Title:

Patient-reported outcome measures in older people with hip fracture: a systematic review of quality and acceptability.

Abstract

Purpose: Hip fracture is the most common serious injury of older people, often resulting in reduced mobility and loss of independence. However, guidance for the use of patient-reported outcome measures (PROMs) does not exist: we describe the first review to apply internationally endorsed criteria in support of PROM quality and acceptability in this group, and make recommendations for future applications.

Methods: Systematic literature searches of major databases (1980-2015) to identify published evidence of the application and quality of clearly defined measures. Evidence of measurement and practical properties, and the extent of active patient involvement, was sought. Study and PROM quality was assessed against recommended criteria.

Results: 71 articles relating to 28 PROMs (Generic n=12; Specific n=16) were included. The SF-36 (v1) and EuroQoL EQ-5D 3L were the most widely evaluated measures with acceptable evidence of measurement properties, but limited evaluations of practical properties or relevance to this group. Evidence was mostly limited for the remaining measures. Hypothesized associations between variables were infrequently evaluated. Evidence of data quality, test-retest reliability, responsiveness, interpretation, acceptability and feasibility was also limited. Active patient involvement in PROM development or evaluation was not reported. There was limited evaluation of proxy completions.

Conclusions: The paucity of robust evaluations is disappointing and prevents clear recommendations for PROM-based assessment. Further research must urgently seek to identify which outcomes really matter to this group. Future PROM selection must be underpinned by research which focuses on methodological quality, including issues of acceptability, relevance, feasibility of application, and proxy completion, whilst seeking to actively incorporate the perspective of patients and their advocates.

Authors:

Haywood KL, Brett J, Tutton E, Staniszewska S.

Title:

Patient-reported outcome measures in older people with hip fracture: a systematic review of quality and acceptability.

Affiliation(s) and addresse(s):

Haywood, Kirstie Louise. DPhil. BSc (Hons). MCSP. SRP. Senior Research Fellow (Patient Reported Outcomes). Royal College of Nursing Research Institute, Division of Health Sciences, Warwick Medical School, Warwick University. Coventry. CV4 7AL.

Brett, Jo. PhD, MSc, MA, BSc (Hons). Senior Research Fellow, AHPD, Faculty of Health and Life Sciences, Oxford Brookes University, Jack Straw's Lane, Marston, Oxford OX3 OFL.

Tutton, Elizabeth. PhD. RN. Senior Research Fellow. Royal College of Nursing Research Institute, Division of Health Sciences, Warwick Medical School, Warwick University. Coventry. CV4 7AL; and Trauma Research, Oxford University Hospitals NHS Foundation Trust, Kadoorie Centre, John Radcliffe Hospital, Oxford, OX3 9DU.

Staniszewska, Sophie. DPhil. Senior Research Fellow. Royal College of Nursing Research Institute, Division of Health Sciences, Warwick Medical School, Warwick University. Coventry. CV4 7AL.

Corresponding author:

Email: k.l.haywood@warwick.ac.uk.

Tel: 024 7615 0616

Fax: 024 7615 0643

Abstract (257/250)

Purpose: Hip fracture is the most common serious injury of older people, often resulting in reduced mobility and loss of independence. However, guidance for the use of patient-reported outcome measures (PROMs) does not exist: we describe the first review to apply internationally endorsed criteria in support of PROM quality and acceptability in this group, and make recommendations for future applications.

Methods: Systematic literature searches of major databases (1980-2015) to identify published evidence of the application and quality of clearly defined measures. Evidence of measurement and practical properties, and the extent of active patient involvement, was sought. Study and PROM quality was assessed against recommended criteria.

Results: 71 articles relating to 28 PROMs (Generic n=12; Specific n=16) were included. The SF-36 (v1) and EuroQoL EQ-5D 3L were the most widely evaluated measures with acceptable evidence of measurement properties, but limited evaluations of practical properties or relevance to this group. Evidence was mostly limited for the remaining measures. Hypothesized associations between variables were infrequently evaluated. Evidence of data quality, test-retest reliability, responsiveness, interpretation, acceptability and feasibility was also limited. Active patient involvement in PROM development or evaluation was not reported. There was limited evaluation of proxy completions.

Conclusions: The paucity of robust evaluations is disappointing and prevents clear recommendations for PROM-based assessment. Further research must urgently seek to identify which outcomes really matter to this group. Future PROM selection must be underpinned by research which focuses on methodological quality, including issues of acceptability, relevance, feasibility of application, and proxy completion, whilst seeking to actively incorporate the perspective of patients and their advocates.

Keywords: systematic review; PROM quality; acceptability; hip fracture; older people

Introduction

Hip fracture is the most common serious injury of older people, often resulting in reduced mobility and loss of independence, and representing one of the greatest challenges to the healthcare community [1]. In 1990, a global incidence of 1.31 million hip fractures was reported, with an associated 740,000 deaths [2]. Hip fractures represent a growing, worldwide socioeconomic burden: current costs to England's NHS are estimated at £1.4 billion, or 1% of the NHS budget [3].

Traditionally, outcome assessment for patients presenting with proximal femoral fracture was focused on mortality / morbidity rates, surgical implant success or operative complications [4, 5]. However, the growing focus on patient-centred care and recognition of the importance of understanding the impact of hip fracture and associated care from the perspective of the patient, has resulted in a shift in how outcomes are assessed in clinical trials, audit and routine practice settings towards the assessment of patient experience and the quality of life achieved [6-8]. The use of well-developed patient-reported outcome measures (PROMs) – single or multi-item questionnaires which seek to assess how patients feel, what they can and cannot do and how they live their lives as a consequence of their health and associated health care, could provide critical information to enhance patient-centred health care [9]. However, guidance for appropriate PROM-based assessment following hip fracture does not exist, and little is known about which outcomes are most important to patients.

Where uncertainly exists, structured reviews of evidence can be essential to informing the selection of relevant and appropriate measures. Three recent articles have reviewed the use and availability of patient-reported and clinician-reported measures following surgical interventions for hip pathology [4] and traumatic hip fracture [5,10]. Ahmad et al, [4] suggested that the outcomes of elective or traumatic hip surgery should be assessed with a clear and concise hip-specific measure that allows consideration of co-morbidities, the use of walking aids, and includes a generic component. However, a selective review of commonly used hip-specific, disease-specific and generic measures highlighted numerous limitations – with none of the reviewed measures fulfilling the suggested requirements. Moreover, few measures had been adequately evaluated, further limiting recommendations. In conclusion, while recommending a combination of hip-specific (Oxford Hip Score - OHS), disease-specific (Western Ontario and McMaster University Osteoarthritis Index - WOMAC) and generic measures, the need for further robust evaluations was emphasised.

Hutchings et al, [5] reviewed fourteen of the most commonly used clinician-reported, performance-based and patient-reported measures in the elderly proximal hip fracture population. They highlighted significant variation in outcome reporting, with no single measure in widespread use in this population. Although the search strategies applied in pursuit of published psychometric evidence were limited [11], concerns pertaining to the limited availability of robust evaluations by which to determine the 'validity' of measures were also raised. Cautious recommendations included: a generic measure, such as the EuroQoL EQ-5D or SF-36; a measure of activities of daily living (ADL), such as the Barthel Index; and a hip-specific measure, such as the OHS, although evidence for the latter was very limited.

A further article provides a limited review of the strengths and caveats of five named measures applicable for use in patients with hip pathology, summarizing, but not directly comparing, their suitability for use in the rheumatology community [12]. However, none of the reviews considered the methodological quality of reviewed studies, thus making it difficult to judge the strength of psychometric evidence underpinning any recommendations [13]. Evidence-based healthcare demands the critical appraisal of study methodological quality; where a study is of poor methodological quality, confidence in the results is reduced [14]. Similarly, an appreciation of the methodological quality of PROM evaluative studies is crucial to data interpretation [11; 15]. Moreover, the reviews of psychometric evidence were often limited, non-transparent and non-systematic.

The aim of this review was to critically appraise, compare and summarize the quality and acceptability of published PROMs evaluated following completion (self, interview, or proxy) by older patients (aged 60 years and above) who had sustained a hip fracture. The results of the review will assist in the selection of a PROM suitable for inclusion in routine practice, audit, or clinical research settings.

Methods

Identification of studies and PROMs: Search strategy

The search strategy sought to retrieve references relating to the development and/or evaluation of multi-item PROMs used in the evaluation of older people (aged 60 years and above) who had sustained a hip fracture. Searches used medical subject headings (MeSH terms) and free text searching to combine terms specific to hip fracture with terms relevant to health measurement and PROM evaluation [11; 16]. Four databases were searched: 1980 to Aug 2015 (MEDLINE (OVID), EMBASE, CINAHL and PsychINFO). Further database searches used names of identified PROMs. Citation lists of included articles and earlier reviews of measures used in hip pathologies or hip fracture [5; 6; 10] were reviewed.

Inclusion/exclusion criteria

All titles and abstracts were assessed for inclusion/exclusion by one reviewer (JB); a sub-set of 10% were double assessed (JB,KH) and agreement checked. Published articles were included if they provided evidence of development/evaluation for clearly defined and reproducible multi-item PROMs, assessing single or multiple domains of health, following self-, interview or proxy completion by older people (aged 60 years and above) who had sustained a hip fracture. Articles relating solely to PROM application without some evidence of measurement and/or practical properties were not included. Included PROMs were categorized as generic (profile or utility), hip-specific (surgeon or patient-completed), condition-specific, or domain-specific [16]. Proxy completion was highlighted. Evaluations in non-English speaking populations published in English language journals were included. Single-item and mobility measures, radiographic and imaging techniques were excluded, as were measures without evidence of reliability or validity.

Data extraction and appraisal

A data extraction form, informed by key psychometric texts [17; 18], guidance for evaluating PROM quality [19], earlier reviews [16; 20] and the requirements of the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist [21; 22] was developed. Data extraction captured both study and PROM-specific information. Evidence for measurement properties included: reliability (internal consistency; test–retest, intra/inter-tester); validity (content; construct including within scale and analyses against external criteria - convergent/divergent; known groups; evidence of explicit hypothesis

testing was sought); responsiveness (criterion-based or construct-based assessment [18; 21] was prioritised; reporting of effect size (ES) statistics was also extracted); interpretation (minimal important difference); precision (data quality; end effects). Evidence for practical properties included acceptability (relevance and respondent burden) and feasibility. The extent of active patient involvement in PROM evaluation was sought [23; 24].

Assessment of study methodological quality

The COSMIN checklist provides a consensus-based framework against which the methodological quality of PROM-based evaluative studies can be judged [21; 22]. Nine specific measurement properties are described: each checklist contains a list of items against which study methodological quality is assessed; items are scored on a 4-point rating scale (that is, excellent, good, fair, poor) [22]. Study methodological quality was evaluated per measurement property and determined by the lowest rating of any of the items in each checklist section. Two reviewers (JB, KH) independently applied the checklist to each article. Agreement was checked and any disagreement was resolved through discussion.

Assessment of PROM quality

A similar consensus-based checklist for the appraisal of PROM quality does not exist. However, a synthesis of various recommendations was described in an earlier review [16] and provided a pragmatic checklist against which the results of PROM testing was judged.

Data synthesis

A qualitative synthesis of data per reviewed PROM informed the overall judgement of quality and acceptability. As per earlier reviews [13; 20], the synthesis considered the following factors: 1) study methodological quality (COSMIN scores); 2) the number of studies reporting evidence per PROM; 3) the results for each measurement property for each PROM; and 4) evidence consistency between reviewed studies. Two element to the data synthesis score are described: First, the *overall quality* of a measurement property was reported as: adequate (+), not adequate (-), conflicting (+/-), or unclear (?). Second, *levels of evidence* for the overall quality of each measurement property was categorized as ‘strong’, ‘moderate’, ‘limited’, ‘conflicting’, or ‘unknown’ [13; 25].

Results

Identification of studies and PROMs

The initial searches (conducted 1980-July 2012) generated more than 9000 articles (Figure 1). Following title and abstract assessment 177 articles were reviewed in full, including eight from citation searches. Update searches (conducted August 2015) generated a further 934 articles; title and abstract assessment resulted in a further 50 articles for full review (Figure 1). No additional articles were identified from updated citation searches.

A total of 71 articles were included in the review (Figure 1) (Appendix Table 1), providing generally limited evidence of measurement quality or acceptability for 28 clearly defined PROMs (Tables 1-3). It was frequently impossible to include measures due to inadequate descriptions or lack of reference.

Characteristics of reviewed measures

Twelve generic measures of health status, quality of life or capability were reviewed. Six were profile measures: the COOP-WONCA Charts [26], Nottingham Health Profile (NHP) [27], Quality of Life Scale (QoLS) [28], Short-Form 36-item Health Survey (version 1)(SF-36 v1) [29], the SF-12 (SF-12) (version 1) [30], and the World Health Organization Quality of Life questionnaire – short form (WHOQOL-BREF) [31]; and one single item measure of quality of life - the EuroQol EQ-thermometer (EQ-VAS) [32]. Four were preference-based utility measures: EuroQoL EQ-5D 3L [32], Health Utility Index 2 (HUI-2) [33] and 3 (HUI-3) [34], and the SF-6D derived from completion of the SF-36 or SF-12 [35]. One measure – the ICEpop CAPability measure for Older people (ICECAP-O), is an older-people specific profile measure of capability and well-being for application in economic evaluations [36].

Three hip-specific measures were reviewed. Two measures – the Charnley Hip Score (CHS) [37] and the Harris Hip Score (HHS) [38] - were designed to be administered by a qualified health professional. They were included in the review due to their widespread use and the inclusion of several patient-based items (pain, mobility, functional activities): it was often difficult to discern how these particular items were completed and the relative contribution of patients to the assessment. The Oxford Hip Score (OHS) [39] is the only patient-completed hip-specific assessment, developed for the assessment of pain and functional ability following a total hip replacement surgery. A measure specific to fragility-hip fracture was not identified.

Two patient-completed, disease-specific measures (Osteoporosis Quality of Life Questionnaire (version 2) (OPAQ-2) [40] and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [41] were reviewed.

