome	23 (34.3)	9 (13.4)	35 (52.2)
7	P12a Statistical approaches for dealing with missing data are explicitly stated	50 (74.6)	6 (9.0)	11 (16.4)
8	C13a: Results: the number of PRO outcome data at baseline and at subsequent time points should be made transparent	3 (4.5)	15 (22.4)	49 (73.1)
9	C15: Baseline table/outcomes showing PRO data when collected	6 (9.0)	10 (14.9)	51 (76.1)
10	C16a: For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups, required for PRO results	3 (4.5)	8 (11.9)	56 (83.6)
11	C17a: For each primary and secondary outcome, results for each group, the estimated effect size and its precision measure (such as 95% CI). For multidimensional PRO results from each domain and time point	22 (32.8)	19 (28.4)	26 (38.8)
12	C18: Results of any other analysis and adjusted analysis distinguishing prespecified from explanatory, including PRO analysis where relevant	3 (4.5)	24 (35.8)	40 (59.7)
13	P20/21: PRO-specific limitations and implications for generalisability and clinical practice	31 (46.3)	22 (32.8)	14 (20.9)
14	C22: PRO data should be interpreted in relation to clinical outcomes including survival data, where relevant	19 (28.4)	35 (52.2)	13 (19.4)

are common, and sometimes unavoidable in PRO-related research.<sup>19</sup> Yet, it is commonly known that missing data can potentially lead to biased estimates and loss of information, study power and interpretability.<sup>19</sup> A number of approaches for dealing with missing data are known, and they have been published by ISOQOL to assist researchers in dealing with missing data. It is worrying that so few studies reported their (methods for dealing with) missing data and that so few journals, regardless of their impact factors, corrected it during peer review. Another finding was that 61.2% of studies did not report appropriate effect size and precision measurement. Most studies only reported p values. Though a p value expresses statistical significance; it provides no information about the magnitude of a possible effect. Reporting p values by themselves can easily lead to misinterpretation of the results and should therefore be avoided.<sup>20 21</sup> An effect size (ie, OR, relative risk, Pearson correlation) shows the *magnitude of the difference between groups* and is therefore more helpful in interpreting results. Along with a precision measure, they are now required in most reporting standards.

Other important PRO criteria that were rarely met were 'interpretation of PRO findings' and 'PRO limitations' (both were met in only ~20% of cases). Failure to report limitations and failure to adequately interpret results can lead to unwarranted conclusions. Since outcomes of PRO studies are used as important arguments in shared-decision making, high reporting quality is important. Nonetheless, this study shows that not only authors, but also journals still fail in adequate reporting. This limits the objectiveness of medical research and more efforts should be made to improve quality of reporting.

Observational studies in this systematic review had significantly lower reporting rates than RCTs, as hypothesised. Moreover, the study design (RCT vs observational study) was a significant predictor for CONSORT adherence. No reporting guidelines for PROs currently exist for observational studies. Hence, when interpreting the results, it should be kept in mind that the observational studies were scored with a scoring system that was not specifically designed for that study design. This could be a reason for the lower adherence scores. RCTs are often favoured over observational studies due to the possible unmeasured bias or confounding that can occur in observational research. However, surgical RCTs are difficult to conduct and, as shown in this article, can lead to misleading results in case of inadequate reporting. Recent literature thus stresses the usefulness of observational studies in trauma surgery.<sup>22</sup> We therefore believe that extending the CONSORT-PRO extension to STROBE (an example of which is found in online supplemental table E) would be valuable for future research.

This study has several limitations. Though we made a thorough search strategy along with an experienced librarian, a publication bias could have occurred along with potential bias due to the exclusion of non-English studies and studies with patients <65 years of age. However, since we included a relatively large amount of studies

from different databases, we do not think additional studies would have significantly changed our conclusion. Another limitation is that we included studies from April 2013 and the protocols of these studies may have been written before CONSORT-PRO publication. This may have caused adherence to be low. Third, we reviewed published articles and online supplemental material of included studies but not study protocols. Some of the CONSORT-PRO criteria may have been addressed in the study protocol. However, CONSORT-PRO advises that the information should be noted in the final report as well. Despite these limitations, this study adds to the existing literature as it is the first to provide a broad overview of quality of reporting PRO in this study population and shows current shortcomings of PRO research. It also serves as a benchmark to monitor quality of PRO reporting in future orthopaedic trauma studies.

## CONCLUSION

This systematic review shows that a wide variety of PRO measures are used to evaluate care in geriatric patients with a hip fracture. To better understand health outcomes, we encourage attempts at standardisation of PRO measures, linking of PRO measures, and creation of standard outcome sets. In addition, while we found slightly higher CONSORT-PRO adherence than previous literature, the reporting of missing data, reporting PRO limitations, and—interpretation remains poor, which impedes adequate interpretation by clinicians and the use of PROs in clinical practice. We therefore advise researchers, reviewers and journal editors to ensure proper PRO reporting. To improve quality of PRO reporting in observational studies, we recommend extending the CONSORT-PRO studies to the STROBE guidelines for observational studies.