Additionally, 11 domain-specific measures were reviewed: three measures of emotional well-being - Geriatric Depression Scale (GDS)[42], Hamilton Rating Scale for Depression (HAM-D) [43] and the Zung Depression Inventory (ZungDI) [44]; and eight measures of (instrumental) activities of daily living (I/ADL): Barthel Index (BI) [45], Modified-Barthel Index [46], Functional Activities Index (FAI) [47], Functional Independence Measure (FIM) [48], Functional Status Questionnaire (FSQ) [49], Lawton IADL scale (Lawton-IADL)[50], Katz Index of Independence in ADL (Katz ADL)[51] and the OARS Multi-dimensional Functional Assessment Questionnaire (OMFAQ)[52]. One of these measures (Katz ADL) is clinician-completed but was included in the review due to its wide-spread use in this patient population. Although most often clinician-completed, the original and Modified versions of the Barthel Index can be self-completed by patients and hence were included in the review.

Patient and study characteristics

Characteristics of included studies are detailed in the Appendix (Appendix Table 1). Sample sizes ranged between 25 and more than 13,000. The mean ages of the patient groups ranged from 69 to 87 years. All patients had sustained a hip fracture. The majority of studies excluded cognitively impaired patients. Only two studies specifically evaluated the impact of proxy completion on PROM performance – the HUI [53] and the FIM [54]. The majority of studies were cohort studies; ten were randomized controlled trials. Several PROM comparative evaluations were included [55-62].

Measurement properties and methodological quality

PROM measurement properties and methodological quality of reviewed studies are summarized in Tables 2 and 3 respectively. Most studies reported validity (mostly known-groups); but few studies formulated a priori hypothesized associations between questionnaires in advance of testing or provided any clarity with regards to the way in which missing data was handled, hence the fair or poor methodological quality rating. Eleven studies reported evidence of reliability (Table 2). With the exception of a just five studies [58; 62-65] which provided acceptable evidence in support of the longitudinal validity of reviewed measures, evidence of measurement responsiveness was largely lacking (Tables 2-

3). Twelve studies generated effect size (ES) statistics (Table 2). Although reflecting the size of change score rather than responsiveness [18; 66], where this was accompanied by a clear (or most often assumed) hypothesis detailing the expected direction and size of effect, this evidence was extracted but not included in the final COSMIN-framed synthesis. The majority of studies reported evidence of statistical significance of change scores (for example, paired t-tests and associated p-values); such evidence is an inappropriate reflection of measurement responsiveness and was not included in the review [16].

Generic measures

The SF-36 (v1) is the most widely evaluated measure in this population group, with moderate to strong evidence supporting measurement validity and responsiveness (Tables 2-3) to change in health following surgical repair of a hip fracture. There is moderate evidence of internal consistency, with some limited evidence of test-retest reliability. Evaluations of the revised SF-36 (version 2) were not identified.

Although lacking evidence of reliability, the EuroQoL EQ-5D 3L has moderate evidence supporting its validity (mostly known-groups) and responsiveness. A strong correlation between the EQ-5D 3L and the OHS has been reported (range 0.70 to 0.74) [55; 67]. Acceptable evidence of responsiveness has been reported following surgical repair of hip fracture (for example, [58; 60; 68]) (Tables 2 and 3). Large standardised effect sizes (range 0.64 to 0.68) have been reported in two large UK-based patient cohorts in the initial 4 to 6-week follow-up period; much smaller values were reported over the longer term (ES range 0.27 at 1 year to 0.32 at 12-weeks) [55; 67]. Acceptable evidence supports the discriminative ability of the EQ-5D: for example, between groups defined by the external clinical criterion ‘good versus less good clinical outcome’ [58]. Few studies provide evidence of change score correlations: where reported between the EQ-5D and SF-36 domains, correlations ranged from 0.03 to 0.45, the strongest being between the EQ-5D index score and the SF-36 domains Body Pain, Vitality, and Physical Function [58].

There were few comparative evaluations of generic measures. Evidence of validity was equally supportive of the SF-36 and EQ-5D [58], although evidence suggests that the EQ-5D may be more responsive where substantial change in health is expected [58]. Comparable ES statistics were reported for the COOP-WONCA charts and the NHP [59] at 4-months post hip-fracture, and for the HUI-2 and HUI-3 at 6-months post hip fracture [63]. For the

remaining generic measures there is little, conflicting, or no evidence of measurement reliability, validity and responsiveness (Tables 2-3).

Specific measures

Hip-specific: Although widely used, there is no evidence in support of the reliability or responsiveness of the Charnley and Harris Hip scores, and evidence of measurement validity is limited. These measures should be used with caution. Three studies provide limited evidence of measurement validity [55; 67; 69] and two report ES statistics [55; 67] for the OHS. Although the hypothesized association between variables was not stated a priori, moderate to strong correlations between the OHS and EQ-5D 3L and comparable ES statistics were reported at 4-weeks (ES 1.14 (OHS)) and 4-months (ES 0.39 (OHS)) post-op. By comparison, small ES statistics were reported for the ICECAP-O at all follow-up points of the same study; correlations with both the OHS and EQ-5D were small [55]. Further comparative evaluations of the OHS, including rigorous evidence of responsiveness (for example, correlation of change scores), relevance and acceptability are required to increase confidence in future applications.

Disease-specific: Although widely evaluated in other conditions, the two disease-specific measures (OPAQ2 and WOMAC) have not been widely evaluated in this population group: there is no evidence of measurement reliability and unknown or limited evidence of validity. Whilst large effect ES have been reported for the OPAQ2 physical and social activity domains at 12-months post hip fracture, small ES were reported for the back pain and tension domains[70]. The OPAQ2 was developed for use with postmenopausal women and so may have limited applicability in the wider hip-fracture population. These measures should be used with caution until further evidence of essential measurement properties, relevance and acceptability are established.

Domain-specific: These measures were further classified as measures of Emotional well-being (4), and measures of (Instrumental) Activities of Daily Living (I/ADL)(8).

Emotional well-being: The most widely evaluated measure was the 30-item Geriatric Depression Scale (GDS). However, evidence of the reliability, validity and responsiveness of the GDS in this group is very limited and further applications should be made with caution. Moreover, evidence suggests limited acceptability of the GDS following interview-administration due to difficulty responding to the ‘yes/no’ response format and a tendency for

responders to ‘digress’ [71]. Although shorter versions exist, which have recently been recommended for use with older people in the UK [72], these have not been evaluated following completion by older people sustaining a hip fracture. There is very limited evidence of essential measurement properties for the remaining measures.

I/ADL: The most widely evaluated measures in this group are the Functional Impact Measure (FIM), the original Barthel Index (BI), and the Katz ADL, with 21, 10, and 8 reviewed evaluations respectively. Only the BI can be self-completed, but it is most often completed by clinicians. The FIM is interview-administered with a trained clinician, and the Katz ADL is completed by a trained clinician. Although not patient-completed measures, these latter two measures were reviewed due to their widespread use as measures of patient-based outcome. The FIM has good evidence of test-retest reliability, moderate evidence of validity (convergent/divergent and known groups), but limited evidence of responsiveness to change following surgical repair of hip fracture [64]. Evidence of acceptability and feasibility of FIM completion was not reported in this group; although only containing 18 items, administration may require between 30 and 60 minutes. Moderate evidence supports the validity of the original Barthel Index; but evidence of reliability is lacking. Moderate to large ES statistics have been reported at 1-, 4-, 6- and 12-months post hip-fracture [59; 73], comparable to values reported for physical mobility domains of the NHP and COOP/WONCA [59]. Containing only 10 items, the BI can be self-completed in 10 minutes (not reported in this population) or clinician-completed in between 5-10 minutes, suggesting better acceptability and feasibility than the FIM. Evidence in support of the Katz ADL and the remaining measures of I/ADL measures is very limited: the majority lack any evidence of reliability and responsiveness, and evidence of validity is restricted to poor quality, known-groups analyses.

Discussion

Despite the large number of studies which now include PROMs in the evaluation of hip fracture in older people, there are disappointingly few robust evaluations from which to draw clear recommendations for PROM selection. Confidence in PROM selection requires evidence of both measurement and practical properties. However, evidence of relevance, test-retest reliability, measurement error, structural validity and score interpretation was not identified for any reviewed measure; just four (SF-36, HUI2, HUI3, FIM) had limited evidence of inter-rater reliability. With the exception of three measures (SF-36, EQ-5D, Barthel Index), evidence of responsiveness was absent or limited. Evidence of acceptability and feasibility was poorly reported; only two studies evaluated the impact of cognitive impairment and the role of proxy completion.

In comparison with earlier PROM reviews for various hip pathologies [4; 5; 10; 12], the strength of this review lies in the first application of transparent appraisal frameworks supporting evaluations of study [22] and PROM quality [16; 20] in this population. These frameworks highlighted significant methodological and quality concerns which must be addressed in future PROM evaluations if robust recommendations are to be made; this is particularly pertinent to the evaluation of measurement responsiveness. The frameworks were independently applied to all included studies by two reviewers (JB, KH) and agreement checked; however, a limitation of the review is that the synthesis score was applied only by a single reviewer (KH). The grading criterion supports synthesis of large amounts of data but, although applied in several recent reviews [13; 20], itself lacks robust evidence of reliability and validity and should therefore be cautiously interpreted. Although only English-language publications were included in the review, a wide-range of questionnaires and language versions were reviewed and any selection bias is unlikely. However, evidence from different countries and language versions was combined, which may fail to take into consideration any cross-cultural variation in performance and should be considered for future reviews [25].

The extensive literature search included the major health databases, and was further supplemented by reference to existing reviews and recent reports. Although only English-language studies were included, the diversity of measures and language versions included in the review suggests that any selection bias is unlikely. Reviewed studies included patients with a lower age of 60 years; no upper-age limit was imposed. However, few studies included cognitively impaired patients or explored the impact of such impairment on PROM

completion and performance. We are confident that the results are generalizable to the wider population of older people who sustain a hip fracture, but may not represent the experience of patients with varying degrees of cognitive impairment.

In keeping with other reviews (for example, [16; 20]), the relevance, content or face validity of the reviewed measures has not been reported in the hip fracture population. The relevance or appropriateness of a measure to the target population is a crucial consideration, particularly if the group differs from the population in which the measure was originally developed [21]. Only one measure – the ICECAP-O, is older people specific; one measure – the Oxford Hip Score (OHS), is intervention-specific (total hip replacement); no measure is specific to hip fracture. Qualitative research which seeks to understand what really matters to older people as an outcome from healthcare following hip fracture is essential to informing appropriate question and PROM content [74; 75]. The active collaboration of the older population in the development and/or evaluation of reviewed PROMs – for example, working in partnership to co-produce knowledge, was not reported.

The estimated range of cognitive impairment in older people with hip fracture is between 31% and 88% (mean 47%) [46; 76; 77], often significantly limiting their ability to self-report [70; 78]. Alternative information sources or proxy respondents such as primary caregivers, close relatives, or health professionals, may be utilized. Few studies have explored the relative impact of proxy-completion in this population [75; 79]. However, as observed with older people more generally [80], evidence suggests that proxy and patient responses are *not* interchangeable, and agreement is higher for more observable health constructs. Consistency of proxy completer (that is, inter-tester reliability) has not been addressed. Anecdotal evidence suggests that many older patients may lack a consistent named proxy – for example, due to lack of dedicated family member or named / regular health professional. The impact of different proxy respondents for both research and clinical practice settings should be considered.

Only the SF-36 and three I/ADL measures (BI, FAI, FIM) have moderate evidence of both convergent and known-groups validity, with clear evidence of a priori hypothesized associations between variables being explored. With the exception of the EQ-5D and Katz ADL, for which moderate evidence of know-groups validity (and limited convergent for the EQ-5D) was reviewed, the majority of the remaining measures had limited or unknown evidence. The majority of studies simply compared the scores on measures between patients

who had sustained a hip fracture and a population-based cohort, or between different types of hip fracture. Evidence of construct validity was limited for all measures by the failure of authors to state a priori hypothesized size and /or direction of associations between variables or known groups. There were no evaluations of structural validity in this population.

Evidence of measurement reliability was very limited; test-retest and measurement error was not reported. Internal consistency reliability was reported only for the SF-36. Limited inter-rater agreement following interview administration of the SF-36 [56] and moderate patient-proxy agreement for the HUI-2 and HUI-3 [79] and the FIM [75] was reported.

As reported by other reviewers (for example, [13; 20; 81]), very few studies with evidence of measurement responsiveness were included. With the exception of the SF-36, EQ-5D, NHP and the FIM evidence of responsiveness was mostly limited or not identified. Although numerous studies described longitudinal change in health or compared the relative benefit of different treatment approaches, the majority reported only the statistical significance of score change. Due to the failure to explore the validity of score change, statistical significance is judged an inappropriate representation of measurement responsiveness [18]. Moreover, due to the difficulties of disentangling issues of responsiveness from the effects of an intervention, judgment on PROM responsiveness in trial-based studies is difficult [18]. Due to the failure to provide a priori hypothesized expectations for the direction and size of score change or correlation, many of the included studies were judged to be of relatively poor methodological quality.

Well-developed PROMs provide essential evidence of the impact of healthcare, contributing the patient perspective to the developing evidence-base. Advances in measurement science and a growing recognition of the importance of capturing the patient perspective has resulted in a substantial growth in PROM availability [82]. However, an historic lack of good practice guidance coupled with a limited requirement for transparency and accountability in PROM development and evaluation has resulted in a large number of measures with dubious development history and limited quality. However, recent internationally endorsed guidance for the transparent development and robust evaluation of PROMs seeks to facilitate the development of high quality, relevant and acceptable PROMs with which to inform decision-making [83]. The end-users of PROMs – including clinicians, health professionals, researchers and patients, should demand that PROM-related data is robust, relevant and acceptable, and that accepted standards for development and evaluation have been adhered to.

Evidence-based healthcare requires the judicious integration of best evidence with clinical expertise and patient experience [84]. Establishing 'best evidence' demands the critical appraisal of study methodological quality; therefore, an appreciation of the methodological quality of PROM evaluative studies is crucial to data interpretation. The development of the COSMIN guidance - consensus-based standards for the evaluation of study methodological quality, provides essential and timely guidance to support and inform greater transparency and methodological rigor in PROM evaluation [11; 15]. The future selection of well-developed PROMs with evidence of essential measurement and practical properties generated from high quality studies will ensure that healthcare is underpinned with satisfactory patient-derived evidence, thus reducing the potential for research waste where evidence is founded upon unacceptable evidence.

Outcomes research for traumatic hip fracture urgently requires methodologically rigorous evaluations of relevant and appropriate PROMs. Evidence suggests that the SF-36 and EQ-5D are candidate measures which require further evaluation. Although long, with evidence of poor-self-completion rates (version 1), the SF-36 (version 2) revised response options could improve acceptability. The EQ-5D benefits from being short (5 questions), with acceptable completion rates in non-cognitively impaired older people; the revised EQ-5D 5L version has not been evaluated in this population, but the improved response categories may improve both relevance and responsiveness. The EQ-5D is the preferred generic measure to inform quality of care assessment by England's Department of Health [85] and has recently been recommended for inclusion in a core outcome set for hip fracture trials [86]. Future evaluations must pay particular attention to the relevance and acceptability of the measures to the target group, to data quality (including missing data), reliability and responsiveness.

However, evaluating the relative benefit of healthcare in patients representing the frailer end of the spectrum and who experience a range of co-morbidities is challenging. Older, more frail patients often view their limitations as a consequence of ageing, or experience difficulties disentangling the impact of the hip fracture from the wide range of co-morbidities they experience [74]. Similar difficulties associated with the impact of multiple co-morbidities have been reported in other patient populations, for example, in mental health [87]. The often complex and diverse nature of health experienced by this population group highlights the need for a well-crafted, relevant and appropriate measures which capture the broad array of important health domains, and have the option for proxy completion.

Moreover, evidence would suggest a benefit to be gained by utilizing both generic and specific measures [79; 88].

The paucity of robust PROM evaluation in this important, diverse, and growing group, is disappointing and prevents clear recommendations for PROM-based assessment. The active engagement of key stakeholders, including patients, carers' and health professionals, should seek to support more collaborative PROMs-related research, the co-production of knowledge and selection of high quality measures that are both relevant and appropriate [23]. Identifying the 'best' measures will require robust comparative evaluations of candidate measures, and should include generic (including the SF-36 (version 2), the EuroQoL EQ-5D (3L and 5L versions)) and domain-specific measures (for example, the Barthel Index), whilst addressing the need for a patient-derived hip-fracture specific measure.

Tables:

Figure 1: Flow diagram for article selection.

Table 1: Characteristics of reviewed generic and specific PROMs evaluated in the hip fracture population (total = 28)

Table 2: Methodological quality (COSMIN) and investigated measurement properties per PROM (n=28) per reviewed article (n=71).

Table 3: Measurement properties and methodological quality: Data synthesis, levels of evidence and overall quality of measurement properties per reviewed PROM (n=28)

Appendices:

Table 1: Characteristics of included studies (n= 71)

Compliance with Ethical Standards:

Funding: This study was funded by the Royal College of Nursing Research Institute.

Conflict of Interest: Author Kirstie L Haywood declares that she has no conflict of interest. Author Jo Brett declares that she has no conflict of interest. Author Liz Tutton declares that she has no conflict of interest. Author Sophie Staniszewska declares that she has no conflict of interest

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. British Orthopaedic Association. (2007). The Care of Patients with Fragility Fracture (Blue Book) BOA, September, http://www.bgs.org.uk/pdf_cms/pubs/.
2. Johnell, O., Kanis, J.A (2004). An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. Osteoporosis International, 15, 897-902.
3. National Hip Fracture Database. (2011). NHFD (2010-11)Annual Report. https://data.gov.uk/dataset/national-hip-fracture-database-annual-report-2010-2011.
4. Ahmad, M. A., Xypnitos, F.N., Giannoudis, P.V. (2011). Measuring hip outcomes: Common scales and checklists. Injury, 42(3), 259-264.
5. Hutchings, L., Fox, R., Chesser, T. (2011). Proximal femoral fractures in the elderly: how are we measuring outcome? Injury, 42(11), 1205–1213.
6. Ahmed, S., Berzon, R., Revicki, D.A., et al. (2012). The use of Patient Reported Outcomes (PRO) within comparative effectiveness research: implications for clinical practice and health care policy. Medical Care, 50(12), 1060-1070.
7. Basch, E., Abernethy, A.P., Mullins, C.D. et al. (2012). Recommendations for incorporating patient-reported outcomes into clincial comparative effectiveness research in adult oncology. Journal of Clinical Oncology, 30(34), 4229-4255.
8. Basch, E. (2012). Beyond the FDA PRO guidance: Steps toward integrating meaningful patient reported outcomes into regulatory trials and US drug labels. Value in Health, 15, 401-403.
9. Barham, L., Devlin, N. (2011). Patient reported outcome measures: implications for nursing. Nursing Standard, Jan 5-11; 25(18), 42-45.
10. Bryant, D. M., Sanders, D.W., Coles, C.P., Petrisor B.A., Jeray, K.J., Laflamme, G.Y. (2009). Selection of outcome measures for patients with hip fracture. J Orthop Trauma. , Jul;23(6), 434-441.
11. Terwee, C. B., Jansma, E.P., Riphagen, I.I., de Vet. H.C (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res., Oct;18(8), 1115-1123.
12. Nilsdotter, A., Bremander, A. (2011). Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire Arthritis Care & Research, 63(Issue Supplement S11), S200–S207.
13. Elbers, R. G., Reitberg, M.B., van Wegen, E.E., Verhoef, J., Kramer, S., Terwee, C.B., Kwakkel, G. (2012). Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. Qual Life Res. , Aug; 21(6), 925-944.
14. Straus, S., Glasziou, M., Richardson, W.,Haynes, R. (2010). Evidence-Based Medicine: How to Practice and Teach It. Churchill Livinstone Elsevier, 4th Ed.
15. Terwee, C. B., Mokkink, L.B., Knol, D.L., Ostelo, R.W., Bouter, L.M., de Vet, H.C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res., May; 21(4), 651-657.
16. Haywood, K. L., Staniszewska, S., Chapman, S. . (2012). Quality and Acceptability of patient reported outcome measures in Chronic Fatigue Syndrome / Myalgic Encephalitis (CFS/ME): a structured review. Qual Life Res., Feb;21(2), 35-52.
17. Streiner, D. L., Norman, G.R. (2014). Health Measurement Scales: A practical guide to their development and use. Oxford University Press, USA, 5th edition.
18. de Vet, H., Terwee, C.B., Mokkink, L.B., Knol, D.L. (2011). Measurement in Medicine. A Practical Guide. Cambridge University Press, Sept, http://dx.doi.org/10.1017/CBO9780511996214.

19. Terwee, C. B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., Bouter, L.M., de Vet, H.C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol., 60(1), 34-42.
20. Haywood, K. L., Collins, S., Crawley, E. (2014). Assessing severity of illness and outcomes of treatment in children with Chronic Fatigue Syndrome/Myalgic Encephalitis (CFS/ME): a systematic review of patient-reported outcome measures. Child Care Health Dev, Nov;40(6), 806-824.
21. Mokkink, L. B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M, de Vet, H.C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol., Jul; 63(7), 737-745.
22. Terwee, C. B., Mokkink, L.B., Knol, D.L., Ostelo, R.W., Bouter, L.M., de Vet, H.C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. , May; 21(4), 651-657.
23. Staniszewska, S., Haywood, K.L., Brett, J., Tutton, L (2012). Patient and public involvement in patient-reported outcome measures: evolution not revolution. Patient 5(2), 79-87.
24. Mattsson, P., Alberts, A., Dahlberg, G., Sohlman, M., Hyldahl, H.C., Larsson, S. (2005). Resorbable cement for the augmentation of internally-fixed unstable trochanteric fractures. A prospective, randomised multicentre study. J Bone Joint Surg Br, 87(9), 1203-1209.
25. Conijn, A. P., Jens, S., Terwee. C.B., Breek, J.C., Koelemay, M.J.W. (2015). Assessing the quality of available patient reported outcome measures for intermittent claudication: a systematic review using the COSMIN checklist. Eur J Vasc Endovasc Surg., 49(3), 316-334.
26. Nelson, E., Wasson, J., Kirk, J., Keller, A., Clark, D., Dietrich, A., Stewart, A., & Zubkoff, M. (1987). Assessment of function in routine clinical practice: description of the COOP Chart method and preliminary findings. Journal Of Chronic Diseases, 40 Suppl 1, 55S-69S.
27. Hunt, S. M., McKenna, S. P., McEwen, J., Backett, E. M., Williams, J., & Papp, E. (1980). A quantitative approach to perceived health status: a validation study. Journal Of Epidemiology And Community Health, 34(4), 281-286.
28. Burkhardt, C., Naderson, K (2003). The quality of life scale: reliability, validity and utilisation. Health and Quality of Life Outcomes, 1, 60.
29. Ware, J. E., Kosinski, M., Keller, S.D. (1994). SF-36 Physical and Mental Health Summary Scales - A Users' Manual. Boston: The Health Institute.
30. Ware, J., Kosinski, M., Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Medical Care, 34(3), 220-233.
31. WHOQOL. Group. (1998). WHOQOL Manual. Division of Mental Health and Prevention of Substance Abuse
32. EuroQol--a new facility for the measurement of health-related quality of life. (1990). Health Policy (Amsterdam, Netherlands), 16(3), 199-208.
33. Torrence, G. W., Feeny, D.H., Furlong, W.J., Barr, R.D., Zhang, Y., Wang, Q. (1996). Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. Medical Care, 34, 702-722.
34. Feeney, D., Furlong, W., Torrence, G.W., Goldsmith, C.H. (2002). Multiattribute and single attribute utility functions for Health Utility Index Mark 3 system. Medical Care, 40, 113-128.
35. Brazier, J. E., Roberts, J.R. (2004). The estimation of a preference-based index from the SF-12. Medical Care. Medical Care, 42(9), 851-859.
36. Coast, J., Flynn, T.N., Natarajan, L. et al. (2008). Valueing the ICECAP capability index for older people. Soc Sci Med, 67, 874-882.
37. Charnley, J. (1972). Long term results of low friction arthroplasty of the hip performed as a primary intervention. J Bone Joint Surg Br, 54, 61-76.
38. Harris, W. H. (1969). Traumatic arthritis of the hip after dislocation in acetabular fractures treatment by mold arthroplasty. J Bone Joint Surg Am, 51, 737-755.

39. Dawson, J., Fitzpatrick, R., Carr, A., Murray, D (1996). Oxford Hip Score: questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br., 78, 185-190.
40. Silverman, S. L. (2000). The Osteoporosis Assessment Questionnaire (OPAQ): a reliable and valid disease-targeted measure of health-related quality of life (HRQOL) in osteoporosis. Qual Life Res., 9, 767–774.
41. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J., Stitt L (1988). Validation study of the WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthopaedic Rheumatology, 1, 95-108.
42. Brink, T. L., Yesavage, J.A., Lum, O., Heersema, P., Adey, M.B., Rose, T.L. (1982). Screening tests for geriatric depression. Clinical Gerontologist, 1, 37-44.
43. Hamilton, M. (1960). A rating scale for depression. Journal of Neurology, Neurosurgery and Psychiatry, 23, 56-62.
44. Zung, W. (1965). A Self-Rating Depression Scale. Arch Gen Psychiatry, 12(1), 63-70.
45. Mahoney, F. I., Barthel, D.W. (1965). Functional evaluation: the Barthel Index. Maryland St. Medical Journal, 14, 61-65.
46. Baker, N. L., Cook, M.N., Arrighi, H.M., Bullock, R (2011). Hip fracture risk and subsequent mortality among Alzheimer's disease patients in the United Kingdom, 1988-2007. Age and Ageing, 40(1), 49-54.
47. Pfeiffer, E., Johnson, T.M., Chiofolo, R.C (1981). Functional assessment of elderly subjects in four service settings. Journal of the American Geriatrics Society, 29, 433-437.
48. Macaulay, W., Nellans, K.W., Garvin, K.L., Iorio, R., Healy, W.L., Rosenwasser, M.P.; other members of the DFACTO Consortium. (2008). Prospective randomized clinical trial comparing hemiarthroplasty to total hip arthroplasty in the treatment of displaced femoral neck fractures: winner of the Dorr Award. J Arthroplasty, 23(6), 2-8.
49. Jette, A. M., Cleary, P.D (1986). Functional Disability Assessment. . Physical Therapy, 67, 1854-1859.
50. Lawton, M. P., Moss, M., Fulcomer, M., Kleban, M.H (1982). A research and service oriented multilevel assessment instrument. Journal of Gerontology, 37, 91-99.
51. Katz, J. N., Phillips, C. B., Poss, R., Harrast, J. J., Fossel, A. H., Liang, M. H., & Sledge, C. B. (1995). The validity and reliability of a Total Hip Arthroplasty Outcome Evaluation Questionnaire. Journal of Bone & Joint Surgery - American Volume, 77(10), 1528-1534.
52. Fillenbaum, G. G., Smyer, M.A. (1981). The development, validity, and reliability of the OARS multidimensional functional assessment questionnaire. J Gerontol, 36(4), 426-434.
53. Jones, C. A., Feeney, D.H. (2005). Agreement Between Patient and Proxy Responses of Health-Related Quality of Life After Hip Fracture. Journal of American Geriatrics Society, 53, 1227-1233.
54. Jones, C. A., Feeny, D.H. (2006). Agreement between patient and proxy responses during recovery after hip fracture: evidence for the FIM instrument. Arch Phys Med Rehabil, 87(10), 1382-1387.
55. Parsons, N., Griffin, X. L., Achten, J., Costa, M. L. (2014). Outcome assessment after hip fracture: is EQ-5D the answer? Bone & Joint Research, 3(3), 69-75.
56. Borgquist, L., Nilsson, L.T.,Lindelöw, G.,Wiklund, I.,Thorngren, K.G. (1992). Perceived health in hip-fracture patients: a prospective follow-up of 100 patients. Age and Aging, 21(2), 109-116.
57. Tidermark, J., Zethraeus, N., Svensson, O., Tornkvist, H., Ponzer, S. (2002b). Quality of life related to fracture displacement among elderly patients with femoral neck fractures treated with internal fixation. J Orthop Trauma, 16, 34-38.
58. Tidermark, J., Bergstrom, G., Svensson, O., Tornkvist, H., Ponzer, S (2003a). Responsiveness of the EuroQol (EQ-5D) and SF-36 in elderly patients with displaced femoral neck fractures. Quality of Life Research, Dec; 12(8), 1069-1079.