**Twitter** Marilyn Heng @marilyn\_heng

**Acknowledgements** Lisa Philpotts, Treadwell Library, Massachusetts General Hospital, Boston, MA, for her guidance in the development of this systematic review's search strategy.

**Contributors** Conceptualisation was performed by: PvdV, SW, E-JV and RMH. Data curation was performed by: PvdV, SW. Formal analysis, investigation, project administration, visualisation, and writing - original draft were performed by: PvdV, SW, MH. Funding acquisition is not applicable. Methodology was performed by: PvdV, MH. Resources and supervision were performed by: MH, RMH, E-JV. Software is not applicable. Validation was performed by: PvdV, SW, MH, RMH. Writing - review and editing was performed by: all authors. PvdV takes full responsibility for the finished work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Map disclaimer** The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Puck van der Vet <http://orcid.org/0000-0003-1577-1660>

Marilyn Heng <http://orcid.org/0000-0002-4047-6818>

#### REFERENCES

- Veronese N, Kolk H, Maggi S. Epidemiology of fragility fractures and social impact 2021:19–34.
- Cummings SR, Melton LJ. Epidemiology and outcomes of osteoporotic fractures. *The Lancet* 2002;359:1761–7.
- Veronese N, Maggi S. Epidemiology and social costs of hip fracture. *Injury* 2018;49:1458–60.
- Gagnier JJ. Patient reported outcomes in orthopaedics. *J Orthop Res* 2017;35:2098–108.
- CMS. Supplemental material to the CMS MMS blueprint patient reported outcome measures, 2021. Available: [cms.gov/files/document/blueprint-patient-reported-outcome-measures.pdf](https://www.cms.gov/files/document/blueprint-patient-reported-outcome-measures.pdf) [Accessed 05 May 2022].
- Weldring T, Smith SMS. Patient-Reported outcomes (pros) and patient-reported outcome measures (PROMs). *Heal Serv Insights* 2013;6:61.
- Peeters CMM, Visser E, Van de Ree CLP, et al. Quality of life after hip fracture in the elderly: a systematic literature review. *Injury* 2016;47:1369–82.
- van der Vet PCR, Kusen JQ, Rohner-Spengler M, et al. The quality of life, patient satisfaction and rehabilitation in patients with a low energy Fracture-Part III of an observational study. *Geriatr Orthop Surg Rehabil* 2021;12:215145932110464. doi:10.1177/21514593211046407
- Talevski J, Sanders KM, Duque G, et al. Effect of clinical care pathways on quality of life and physical function after fragility fracture: a meta-analysis. *J Am Med Dir Assoc* 2019;20:926.e1–926.e11. doi:10.1016/j.jamda.2019.02.022
- Haywood KL, Brett J, Tutton E, et al. Patient-Reported outcome measures in older people with hip fracture: a systematic review of quality and acceptability. *Qual Life Res* 2017;26:799–812.
- Brundage M, Bass B, Davidson J, et al. Patterns of reporting health-related quality of life outcomes in randomized clinical trials: implications for clinicians and quality of life researchers. *Qual Life Res* 2011;20:653–64.
- Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT pro extension. *JAMA* 2013;309:814–22.
- Page M, McKenzie J, Bossuyt P. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372. doi:10.1136/bmj.n71
- Slim K, Nini E, Forestier D, et al. Methodological index for non-randomized studies (minors): development and validation of a new instrument. *ANZ J Surg* 2003;73:712–6. doi:10.1046/j.1445-2197.2003.02748.x
- Mercieca-Bebber R, Rouette J, Calvert M, et al. Preliminary evidence on the uptake, use and benefits of the CONSORT-PRO extension. *Qual Life Res* 2017;26:1427–37.
- Herrera-Escobar JP, Osman SY, Das S, et al. Long-Term patient-reported outcomes and patient-reported outcome measures after injury: the National trauma research action plan (NTRAP) scoping review. *J Trauma Acute Care Surg* 2021;90:891–900. doi:10.1097/TA.0000000000003108
- Bylicki O, Gan HK, Joly F, et al. Poor patient-reported outcomes reporting according to consort guidelines in randomized clinical trials evaluating systemic cancer therapy. *Ann Oncol* 2015;26:231–7.
- Stevens DJ, Blencowe NS, McElnay PJ, et al. A systematic review of patient-reported outcomes in randomized controlled trials of unplanned general surgery. *World J Surg* 2016;40:267–76.
- Mercieca-Bebber R, Palmer MJ, Brundage M, et al. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (pro) data: a systematic review. *BMJ Open* 2016;6:e010938.
- Sullivan GM, Feinn R. Using effect Size-or why the P value is not enough. *J Grad Med Educ* 2012;4:279–82.
- Tanha K, Mohammadi N, Janani L. P-value: what is and what is not. *Med J Islam Repub Iran* 2017;31:377–8.
- Beks RB, Bhashyam AR, Houwert RM, et al. When observational studies are as helpful as randomized trials: examples from orthopedic trauma. *J Trauma Acute Care Surg* 2019;87:730–2.