59. Van Balen, R., Essink-Bot, M.L., Steyerberg, E., Cools, H., Habbema, D.F (2003). Quality of life after hip fracture: a comparison of four health status measures in 208 patients. Disability and Rehabilitation, 25(10), 507-519.
60. Inngul, C., Hedbeck, C.-J., Blomfeldt, R., Lapidus, G., Ponzer, S., & Enocson, A. (2013). Unipolar hemiarthroplasty versus bipolar hemiarthroplasty in patients with displaced femoral neck fractures: a four-year follow-up of a randomised controlled trial. International Orthopaedics, 37(12), 2457-2464.
61. Cranney, A. B., Coyle, D.,Hopman, W.M.,Hum, V.,Power, B.,Tugwell, P.S. (2005). Prospective Evaluation of Preferences and Quality of Life in Women with Hip Fractures. The Journal of Rheumatology, 32(12), 2393-2399.
62. Latham, N. K., Mehta, V., Nguyen, A.M., Jette, A.M., Olarsch, S., Papanicolaou, D., Chandler, J. (2008). Performance-based or self-report measures of physical function: which should be used in clinical trials of hip fracture patients? Arch Phys Med Rehabil, 89(11), 2146-2155.
63. Jones, C. A., Pohar, S. L., Feeny, D. H., & Eng, K. (2014). Longitudinal construct validity of the Health Utilities Indices Mark 2 and Mark 3 in hip fracture. Quality Of Life Research: An International Journal Of Quality Of Life Aspects Of Treatment, Care And Rehabilitation, 23(3), 805-813.
64. Mendelsohn, M. E., Leidl, D.S., Overend, T.J., Petrella, R.J. (2003). Specificity of functional mobility measures in older adults after hip fracture: a pilot study. Am J Phys Med Rehabil, 82(10), 766-774.
65. Tidermark, J., Bergström, G. (2007). Responsiveness of the EuroQol (EQ-5D) and the Nottingham Health Profile (NHP) in elderly patients with femoral neck fractures. Qual Life Res., 16(2), 321-330.
66. Hobart, J. C., Cano, S.J (2010). Effect sizes can be misleading: Is it time to change the way we measure change? Quality of Life Research, 19, 41.
67. Griffin, X. L., Parsons, N., Achten, J., Fernandez, M., Costa, M.L. (2015). Recovery of health-related quality of life in a United Kingdom hip fracture population: the Warwick Hip Trauma Evaluation - a prospective cohort study. Bone Joint J. 2015 Mar;97-B(3):372-82. doi: 10.1302/0301-620X.97B3.35738.
68. Buecking, B., Struewer, J., Waldermann, A., Horstmann, K., Schubert, N., Balzer-Geldsetzer, M., Dodel, R., Bohl, K., Ruchholtz, S., Bliemel C. (2014). What determines health-related quality of life in hip fracture patients at the end of acute care?--a prospective observational study. Osteoporos Int. 2014 Feb;25(2):475-84. doi: 10.1007/s00198-013-2415-5. Epub 2013 Jun 20.
69. Mishra, V., Thomas, G., Sibly, T.F. (2004). Results of displaced subcapital fractures treated by primary total hip replacement. Injury, 35(2), 157-160.
70. Randell, A. J., Nguyen, T.V.,Bhalerao, N.,Silverman, S.L.,Sambrook, P.N.,Eisman, J.A. (2000). Deterioration in Quality of Life Following Hip Fracture: A Prospective Study. Osteoporosis International, 11, 460-466.
71. Rohde, G., Haugeberg, G., Mengshoel, A.M., Moum, T., Wahl, A.K. (2010). Two-year changes in quality of life in elderly patients with low-energy hip fractures. A case-control study. BMC Musculoskelet Disord, 11, 226.
72. Hall, S. E., Williams, J.A., Senior, J.A., Goldswain, P.R., Criddle, R.A. (2000). Hip fracture outcomes: quality of life and functional status in older adults living in the community. Aust N Z J Med, 30(3), 327-332.
73. Shepherd, S. M., Prescott, R.J. (1996). Use of standardised assessment scales in elderly hip fracture patients. J R Coll Physicians Lond, 30(4), 335-343.
74. Griffiths, F., Mason, V., Boardman, F., Dennick, K., Haywood, K. et al. (2015). Evaluating recovery following hip fracture: a qualitative study of what is important to patients BMJ Open, Jan 6, 5(1), e005406.

75. Haywood, K., Brett, J., Salek, S., Marlett, N., Penman, C., Shklarov, S., Norris, C., Jose Santana, M., Staniszewska, S (2015). Patient and Public Engagement in Health-Related Quality of Life and Patient-Reported Outcomes Research: What's Important and Why Should We Care? Quality of Life Research., May; 24(5), 1069-1076.
76. Seitz, D. P., Adunuri, N., Gill, S.S., Rochon, P.A. . (2011). Prevalence of Dementia and Cognitive Impairment Among Older Adults With Hip Fractures. Journal of the American Medical Directors Association, Mar 7.
77. Allan, L. M., Ballard, C.G., Rowan, E.N., Kenny, R.A (2009). Incidence and prediction of falls in dementia: a prospective study in older people. PLoS One, 4(5), e5521.
78. Jones, C. A., Feeny, D.H. (2006). Agreement Between Patient and Proxy Responses During Recovery After Hip Fracture: Evidence for the FIM Instrument. Archives of Physical Medicine and Rehabilitation, 87(Oct 10), 1382–1387.
79. Fitzpatrick, R., Davey, C., Buxton, M.J., and Jones, D.R. (1998). Evaluating patient-based outcome measures for use in clinical trials. Health Technology Assessment (Winchester, England), 2, 14.
80. Binder, E. F., Brown, M., Sinacore, D.R. (2004). Effects of Extended Outpatient Rehabilitation After Hip Fracture A Randomized Controlled Trial. JAMA, 292(7), 837-846.
81. Pusic, A. L., Chen, C.M., Cano, S., Klassen, A., McCarthy, C., Collins, E.D., Cordeiro, P.G. (2007). Measuring quality of life in cosmetic and reconstructive breast surgery: a systematic review of patient-reported outcomes instruments. Plast Reconstr Surg., Sep 15;120(4), 823-837.
82. Garratt, A., Schmidt, L., Mackintosh, A., Fitzpatrick, R. (2002). Quality of life measurement: bibliography study of patient assessed health outcome measures. BMJ, 324(7351), 1417.
83. U.S. Department of Health and Human Services Food and Drug Administration. (2009). Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf.
84. Sackett, D. L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., Richardson, W.S. (1996). Evidence based medicine: what it is and what it isn't. BMJ, 312(7023), 71-72.
85. Boonen, S., Autier, P.,Barette, M.,Vanderschueren, D.,Lips, P.,Haentjens, P. (2004). Functional outcome and quality of life following hip fracture in elderly women: a prospective controlled study. Osteoporos Int., 15, 87-94.
86. Haywood, K. L., Griffin, X.L., Achten, J., Costa ML (2014). Developing a care outcome set for hip fracture trials. Bone Joint J. , 96-B(8), 1016-1023.
87. Connell, J., O'Cathain, A., Brazier, J (2014). Measuring quality of life in mental health:are we asking the right questions? Soc Sci Med, Nov;120, 12-20.
88. Duppils, G. S., Wikblad, K. (2004). Cognitive Function and Health-Related Quality of Life After Delirium in Connection With Hip Surgery: A Six-Month Follow-Up. Orthopaedic Nursing, 23(3), 195-203.

Table 1: Characteristics of reviewed generic and specific PROMs evaluated in the hip fracture population (total = 28)

PROM *(Author; web-link; completion format[b])* [c]	n[a]	Construct	Recall Period	Domains (number of items)	Response options (range)	Score range	Administration [b] *(mode and time)*
Generic – health status (12)							
Profile measures (6/12)							
COOP-WONCA Charts (Nelson et al, 1987)[1] [c] (6 items) http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239076/?page=3	1	Health-related Quality of Life and functional capacity	Current	6 domains (6 items; 1 item per domain): Daily Activities (1) Feelings (mental well-being) (1) Physical Fitness (1) Social Activities (1) Overall Health (1) Change in health status (1)	5-point descriptive; where 1= best, 5= worst health status. Pictorial health charts used for functional health status items: physical fitness, feelings, daily activities, social activities; plus for change in health and overall health.	Domains scores range 1 to 5, where higher scores reflect worse health. Score profile (per domain) not index.	*Self-complete or interview-administration* <5mins *Not reported in elderly hip fracture population*
Nottingham Health Profile (NHP) (Hunt et al, 1980)[2] http://www.proqolid.org/instruments/nottingham_health *Self-completion or interview administered*	4	Health-related Quality of Life (HRQL): Aims to provide a brief indication of an individual's perceived emotional, social, and physical health problems	At the moment	Total: 45 items Part 1: 6 domains (38 items) Physical mobility (8) Pain (8) Sleep (5) Social isolation (5) Emotional reactions (9) Energy levels (3) Part 2 (7 items): Effect of health problems on occupation, jobs around the house, personal relationships, social life, sex life, hobbies and holidays	Dichotomous: Yes/No Each item is weighted: weights are derived from patients and non-patients.	Domain scores range 0-100, where higher scores reflect worse health. Score profile (per domain) not index.	*Self-complete or interview-administration* 5-10 mins *Not reported in elderly hip fracture population*

Quality of Life Scale (QoLS) (Burkhardt and Anderson, 2003)[3] http://www.hqlo.com/conte nt/1/1/60 burckhac@ohsu.edu *Self-completion or interview administered*	1	Quality of Life		5 domains (16 items): Material and physical well-being Relationships with other people Social, community and civic activities Personal development and fulfilment Recreation	7-point descriptive: range 7= delighted 6= pleased 5= mostly satisfied 4= mixed 3= mostly dissatisfied (3 2= unhappy 1= terrible	Item summation to produce an index score. Score range 16 to 112, where higher scores indicate better quality of life.	*Self-complete or interview-administration* Approx 5 minutes. *Not reported in elderly hip fracture population*
Short Form 36-item Health Survey (SF-36) (version 1 (v1)) [Ware, Kosinski, & Keller, 1994][4] http://www.sf-36.org/ http://www.sf-36.org/tools/sf36.shtml#CONSTRUCT *Self-completion or interview administered*	17	Health Status	Recall: standard 4 weeks, acute 1 week	8 domains (36 items) Bodily pain (BP)(2) General health (GH)(5) Mental health (MH) (5) Physical functioning (PF)(10) Role limitation-emotional (RE)(3) Role limitation-physical (RP)(4) Social functioning (SF)(2) Vitality (V)(4)	Categorical: 2-6 options	Requires scoring algorithm. Creates a domain profile: each domain score ranges 0-100, where higher scores indicate better health. Two summary scores: Physical (PCS), Mental (MCS): norm-based scores calculated (mean 50, sd 10), where scores higher than 50 suggest a health state better than the population mean.	*Self-complete or interview-administration* 15 to 30 mins *Not reported in elderly hip fracture population*
Short Form 12-item Health	3	Health Status: 8	Recall:	8 domains (12 items)	Categorical: 2-6 options	Requires scoring	*Self-complete or*

Table 1: Characteristics of Reviewed Measures TC R1 130616

Survey (SF-12)(v1) [Ware, Kosinski, Keller 1996][5] http://www.sf-36.org/ http://www.sf-36.org/tools/sf36.shtml#CONSTRUCT *Self-completion or interview administered*		domains	standard 4 weeks, acute 1 week	Bodily pain (BP)(1) Energy/Vitality (V)(1) General health (GH)(1) Mental health (MH)(2) Physical functioning (PF)(2) Role limitation-emotional (RE)(2) Role limitation-physical (RP)(2) Social functioning (SF)(1)		algorithm (US population). Produces two summary scores: Physical (PCS-12), Mental (MCS-12): norm-based scores calculated (mean 50, sd 10), where scores higher than 50 suggest a health state better than the population mean.	*interview-administration* 5 to 10 mins *Not reported in elderly hip fracture population*
World Health Organisation Quality of Life questionnaire (WHOQOL) (The WHOQOL Group, 1998)[6] http://www.who.int/mental_health/media/68.pdf WHOQOL@who.int *Self-complete or interview-administration*	1	WHOQoL 100: General QOL International cross-culturally comparable quality of life assessment instrument. It assesses the individual's perceptions in the context of their culture and value systems, and their personal goals, standards and concerns.	4-weeks	6 domains (24 facets; 100 items) Physical health (energy and fatigue; pain and discomfort; sleep and rest) Psychological (body image; negative / positive feelings; self-esteem; thinking, learning, memory, concentration) Level of independence (mobility, ADL, dependency (medicine; aids), work capacity) Social relations (personal relationships; social support; sexual activity) Environment (includes finance; freedom; social care; environment; transport)	5-point descriptive: range 1= very poor/ very dissatisfied/ an extreme amount / always to 5= very good / very satisfied / not at all /completely / never	Facet scores: item summation (reverse score negative items): where higher scores suggest better quality of life. Domain scores: facet summation (domain score divided by no. of facets to facilitate comparison between domains): where higher scores suggest better quality of life.	*Self-complete or interview-administration* 100 items: estimated at 15mins to 20 minutes *Not reported in elderly hip fracture population*

Table 1: Characteristics of Reviewed Measures TC R1 130616

				Spirituality/religious/ personal beliefs (single facet – religion, spirituality and personal beliefs)			
Single item VAS (1/12)							
EuroQoL EQ-5D Visual Analogue Scale (VAS) (EuroQoL Group, 1990)[7] http://www.euroqol.org/ *Self-completion or interview administered*	5	Quality of Life: 'Your own health state' 'EuroQoL thermometer'	Today	General quality of life (1 item)	Vertical Visual Analogue Scale (VAS) 'Thermometer': anchored 0 'worst imaginable health' and 100 'best imaginable health state'.	0-100, where 0 is worst health and 100 is best health.	*Self-complete or interview-administration* 2 minutes *Not reported in elderly hip fracture population*
Preference-based Utility Measures (4/12)							
EuroQoL EQ-5D (3L) (EuroQoL Group, 1990)[7] http://www.euroqol.org/ *Self-completion or interview administered*	16	Quality of Life	Today	5 domains (5 items) Mobility Self-care Usual activities Pain/discomfort Anxiety/depression	3-point descriptive: no problems some problems severe problems.	Utility index value (society assigned value system algorithm): -0.59 to 1.00 where 1.00 is perfect quality of life, 0 is death, and <0 is a health state worse than death.	*Self-complete or interview-administration* 2 to 5 mins *Not reported in elderly hip fracture population*
Health Utility Index – 2 (HUI-2) (Torrence et al., 1996)[8]	3	Multi-attribute health status classification system	Varies: 1/52,2/52,	7 domains (attributes) (7 items): Sensation (vision, hearing, speech)	3 to 5 descriptive per attribute / domain; where 1 is best health.	Standard algorithms. 0 to 1.00 where 1.00 is perfect	*Self-complete or interview-administration*

Table 1: Characteristics of Reviewed Measures TC R1 130616

http://www.healthutilities.com/hui2.htm *Self-completion or interview administered*		Describes the comprehensive health state of an individual as a 7-element vector *Originally developed to assess outcomes among survivors of childhood cancers)*	4/52	Mobility Emotion Cognitive Self-care Pain Fertility		QoL	*Not reported in elderly hip fracture population* *Likely to be relative quick*
Health Utility Index – 3 (HUI-3) (Feeney et al, 2002)[9] http://www.healthutilities.com/hui3.htm *Self-completion or interview administered*	2	Multi-attribute health status classification system Describes the comprehensive health state of an individual as an 8-element vector	Current	8 domains (attributes) (8 items): Vision Hearing Speech Ambulation Dexterity Emotion Cognition Pain	1 to 5 or 1 to 6 descriptive response options per attribute / domain; where 1 is best health, 5 or 6 is worst health.	Standard algorithms. 0 to 1.00 where 1.00 is perfect QoL	*Self-complete or interview-administration* 5 mins *Not reported in elderly hip fracture population*
SF-6D (Brazier et al, 2004)[10] *Self-complete or interview-administration* *https://www.shef.ac.uk/scharr/sections/heds/mvh/index* *Self-completion or interview administered*	1	6-dimensional health state classification; preference-based measure of health derived from the SF-36.	Recall: standard 4 weeks, acute 1 week	Description of health derived from 6 multi-level dimensions from the SF-36 or SF-12 (6 items).	Categorical: 2-6 options General population generated preference weights (standard gamble)	Standard algorithms. Utility score: 0 to 1.00 where 1.00 is perfect QoL	Requires completion of SF-36 or SF-12 to generate SF-6D score. *Not reported in elderly hip fracture population*
Capability measure (1/12)							
ICEpop CAPability measure for Older people	1	Profile measure of capability and general	Current	5 descriptive attributes important to older people	1 to 4 ordered levels of capability: where 1 is	Scaled using country-	5 mins

Table 1: Characteristics of Reviewed Measures TC R1 130616

(ICECAP-O) (Al-Janabi, Flynn, Coast 2008)[11] (Coast et al., 2008)[12] http://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/HE/ICECAP/ICECAP-O/index.aspx *Self-complete or interview administered*		well-being for use in economic evaluations		(10 items): Attachment (love and friendship) (2); Security (thinking about the future without concern) (2); Role (doing things that make you feel valued) (2); Enjoyment (enjoyment and pleasure) (2); Control (independence) (2)	lowest capability and 4 is the highest.	index values / algorithms. 0 to 1.00 where 1.00 is full capacity and 0 is no capacity	*Not reported in elderly hip fracture population*
Specific measures (16)							
Hip-specific (3): Surgeon-based assessment (2/3)							
Charnley Hip Score* (Charnley, 1972) [13] *Assessment entirely by the clinician – to represent the opinion / perspective of both clinician and patient* http://www.bjj.boneandjoint.org.uk/content/54-B/1/61.full.pdf	4	For the assessment of hip surgery. To evaluate hip disabilities and methods of treatment	Current	3 domains (3 items): Pain (1) - severity Mobility (1): sum of the range of movement (ROM) in the 3 standard directions (flexion; extension; abduction). Walking (1): hip function re ability to walk. Detail very limited.	Pain: 6-point descriptive; where 1 is the worst pain (severe and spontaneous) and 6 is the best (No pain). Mobility: 6-point categorical; where 1= severely limited (0-30 degrees) and 6 is good movement (260 degrees) *(No further detail re ROM).* Walking: 6-point descriptive; where 1= bedridden and 6= normal.	3 domain scores: where lower scores (1) indicate greater pain / most limited mobility / greater walking disability.	5 mins *Not reported in elderly hip fracture population*
Harris Hip Score* (HHS) (Harris, 1969)[14]	8	For the assessment of hip surgery. To	Current	4 domains (10 items)	Each item has a unique numerical score which	Score range 0 – 100: where	5 mins

Table 1: Characteristics of Reviewed Measures TC R1 130616

http://www.orthopaedicscore.com/scorepages/harris_hip_score.html *Clinician-based outcome measure administered by a qualified health professional (Nilsdotter & Bremander, 2011)* ***Pain and function items could be patient reported - often not clear how the measure was completed?***		evaluate hip disabilities and methods of treatment. 8 questions and a physical examination (ROM of hip)		Section 1 (8 items): Pain: severity, effects on activities, need for medication. Function: daily activities (stair use, using public transport, sitting, managing shoes/socks), gait (limp, support needed, walking distance) Section 2 and 3 (3 items) Absence of deformity (hip flexion, adduction, internal rotation and extremity leg discrepancy) Range of movement (ROM): hip flexion, abduction, external/internal rotation, adduction.	corresponds to a descriptive response. Number of response options and number of points varies by item. Section 1: 1.1 Pain: 6 descriptive options range 'None / ignores it' to 'Totally disabled, crippled, pain in bed, bedridden'. 1.2 Support (when walking): 6 descriptive options range 'None' to 'Two crutches or not able to walk'. 1.3 Distance walked: 5 descriptive options range 'Unlimited' to 'Bed and chair only'. 1.4 Presence of Limp: 4 descriptive options range 'None' to 'Severe or unable to walk' 1.5 Activities – ability to put on shoes / socks: 3 options range 'with ease' to 'unable to fit or tie' 1.6 Stairs: 4 options range 'normally without using a railing' to 'unable to do stairs'	higher scores suggest less dysfunction. <70 Poor 70-79 Fair 80-89 Good 90-100 Excellent	*Not reported in elderly hip fracture population*

Table 1: Characteristics of Reviewed Measures TC R1 130616

					1.7 Public transportation: 2 options range 'Able to use transportation (bus)' to 'Unable to use public transportations (bus)' 1.8 Sitting: 3 options range 'comfortably, ordinary chair for one hour' to 'unable to sit comfortably on any chair'. Sections 2 and 3: select ROM range for specific motion (hip flexion, abduction, external rotation, adduction).		
Patient-completed assessment (1/3)							
Oxford Hip Score (OHS) (Dawson et al, 1996)[15] http://phi.uhce.ox.ac.uk/pdf/OxfordScores/hip_score_guide.pdf. *Self-completion or interview administered*	3	'Intervention-specific measure' – to assess outcome after Total Hip Replacement (THR) Assess pain and function of the hip in relation to daily activities.	4 wks	12 items: Function (6) Pain (6)	(Revised scoring system 2007): 5-point descriptive scale: range 0= worst health to 4= best health.	Item summation. Range 0 to 48, where 48 is best health status.	Approx 5 mins *Not reported in elderly hip fracture population*
Disease-specific (2)							
Osteoporosis Quality of Life Questionnaire (version 2) (OPAQ-2) (Silverman, 2000)[16]	1	Health-related quality of life in post-menopausal women with osteoporosis and fracture.	2 weeks	54 items grouped into 14 domains 'across 4 major dimensions of health status': Physical function (6	5-point Likert scale: where 1= no impairment and 5= constant impairment	Item summation. Profile / domain scores. Normalisation procedure so that	20-30 mins *Not reported in elderly hip*

Table 1: Characteristics of Reviewed Measures TC R1 130616

Self-completion or interview administered		Shortened version of original the original OPAQ (79 items (across 18 domains and 4 overall dimensions of health) +5 (overall well-being) items = 84 items (+ 18 items on satisfaction) Also OPAQ SV (short version: 34 items across 3 domains (physical function, emotional status, symptoms)		domains): walking/bending, standing/sitting, dressing/reaching, household/self-care, transfers, usual work. Emotional status (4 domains): fear of falls, level of tension, body image, independence. Symptoms: 2domains: back pain, fatigue. Social interaction (2 domains): social activity and support of family and friends. Plus 6 additional items on general health. Overall HRQoL and change in HRQoL over the last year.		all domains are scored 0-10, where 0= worst possible health status Four dimension scores: domain scores summed within the same dimension and normalised to a 0-100 score, were 0= worst health status,	*fracture population*
Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (Bellamy et al, 1988)[17] Prof Nick Bellamy: n.bellam@ug.edu.a *Self, interview or telephone administration* *www.womac.org*	4	To assess pain, stiffness, and physical function in patients with hip and / or knee osteoarthritis (OA) *Self, interview or telephone administration*	? unclear	3 domains (24 items) Pain (5): during walking, using stairs, in bed, sitting or lying, and standing Stiffness (2): after first waking and later in the day Physical Function (PF) (17): stair use, rising from sitting, standing, bending, walking, getting in / out of a car, shopping, putting on / taking off socks, rising from bed, lying in bed, getting in / out of bath, sitting, getting	5-point categorical (0 to 4); where 0= is extreme pain / stiffness / impairment and 4 is none.	Item summation per domain. Pain: range 0-20 Stiffness: range 0-8 PF: range 0-68 Where higher scores suggest between health status. Index score (%) 0-100%: where higher scores suggest better health status.	12 mins *Not reported in elderly hip fracture population*

Table 1: Characteristics of Reviewed Measures TC R1 130616

http://www.rheumatology.org/practice/clinical/clinicianresearchers/outcomes-instrumentation/WOMAC.asp				on / off toilet, heavy household duties, light household duties			
Domain-specific (11)							
Emotional well-being (3/11)							
Geriatric Depression Scale (GDS) (Brink and Yesavage, 1982)[18] http://www.sabp.nhs.uk/Documents/D1.3d6.pdf *Self-completion or interview administered*	4	Depression – suitable for screening for depressive symptoms and monitoring treatment. Revised 15 and 4-item short forms also available http://www.thementalelf.net/mental-health-conditions/depression/the-geriatric-depression-scale-is-the-best-screening-tool-for-depression-in-older-people-in-acute-hospital-settings/	Past week	30 items (15 in short version)	Dichotomous: Yes/No	Item summation. Index score ranges 1 to 30, where lower scores indicate less depression: 0 to 10 = normal, 11 to 20 = mild depression, 21 to 30 = moderate to severe depression	8-10 mins (short version 5-7 mins) *Not reported in elderly hip fracture population*
Hamilton Rating Scale for Depression (HAM-D) (Hamilton, 1960)[19] *Interview-administered (trained)*	1	Severity of depression	Current	21 items, but scoring based on items 1-17	Items 1-3, 7-11, 15, 16, 19, 20: 5-point descriptive scale (0-4), where 0= best and 4= worst Items 16, 20: 4-point descriptive scale (0-3) where 0= best and 3= worst	Item summation. Index score range 0 to 50, where lower scores indicate less depression: Score range: 14-18 = moderate depression,	10-15 mins *Not reported in elderly hip fracture population*

Table 1: Characteristics of Reviewed Measures TC R1 130616

http://healthnet.umassmed.edu/mhealth/HAMD.pdf					Items 4-6, 12-14, 17, 18, 21: 3-point descriptive scale (0-2) where 0= best and 2= worst	19-22 = severe depression, ≥ 23 = very severe depression	
Zung Depression Inventory (Zung DI) (Zung, 1965)[20] *Self-completion or interview administered*	1	Severity of depression	Current	20 items.	4-point categorical scale (1 to 4): range 'None' / 'A little of the time' to 'Most' / 'All of the time'. Reverse scoring for items: 2,5,6,11,12,14,16-18,20.	Item summation: Index score range 20-80, where higher scores indicate more depression. Convert to a 25-100 scale by dividing total by 0.8: lower scores suggest better health / less depression. Interpretation: <50 normal; 50-59: minimum/ mild depression; 60-69 moderate /marked depression; >70 severe depression.	10-15 mins. *Not reported in elderly hip fracture population*
Activities of Daily Living / Instrumental Activities of Daily Living (ADL/IADL)(8/11)							
Barthel Index (BI)	10	Functional independence in personal care and	Current	10 items: Personal care and Mobility	Rated in terms of ability of patient to complete the activity independently,	Item summation: range 0 to 100, where 0 to	Clinician-completed 2-5 minutes

Table 1: Characteristics of Reviewed Measures TC R1 130616

(Mahoney and Barthel, 1965)[21] http://www.healthcare.uiowa.edu/igec/tools/function/barthelADLs.pdf		mobility			with some assistance, or is dependent on help. 0, 5,10 Or 0, 5,10,15 0 is worst, 10 or 15 is best.	20=Total dependency, 21-60=severe independency 61 -90= moderate dependency, & 91-99 slight dependency. 100=independent	Self-completed in approx 10 mins *Not reported in elderly hip fracture*
Modified-BI (MBI) (Granger et al, 1979; Fortinsky 1981)[22; 23] http://a4ebm.org/sites/default/files/Measuring%20Health.pdf *Clinician-completed – from direct observation or from medical records* *May also be self-completed*	2	Functional dependency in personal care and mobility	Current	15 items: Personal care and Mobility	4 response options: Dependent: Null (IV), Helper (III); Independent: Limited (II). Intact (I).	Item summation: index score. Range 0 to 100, where 0 is worst score and 100 is best score: 0 to 20 = total dependency; 21-60 = severe dependency; 61-90 = moderate dependency; 91-99 = slight dependency; 100 = independent	Clinician-completed 2-5 minutes Self-completed approx 10 mins *Not reported in elderly hip fracture population*
Functional Activities Index (FAI) (Pfeiffer et al, 1981)[24] *Shortened version of OARS OMFAQ*	3	Functional status Revision of the OMFAQ (OARS Multi-dimensional Functional Assessment	Current (part B mixed)	5 domains (30 items): ADL impairment Economic resources Mental health Physical health	3-point descriptive: range 0= worst function to 2= best function.	Item summation: Index score range 0 to 60, where 0 is worst function.	30 to 40 mins (vs > 40 mins for the OMFAQ)

Table 1: Characteristics of Reviewed Measures TC R1 130616

Interview-administered (clinician)		Questionnaire): items about medical services removed; items about life satisfaction and self-esteem added.		Social resources			
Functional Independence Measure (FIM) (Keith, Granger and Hamilton, 1987)[25] http://www.rehabmeasures.org/lists/rehabmeasures/dispform.aspx?id=889 *Interview-administered (clinician or non-clinician). NOT self-completed.* *Classified as a PRO – but NOT patient completed* *For copy of FIM (and copyright) contact:* Uniform Data System for Medical Rehabilitation 270 Northpointe Parkway, Suite 300 Amherst, New York 14228 (716) 817-7800 FAX (716) 568-0037 email: info@udsmr.org web site: http://www.udsmr.org	21	Activities of daily living (ADL): assessment underpinned by reference to the International Classification of Impairment, Disabilities and Handicaps. Assesses disability level and the amount of assistance required for an individual to carry out activities of daily living Includes a focus on the burden of care	Current	6 domains (18 items): Cognitive tasks (CoGFIM)(5 items): social cognition, problem solving, communication. Motor tasks (MotorFIM)(13 items): self-care, sphincter control, mobility, locomotion.	7-point ordinal scale: range from complete dependency (scores 1 and 2) to complete independence (score 7) 1 = total dependency / assistance 2 = maximum assistance 3 = moderate assistance 4 = minimal contact assistance 5 = supervision or setup 6 = modified independence 7 = complete independence	Item summation: Index score range 18 to 126; where 18 is greatest dependency in ADL, and 126 is independent in ADL.	30 to 60 minutes. *Not reported in elderly hip fracture population*
Functional Status Questionnaire (FSQ)	1	Physical, psychological and social role	1 mth	Total 34 items:	4, 5 and 6-point descriptive ratings: range	Item summation. 6 summary scores	15 mins

Table 1: Characteristics of Reviewed Measures TC R1 130616

(Jette, 1986)[26] *Self-completed*		functioning in ambulatory patients. *Designed as a screening assessment for disability and to monitor functional change in a primary care setting.*		6 domains (28 items): Physical Function: ADL (3) and Instrumental ADL (6); Psychological function (5); Work performance (6); Social activity (3); Quality of social interaction (5) 6 additional items: work status, bed disability days, activity reductions, satisfaction with sexual relationships, interpersonal relationships, feelings about health.	1 (all of the time) to 4/5/6 (none of the time), where higher score is better function.	and 6 single item scores. Also standardised to 0 to 100, where higher scores are best functional status. Includes one-page summary report highlighting areas of clinical concern.	*Not reported in elderly hip fracture population*
Lawton Instrumental Activities of Daily Life Scale (Lawton-IADL) (Lawton and Brody, 1969)[27] *http://www.strokecenter.org/wp-content/uploads/2011/08/lawton_IADL_Scale.pdf* *Interview-administered (clinician)* **Copyright** (c) The Gerontological Society of America	5	Functional impact of emotional, cognitive, and physical impairment.	Current	1 domain (ADL) (8 items): Ability to use telephone Shopping Laundry Food preparation Mode of transportation Responsibility for own medication Housekeeping Ability to handle finances	Dichotomous response options: 0 (impaired) or 1 (unimpaired)	Item summation: *For females:* range 0 to 8, where higher scores suggest better levels of ADL / less impairment. *For males:* range 0 to 5. Score excludes food preparation, laundering and housekeeping.	5 mins *Not reported in elderly hip fracture population*

Table 1: Characteristics of Reviewed Measures TC R1 130616

Katz Index of Independence in Activities of Daily Life (Katz ADL) (Katz et al, 1963)[28] *Clinician-completed – through observation and interview (training required)* http://www.npcrc.org/usr_doc/adhoc/functionalstatus/Katz%20Index%20of%20Independence%20in%20Activities%20of%20Daily%20Living.pdf	8	Ability to complete basic activities of daily living and to live independently. Focus on individuals with chronic illness and the elderly (originally developed following an evaluation of older people with hip fracture).	2-weeks	1 domain (ADL) (6 items): Bathing Dressing Using the toilet Transferring from bed to chair Continence Feeding Items reflect a hierarchical order of functional difficulty in ADL. Independence in these activities is assessed.	3-point scale of independence: range 'No assistance' (0) to 'Maximum assistance' (2). Response per item includes a detailed descriptive response. Scores translated into a 'dependent/independent' classification: A,B,C,D,E,F,G,O Where A = total independence, and G = total dependence; O = dependent in 2 ADLs	*Original scoring:* Overall level of performance summarised on an 8-point scale: 8 levels of dependency ranging from A = independent, to G = total dependence; O=dependent in 2 or more. *Simplified scoring:* Number of activities in which individual is dependent on scale 0 to 6, where 0=independent and 6=dependent. Higher scores suggest greater dependency.	10-15 mins *Not reported in elderly hip fracture population*
OARS Multi-dimensional Functional Assessment Questionnaire (OMFAQ) (Fillenbaum, 1983)[29] *Interview-administration only (training required – 2 day training course)*	1	An assessment of overall functional status and service use of adults – in particular older people.	Current (part B mixed)	Part A: 5 domains (120 items): ADL (14; IADL 7/14), Economic resources (15), Mental health (21), Physical health (16) Social resources (9) Demographic items (11) Informant items (10)	Categorical, some with written answers. Interviewer: 5-point categorical	Part A: 5 summary scores or coding scheme (algorithm) Index: Cumulative Impairment Score 5-30, where 30 is maximum	Part A: 30mins Part B:45mins *Not reported in elderly hip fracture population*

Table 1: Characteristics of Reviewed Measures TC R1 130616

http://centerforaging.duke.e du/services/141				Interview section: Interview-specific (4), Interviewer assessments (15), Interview ratings (5) Short Portable Mental Status Questionnaire (10) Part B: Services Assessment (24)		impairment Part B: where higher scores suggest better care services .	

Footnote:

[a] Number of evaluations in elderly hip fracture population and included in review.

[b] ***Completion format:***

- *Clinician / surgeon completed only (3/29): Charnley Hip Score, Harris Hip Score, Katz ADL.
- Interview-administration only (5/29): HAM-D, FAI, FIM, Lawton ADL, OMFAQ.
- Clinician OR self-completed (2/29): Barthel Index and the Modified Barthel Index.
- Self or interview-completion (18/29): COOP-WONCA, EQ-VAS, NHP, QoLS, SF-6D, SF-36, SF-12, WHOQOL-BREF, EuroQoL EQ-5D, HUI-2, HUI-3, ICECAP-O, OHS, OPAQ-2, WOMAC, BDI-II, GDS, PGCMS, Zung DI, FSQ.

[c] References:

1. Nelson, E., Wasson, J., Kirk, J., Keller, A., Clark, D., Dietrich, A., Stewart, A., & Zubkoff, M. (1987). Assessment of function in routine clinical practice: description of the COOP Chart method and preliminary findings. Journal Of Chronic Diseases, 40 Suppl 1, 55S-69S.
2. Hunt, S. M., McKenna, S. P., McEwen, J., Backett, E. M., Williams, J., & Papp, E. (1980). A quantitative approach to perceived health status: a validation study. Journal Of Epidemiology And Community Health, 34(4), 281-286.
3. Burkhardt, C., Naderson, K (2003). The quality of life scale: reliability, validity and utilisation. Health and Quality of Life Outcomes, 1, 60.
4. Ware, J. E., Kosinski, M., Keller, S.D. (1994). SF-36 Physical and Mental Health Summary Scales - A Users' Manual. Boston: The Health Institute.
5. Ware, J., Kosinski, M., Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Medical Care, 34(3), 220-233.
6. WHOQOL. Group. (1998). WHOQOL Manual. Division of Mental Health and Prevention of Substance Abuse
7. The EuroQol Group. (1990). EuroQol—a new facility for the measurement of health related quality of life. Health Policy (Amsterdam, Netherlands), 16, 199-208.

Table 1: Characteristics of Reviewed Measures TC R1 130616

8. Torrence, G. W., Feeny, D.H., Furlong, W.J., Barr, R.D., Zhang, Y., Wang, Q. (1996). Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. Medical Care, 34, 702-722.
9. Feeney, D., Furlong, W., Torrence, G.W., Goldsmith, C.H. (2002). Multiattribute and single attribute utility functions for Health Utility Index Mark 3 system. Medical Care, 40, 113-128.
10. Brazier, J. E., Roberts, J.R. (2004). The estimation of a preference-based index from the SF-12. Medical Care. Medical Care, 42(9), 851-859.
11. Al-Janabi H, F. T., Coast J. (2012). Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Quality of Life Research, 21, 167-176.
12. Coast, J., Flynn, T.N., Natarajan, L. et al. (2008). Valueing the ICECAP capability index for older people. Soc Sci Med, 67, 874-882.
13. Charnley, J. (1972). Long term results of low friction arthroplasty of the hip performed as a primary intervention. J Bone Joint Surg Br, 54, 61-76.
14. Harris, W. H. (1969). Traumatic arthritis of the hip after dislocation in acetabular fractures treatment by mold arthroplasty. J Bone Joint Surg Am, 51, 737-755.
15. Dawson, J., Fitzpatrick, R., Carr, A., Murray, D (1996). Oxford Hip Score: questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br., 78, 185-190.
16. Silverman, S. L. (2000). The Osteoporosis Assessment Questionnaire (OPAQ): a reliable and valid disease-targeted measure of health-related quality of life (HRQOL) in osteoporosis. Qual Life Res., 9, 767–774.
17. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J., Stitt L (1988). Validation study of the WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthopaedic Rheumatology, 1, 95-108.
18. Brink, T. L., Yesavage, J.A., Lum, O., Heersema, P., Adey, M.B., Rose, T.L. (1982). Screening tests for geriatric depression. Clinical Gerontologist, 1, 37-44.
19. Hamilton, M. (1960). A rating scale for depression. Journal of Neurology, Neurosurgery and Psychiatry, 23, 56-62.
20. Zung, W. (1965). A Self-Rating Depression Scale. Arch Gen Psychiatry, 12(1), 63-70.
21. Mahoney, F. I., Barthel, D.W. (1965). Functional evaluation: the Barthel Index. Maryland St. Medical Journal, 14, 61-65.
22. Granger, C. V., Albrecht, G.L., & Hamilton, B.B. . (1979). Outcome of comprehensive medical rehabilitation: measurement by PULSES Profile and the Barthel Index. . Archives of Physical Medical Rehabilitation, 60, 145-154.
23. Fortinsky, R. H., Granger, C.V., & Seltzer, G.B. (1981). The use of functional assessment in understanding home care needs. Medical Care, 19, 489-497.
24. Pfeiffer, E., Johnson, T.M., Chiofolo, R.C (1981). Functional assessment of elderly subjects in four service settings. Journal of the American Geriatrics Society, 29, 433-437.
25. Keith, R. A., Granger, C.V., Hamilton, B.B., Sherwin, F.S. (1987). The functional independence measure: a new tool for rehabilitation. Adv Clin Rehabil., 1, 6-18.
26. Jette, A. M., Cleary, P.D (1986). Functional Disability Assessment. . Physical Therapy, 67, 1854-1859.
27. Lawton, M. P., and Brody, E.M. . (1969). Assessment of older people: Self-maintaining and instrumental activities ofdaily living. Gerontologist, 9, 179-186.
28. Katz, S., Ford, A.B., Moskowitz, R., et al. (1963). Studies of illness in the aged: The index of ADL – a standardised measure of biological and psychosocial function. Journal of the American Medical Association, 185, 914-919.
29. Fillenbaum, G. G., Smyer, M.A. (1981). The development, validity, and reliability of the OARS multidimensional functional assessment questionnaire. J Gerontol, 36(4), 426-434.

Table 1: Characteristics of Reviewed Measures TC R1 130616

Table 2 Methodological quality (COSMIN) and investigated measurement properties per PROM (n=28) per reviewed article (n=71).

Study[a] / PROM[b] [References Table 2[i]]	Country	Patient (n)	Reliability			Validity				Responsiveness[e]	
			Test-retest	Internal reliability	Measurement error	Content	Convergent/ divergent	Known groups	Structural	Responsiveness - COSMIN	Responsiveness (other)
Generic – health status (12/28)											
Profile measures (6/12)											
COOP/WONCA											
Van Balen (2003)[1]	Holland	208	-	-	-	-	Good	Poor	-	-	ES
Nottingham Health Profile (NHP)											
Borgquist (1992)[2]	Sweden	100	-	-	-	-	-	Poor	-	-	-
Van Balen (2003)[1]	Holland	208	-	Poor	-	-	Good	Fair	-	-	ES
Tidermark (2007)[3]	Sweden	59	-	-	-	-	Fair	-	-	Good	SRM
Quality of Life Scale (QOLS)											
Rohde (2010)[4]	Norway	61	-	Fair	-	-	-	Poor	-	-	ES
SF-36 (v1)											
Hall (2000)[5]	Australia	184	-	-	-	-	Fair	Fair	-	-	-
Randell (2000)[6]	Australia	32	-	-	-	-	Fair	-	-	-	SRM
Tosteson (2001)[7]	Sweden	67	-	-	-	-	-	Fair	-	-	-
Tidermark (2003)[8]	Sweden	95	-	-	-	-	-	Good	-	Fair	ES; SRM; Group discrimination[h]
Binder (2004)[9]	USA	90	-	-	-	-	-	Fair	-	-	-
Boonen (2004)[10]	Belgium	134	-	-	-	-	-	Fair	-	-	-
Duppils & Wikblad (2004)[11]	Sweden	115	Poor[c]		Poor		Poor	Poor		-	-
Hallberg (2004)[12]	Sweden	40	-	-	-	-	-	Fair	-	-	-
Shyu (2004a)[13]	Taiwan	116	Poor[c]	Good	-	-	Poor	Fair	-	-	-
Cranney (2005)[14]	Canada	40	-	-	-	-	Poor	Fair	-	-	-
Mattsson (2005)[15]	Sweden	112	-	-	-	-	-	Poor	-	-	-
MaCaulay (2008)[16]	USA	40	-	-	-	-	-	Poor	-	-	-
Zlowodski (2008) [17]	USA	70	-	-	-	-	-	Good	-	-	-
Hallberg (2009)[18]	Sweden	25	-	-	-	-	-	Fair	-	-	-
Rohde (2010)[4]	Norway	61	-	Fair	-	-	-	Fair	-	-	ES
Ziden (2010)[19]	Sweden	102	-	-	-	-	-	Poor	-	-	-
Shyu (2013)[20]	Taiwan	299	-	-	-	-	-	Fair	-	-	-
SF-36 (v1) PF											
Latham (2008)[21]	Multiple	108	-	-	-	-	Fair	Fair	-	Fair	ES, SRM, MDC_{90}
SF-36 (v1) PF/RP/BP											
Jongjit (2003)[22]	Thailand	60	-	-	-	-	Fair	Fair	-	-	-

SF-12 (v1)											
Vergara (2014)[23]	Spain	638	-	-	-	-	-	Fair	-	-	-
Orive (2015)[24]	Spain	891	-	-	-	-	-	Fair	-	-	-
SF-12 (v1) PCS											
Mishra (2004)[25]	UK	51	-	-	-	-	Fair	-	-	-	-
WHOQoL-BREF											
Tsauo (2005)[26]	Taiwan	25	-	-	-	-	-	Poor	-	-	-
Single item VAS (1/12)											
EuroQoL EQ-5D VAS											
Frihagen (2007)[27]	Sweden	137	-	-	-	-	-	Poor	-	-	-
Frihagen (2008)[28]	Sweden	79	-	-	-	-	-	Poor	-	-	-
Beucking (2014)[29]	Germany	227	-	-	-	-	Poor	-	-	-	-
Graham (2014)[30]	USA	194	-	-	-	-	-	Poor	-	-	-
Parsons (2014)[31]	England	225	-	-	-	-	-	-	-	-	ES
Preference-based Utility Measures (4/12)											
EuroQoL EQ-5D 3L											
Tidermark (2002a)[32]	Sweden	90	-	-	-	-	-	Fair	-	-	-
Tidermark (2002b)[33]	Sweden	90	-	-	-	-	-	Fair	-	-	-
Tidermark (2003)[8]	Sweden	95	-	-	-	-	-	Fair	-	Fair	ES; SRM; Group discrimination [h]
Blomfeldt (2006) [34]	Sweden	84					-	Fair		-	-
Soderqvist (2006)[35]	Sweden	213	-	-	-	-	-	Fair	-	-	-
Tidermark (2007)[3]	Sweden	59	-	-	-	-	Fair	-	-	Good	SRM
Frihagen (2007)[27]	Sweden	137	-	-	-	-	-	Poor	-	-	-
Frihagen (2008)[28]	Sweden	79	-	-	-	-	-	Poor	-	-	-
Zlowodski (2008) [17]	USA	70	-	-	-	-	-	Fair	-	-	-
Gjertsen (2011)[36]	Norway	1948	-	-	-	-	-	Poor	-	-	-
Hajbaghery (2013)[37]	Iran	140	-	-	-	-	-	Poor	-	-	-
Inngul (2013)[38]	Sweden	59	-	-	-	-	-	Fair	-	-	ES
Graham (2014)[30]	USA	194	-	-	-	-	-	Poor	-	-	-
Beucking (2014)[29]	Germany	227	-	-	-	-	Poor	-	-	-	ES
Parsons (2014)[31]	England	225	-	-	-	-	Fair	-	-	-	ES
Griffin (2015)[39]	England	403	-	-	-	-	-	Fair	-	-	-
HUI-2											
Cranney (2005)[14]	Canada	40	-	n/a	-	-	Poor	Fair	n/a	-	SRM
Jones (2005)[40]	Canada	245	Good[d]	n/a	-	-	-	-	-	-	-
Jones (2014)[41]	Canada	278	-	n/a	-	-	-	-	-	-	ES
HUI-3											
Jones (2005)[40]	Canada	245	Good[d]	n/a	-	-	-	-	-	-	-
Jones (2014)[41]	Canada	278	-	n/a	-	-	-	-	-	-	ES

Table 2 Meth Qual per Study and per PROM TC R! 130616

SF-6D											
Cranney (2005)[14]	Canada	40	-	n/a	-	-	Poor	Fair	-	-	SRM
Capability measure (1/12)											
ICECAP-O											
Parsons (2014)[31]	England	225	-	-	-	-	Poor	-	-	-	ES
			Test-retest	**Internal reliability**	**Measurement error**	**Content**	**Convergent/ divergent**	**Known groups**	**Structural**	**Responsiveness - COSMIN**	**Responsiveness (other)**
Specific measures (16/28)											
Hip-specific (3/16): Surgeon-based assessment (2/3)											
Charnley Hip Score											
Blomfeldt (2006) [34]	Sweden	84	-	-	-	-	-	Fair	-	-	-
Soderqvist (2006)[35]	Sweden	213	-	-	-	-	-	Fair	-	-	-
Tidermark (2003)[8]	Sweden	95	-	-	-	-	-	Poor	-	-	-
Harris Hip Score											
Tsauo (2005)[26]	Taiwan	25	-	-	-	-	-	Poor	-	-	-
Frihagen (2007)[27]	Sweden	137	-	-	-	-	-	Poor	-	-	-
Frihagen (2008)[28]	Sweden	79	-	-	-	-	-	Poor	-	-	-
MaCauley (2008)[16]	USA	40						Poor			
Mouzopoulos (2008)[42]	Greece	62	-	-	-	-	Fair	Fair	-	-	-
Inngul (2013)[38]	Sweden	59	-	-	-	-	-	Fair	-	-	-
Graham (2014)[30]	USA	194	-	-	-	-	-	Poor	-	-	-
Patient-completed assessment (1/3)											
Oxford Hip Score											
Mishra (2004)[25]	UK	51	-	-	-	-	Poor	-	-	-	-
Parsons (2014)[31]	England	225	-	-	-	-	Fair	-	-	-	ES
Griffin (2015)[39]	England	403	-	-	-	-	-	Fair	-	-	-
Disease-specific (2/16)											
OPAQ-2											
Randell (2000)[6]	Australia	32	-	-	-	-	Fair	-	-	-	SRM
WOMAC											
MaCauley (2008)[16]	USA	40	-	-	-	-	-	Poor	-	-	-
Vergara (2014)[23]	Spain	638	-	-	-	-	-	Fair	-	-	-
Zielinski (2014)[43]	Holland	248	-	-	-	-	-	Fair	-	-	-
Orive (2015)[24]	Spain	891	-	-	-	-	-	Fair	-	-	-
Domain-specific (11/16)											
Emotional well-being (3/11)											
Geriatric Depression Scale (GDS)											
Shepherd (1996)[44]	England	270	-	-	-	-	-	-	-	-	ES; SRM

Table 2 Meth Qual per Study and per PROM TC R! 130616

Bellelli (2008)[45]	Italy	211	-	-	-	-	-	Fair	-	-	-
Shyu (2008)[46]	Taiwan	162	-	-	-	-	-	Fair	-	-	-
Shyu (2009)[47]	Taiwan	147	-	Poor	-	-	-	-	-	-	-
Hamilton Rating Scale for Depression (HAM-D)											
Lenze (2007)[48]	USA	126	-	Fair	-	-	-	-	-	-	-
Zung Depression Inventory (Zung DI)											
Arinzon (2007)[49]	Israel	63	-	-	-	-	Poor	-	-	-	-
Activities of Daily Living / Instrumental Activities of Daily Living (ADL/IADL)(8/11)											
Barthel Index (BI)											
Van Balen (2003)[1]	Holland	208	-	Poor	-	-	Good	-	-	-	ES
Shyu (2004a)[13]	Taiwan	116	-	Poor	-	-	Poor	Fair	-	-	-
Shyu (2004b)[50]	Taiwan	110	-	Poor	-	-	-	-	-	-	-
Frihagen (2007)[27]	Sweden	137	-	-	-	-	-	Poor	-	-	-
Bellelli (2008)[45]	Italy	211	-	-	-	-	-	Fair	-	-	-
Mouzopoulos (2008)[42]	Greece	62	-	-	-	-	Fair	Fair	-	-	-
Shyu (2008)[46]	Taiwan	162	-	-	-	-	-	Fair	-	-	-
Shyu (2009)[47]	Taiwan	147	-	Poor	-	-	-	-	-	-	-
Vergara (2014)[23]	Spain	638	-	-	-	-	-	Fair	-	-	-
Orive (2015)[24]	Spain	891	-	-	-	-	-	Fair	-	-	-
Modified-Barthel Index (M-BI)											
Hall (2000)[5]	Australia	184	-	-	-	-	Fair	Fair	-	-	-
Beaupre (2005)[51]	Canada	919	-	-	-	-	-	Poor	-	-	-
Functional Activities Index (FAI)											
Hall (2000)[5]	Australia	184	-	-	-	-	Fair	Fair	-	-	-
Jongjit (2003)[22]	Thailand	60	-	-	-	-	Fair	Fair	-	-	-
Zidan (2010)[19]	Sweden	102	-	-	-	-	-	Fair	-	-	-
Functional Independence Measure (FIM)											
FIM – total											
Goldstein (1997)[52]	USA	58	-	-	-	-	Poor	Poor	-	-	-
Adunsky (2001)[53]	Israel	217	-	-	-	-	-	Poor	-	-	-
Adunsky (2002)[54]	Israel	143	-	-	-	-	Poor	-	-	-	-
Beloosesky (2002)[55]	Israel	153	-	-	-	-	-	Fair	-	-	-
Adunsky (2001)[53]	Israel	217	-	-	-	-	-	Poor	-	-	-
Jones (2002)[56]	Canada	100	-	-	-	-	-	Poor	-	-	-
Jongjit (2003)[22]	Thailand	60	-	-	-	-	Fair	Fair	-	-	-
Mendelsohn (2003)[57]	Canada	40	-	-	-	-	Fair	-	-	Poor	-
Beloosesky (2004)[58]	Israel	123	-	-	-	-	-	Fair	-	-	-
Cornwall (2004)[59]	USA	537	-	-	-	-	-	Good	-	-	-
Rolland (2004)[60]	France	61	-	-	-	-	-	Poor	-	-	-
Jones (2006)[61]	Canada	137	Good[d]	-	-	-	-	-	-	-	-

Table 2 Meth Qual per Study and per PROM TC R! 130616

Nguyen-Oghalai (2006)[62]	USA	13394	Poor[c]	-	-	-	-	Fair	-	-	-
Arinzon (2007)[49]	Israel	63	-	-	-	-	Poor	-	-	-	-
Hershkovitz (2007)[63]	Israel	133	-	-	-	-	Poor	Fair	-	-	-
Mizrahi (2007)[64]	Israel	460	-	-	-	-	-	Fair	-	-	-
Graham (2008)[65]	USA	6970	-	-	-	-	-	Poor	-	-	-
Zidan (2010)[19]	Sweden	102	-	-	-	-	-	Fair	-	-	-
FIM – cognition											
Heruti (1999)[66]	Israel	204	-	-	-	-	Fair	Fair	-	-	-
Deutsch (2005)[67]	USA	29,793	-	Poor	-	-	-	-	-	-	-
McGilton (2009)[68]	Canada	31	-	-	-	-	-	Poor	-	-	-
FIM - mobility											
Mendelsohn (2003)[57]	Canada	40	-	-	-	-	Fair	-	-	Poor	-
FIM - motor											
Heruti (1999)[66]	Israel	204					-	Poor	-	-	-
Dorra (2002)[69]	USA	137	-	-	-	-	-	Fair	-	-	-
Mendelsohn (2003)[57]	Canada	40	-	-	-	-	Fair	-	-	Poor	-
Deutsch (2005)[67]	USA	29,793	-	Poor	-	-	-	-	-	-	-
Hershkovitz (2007)[63]	Israel	133	-	-	-	-	Poor	Fair	-	-	-
Mizrahi (2007)[64]	Israel	460	-	-	-	-	-	Fair	-	-	-
McGilton (2009)[68]	Canada	31	-	-	-	-	-	Poor	-	-	-
Functional Status Questionnaire (FSQ)											
Binder (2004)[9]	USA	90	-	-	-	-	-	Poor	-	-	-
Lawton Instrumental Activities of Daily Life Scale (Lawton-IADL)											
Shyu (2004a)[13]	Taiwan	116	-	-	-	-	Fair	-	-	-	-
Bellelli (2008)[45]	Italy	211	-	-	-	-	-	Poor	-	-	-
Vergara (2014)[23]	Spain	638	-	-	-	-	-	Fair	-	-	-
Orive (2015)[24]	Spain	891	-	-	-	-	-	Fair	-	-	-
Katz Index of Independence in Activities of Daily Life (Katz ADL)											
Beloosky (2002)[55]	Israel	153	-	-	-	-	-	Poor	-	-	-
Kirke (2002)[70]	Ireland	106	-	-	-	-	-	Poor	-	-	-
Tidermark (2002a)[32]	Sweden	90	-	-	-	-	-	Fair	-	-	-
Tidermark (2002b)[33]	Sweden	90	-	-	-	-	-	Fair	-	-	-
Tidermark (2003)[8]	Sweden	95	-	-	-	-	-	Poor	-	-	-
Vidan (2005)[71]	Spain	250	-	-	-	-	-	Poor	-	-	-
Soderqvist (2006)[35]	Sweden	213	-	-	-	-	-	Fair	-	-	-
Arinzon (2007)[49]	Israel	63	-	-	-	-	Poor	-	-	-	-
OARS Multi-dimensional Functional Assessment Questionnaire (OMFAQ)											
Binder (2004)[9]	USA	90	-	-	-	-	-	Poor	-	-	-

Table 2 Meth Qual per Study and per PROM TC R! 130616

Footnote: [a] Reviewed article reports measurement / practical properties for listed PROM.

[b] PROMs (acronyms; alphabetical order): BDI-II - Beck Depression Inventory – version II; BI – Barthel Index; Charnley HS – Charnley Hip Score; COOP/W – COOP/WONCA Charts; EQ-5D - EuroQoL EQ-5D; FAI – Functional Activities Index; FIM – Functional Independence Measure; FSQ - Functional Status Questionnaire; GDS - Geriatric Depression Scale; HAM-D - Hamilton Rating Scale for Depression (HAM-D); HHS – Harris Hip Score (self-report domains included if reported separately); HUI 2/3 – Health Utility Index; ICECAP-O - ICEpop CAPability measure for Older people; Katz ADL - Katz Index of Independence in Activities of Daily Life; Lawton IADL - Lawton-Instrumental Activities of Daily Life Scale; MBI – Modified Barthel Index; NHP – Nottingham Health Profile; OHS - Oxford Hip Score; OMFAQ – Older Americans Resources and Services (OARS) Multi-dimensional Functional Assessment Questionnaire; OPAQ2 – Osteoporosis Quality of Life Questionnaire (version 2); QLS - Quality of Life Scale; SF-36/12 – Short-Form 36/12-item Health Survey; WHOQOL – World Health Organisation Quality of Life Questionnaire; WOMAC - Western Ontario and McMaster Universities Osteoarthritis Index; ZDI – Zung Depression Inventory.

[c] Inter-rater reliability (patients and raters – detail re raters not clear).

[d] Inter-rater agreement between patients and proxy

[e] Responsiveness: Studies were awarded a COSMIN rating where evidence of responsiveness / longitudinal validity conformed with the standards of responsiveness as defined by the COSMIN initiative. For example, correlation of score change with change in a criterion measure; correlation with changes in other similar measures; Receiver Operating Curve (ROC) / Area Under the Curve (AUC)[72]. COSMIN ratings do not include distribution-based assessments - for example, Effect size (ES), Standardised Response Mean (SRM): reporting of such evidence was not awarded a COSMIN score and is reported in a separate column.

[f] Minimal Detectable Change (MDC_{90})

[g] Paper included for (limited) evidence of feasibility and acceptability.

[h] Group discrimination - between groups (external criterion: good vs less good outcome).

[i] References for Table 2

1. Van Balen, R., Essink-Bot, M.L., Steyerberg, E., Cools, H., Habbema, D.F (2003). Quality of life after hip fracture: a comparison of four health status measures in 208 patients. Disability and Rehabilitation, 25(10), 507-519.
2. Borgquist, L., Nilsson, L.T.,Lindelöw, G.,Wiklund, I.,Thorngren, K.G. (1992). Perceived health in hip-fracture patients: a prospective follow-up of 100 patients. Age and Aging, 21(2), 109-116.
3. Tidermark, J., Bergström, G. (2007). Responsiveness of the EuroQol (EQ-5D) and the Nottingham Health Profile (NHP) in elderly patients with femoral neck fractures. Qual Life Res., 16(2), 321-330.
4. Rohde, G., Haugeberg, G., Mengshoel, A.M., Moum, T., Wahl, A.K. (2010). Two-year changes in quality of life in elderly patients with low-energy hip fractures. A case-control study. BMC Musculoskelet Disord, 11, 226.
5. Hall, S. E., Williams, J.A., Senior, J.A., Goldswain, P.R., Criddle, R.A. (2000). Hip fracture outcomes: quality of life and functional status in older adults living in the community. Aust N Z J Med, 30(3), 327-332.
6. Randell, A. G., Nguyen, T. V., Bhalerao, N., Silverman, S. L., Sambrook, P. N., & Eisman, J. A. (2000). Deterioration in quality of life following hip fracture: a prospective study. Osteoporosis International, 11(5), 460-466.
7. Tosteson, A. N., Gabriel, S.E.,Grove, M.R.,Moncur, M.M.,Kneeland, T.S., Melton, L.J. (2001). Impact of Hip and Vertebral Fractures on Quality-Adjusted Life Years. Osteoporos Int. , 12, 1042-1049.
8. Tidermark, J., Blomfeldt, R., Ponzer, S., Soderqvist, A., & Tornkvist, H. (2003). Primary total hip arthroplasty with a Burch-Schneider antiprotrusion cage and autologous bone grafting for acetabular fractures in elderly patients. Journal of Orthopaedic Trauma, 17(3), 193-197.
9. Binder, E. F., Brown, M., Sinacore, D.R. (2004). Effects of Extended Outpatient Rehabilitation After Hip Fracture A Randomized Controlled Trial. JAMA, 292(7), 837-846.
10. Boonen, S., Autier, P.,Barette, M.,Vanderschueren, D.,Lips, P.,Haentjens, P. (2004). Functional outcome and quality of life following hip fracture in elderly women: a prospective controlled study. Osteoporos Int., 15, 87-94.
11. Duppils, G. S., Wikblad, K. (2004). Cognitive Function and Health-Related Quality of Life After Delirium in Connection With Hip Surgery: A Six-Month Follow-Up. Orthopaedic Nursing, 23(3), 195-203.
12. Hallberg, I., Rosenqvist, A. M., Kartous, L., Lofman, O., Wahlstrom, O., & Toss, G. (2004). Health-related quality of life after osteoporotic fractures. Osteoporosis International, 15(10), 834-841.
13. Shyu, Y., Lu, J.F., Liang, J. (2004a). Evaluation of Medical Outcomes Study Short Form-36 Taiwan version in assessing elderly patients with hip fracture. Osteoporos Int., 15(7), 575-582.

14. Cranney, A. B., Coyle, D.,Hopman, W.M.,Hum, V.,Power, B.,Tugwell, P.S. (2005). Prospective Evaluation of Preferences and Quality of Life in Women with Hip Fractures. The Journal of Rheumatology, 32(12), 2393-2399.
15. Mattsson, P., Alberts, A., Dahlberg, G., Sohlman, M., Hyldahl, H.C., Larsson, S. (2005). Resorbable cement for the augmentation of internally-fixed unstable trochanteric fractures. A prospective, randomised multicentre study. J Bone Joint Surg Br, 87(9), 1203-1209.
16. Macaulay, W., Nellans, K.W., Garvin, K.L., Iorio, R., Healy, W.L., Rosenwasser, M.P.; other members of the DFACTO Consortium. (2008). Prospective randomized clinical trial comparing hemiarthroplasty to total hip arthroplasty in the treatment of displaced femoral neck fractures: winner of the Dorr Award. J Arthroplasty, 23(6), 2-8.
17. Zlowodzki, M., Brink. O., Switzer, J., Wingerter, S., Woodall, J., Petrisor, B.A., Kregor, P.J., Bruinsma, D.R., Bhandari, M. (2008). The effect of shortening and varus collapse of the femoral neck on function after fixation of intracapsular fracture of the hip: a multi-centre cohort study. J Bone Joint Surg Br., 90(11), 1487-1494.
18. Hallberg, I., Bachrach-Lindström, M., Hammerby, S., Toss, G., Ek, A.C. (2009). Health-related quality of life after vertebral or hip fracture: a seven-year follow-up study. BMC Musculoskelet Disord, 10, 135.
19. Zidén, L., Kreuter, M., Frändin, K. (2010). Long-term effects of home rehabilitation after hip fracture - 1-year follow-up of functioning, balance confidence, and health-related quality of life in elderly people. Disabil Rehabil, 32(1), 18-32.
20. Shyu, Y., Liang, J., Tseng, M.Y., Li, H.J., Wu, C.C., Cheng, H.S., Chou, S.W., Chen, C.Y., Yang, C.T. (2013). Comprehensive and subacute care interventions improve health-related quality of life for older patients after surgery for hip fracture: a randomised controlled trial. International Journal Of Nursing Studies, 50(8), 1013-1024.
21. Latham, N. K., Mehta, V., Nguyen, A.M., Jette, A.M., Olarsch, S., Papanicolaou, D., Chandler, J. (2008). Performance-based or self-report measures of physical function: which should be used in clinical trials of hip fracture patients? Arch Phys Med Rehabil, 89(11), 2146-2155.
22. Jongjit, J., Komsopapong, L., Songjakkaew, P., Kongsakon, R. (2003). Health-related quality of life after hip fracture in the elderly community-dwelling. Southeast Asian J Trop Med Public Health, 34(3), 670-674.
23. Vergara, I., Vrotsou, K., Orive, M., Gonzalez, N., Garcia, S., Quintana, J.M. (2014). Factors related to functional prognosis in elderly patients after accidental hip fractures: a prospective cohort study. BMC Geriatrics, 14, 124-124.
24. Orive, M., Aguirre, U., García-Gutiérrez, S., Las Hayas, C., Bilbao, A., González, N., Zabala, J., Navarro, G., Quintana, J.M. (2015). Changes in health-related quality of life and activities of daily living after hip fracture because of a fall in elderly patients: a prospective cohort study. Int J Clin Pract. 2015 Apr;69(4):491-500. doi: 10.1111/ijcp.12527. Epub 2015 Feb 27.
25. Mishra, V., Thomas, G., Sibly, T.F. (2004). Results of displaced subcapital fractures treated by primary total hip replacement. Injury, 35(2), 157-160.
26. Tsauo, J. Y., Leu, W.S., Chen, Y.T., Yang, R.S. (2005). Effects on Function and Quality of Life of Postoperative Home-Based Physical Therapy for Patients With Hip Fracture. Arch Phys Med Rehabil, 86, 1953-1957.
27. Frihagen, F., Nordsletten, L., Madsen, J.E. (2007). Hemiarthroplasty or internal fixation for intracapsular displaced femoral neck fractures: randomised controlled trial. BMJ, 335(7632), 1251-1254.
28. Frihagen, F., Grotle, M., Madsen, J.E., Wyller, T.B., Mowinckel, P., Nordsletten, L. (2008). Outcome after femoral neck fractures: a comparison of Harris Hip Score, Eq-5d and Barthel Index. Injury, 39(10), 1147-1156.
29. Buecking, B., Waldermann, A., Horstmann, K., Schubert, N., Balzer-Geldsetzer, M., Dodel, R., Bohl, K., Ruchholtz, S., Bliemel, C. . (2014). What determines health-related quality of life in hip fracture patients at the end of acute care?--a prospective observational study. Osteoporos Int., 25(2), 475-484.
30. Graham, J., Bowen, T. R., Strohecker, K. A., Irgit, K., & Smith, W. R. (2014). Reducing mortality in hip fracture patients using a perioperative approach and "Patient- Centered Medical Home" model: a prospective cohort study. Patient Safety In Surgery, 8(1), 7-7.
31. Parsons, N., Griffin, X. L., Achten, J., Costa, M. L. (2014). Outcome assessment after hip fracture: is EQ-5D the answer? Bone & Joint Research, 3(3), 69-75.
32. Tidermark, J., Zethraeus, N., Svensson, O., Tornkvist, H., Ponzer, S. . (2002a). Femoral neck fractures in the eldery: functional outcome and quality of life according to the EuroQol. Quality of Life Research, 11, 473-481.
33. Tidermark, J., Zethraeus, N., Svensson, O., Tornkvist, H., Ponzer, S. (2002b). Quality of life related to fracture displacement among elderly patients with femoral neck fractures treated with internal fixation. J Orthop Trauma, 16, 34-38.
34. Blomfeldt, R., Törnkvist, H., Ponzer, S., Söderqvist, A., Tidermark, J. (2006). Displaced femoral neck fracture: comparison of primary total hip replacement with secondary replacement after failed internal fixation: a 2-year follow-up of 84 patients. Acta Orthopaedica, 77(4), 638-643.
35. Söderqvist, A., Miedel, R., Ponzer, S., Tidermark, J. (2006). The influence of cognitive function on outcome after a hip fracture. J Bone Joint Surg Am, 88(10), 2115-2123.
36. Gjertsen, J. E., Fevang, J.M., Matre, K., Vinje, T., Engesæter, L.B. (2011). Clinical outcome after undisplaced femoral neck fractures. Acta Orthopaedica, 82(3), 268-274.
37. Hajbaghery, M. A., Abbasinia, M. (2013). Quality of Life of the Elderly after Hip Fracture Surgery: A Case-Control Study. Journal of Caring Sciences, 2(1), 53-59.

38. Inngul, C., Hedbeck, C.-J., Blomfeldt, R., Lapidus, G., Ponzer, S., & Enocson, A. (2013). Unipolar hemiarthroplasty versus bipolar hemiarthroplasty in patients with displaced femoral neck fractures: a four-year follow-up of a randomised controlled trial. International Orthopaedics, 37(12), 2457-2464.
39. Griffin, X. L., Parsons, N., Achten, J., Fernandez, M., Costa, M.L. (2015). Recovery of health-related quality of life in a United Kingdom hip fracture population: the Warwick Hip Trauma Evaluation - a prospective cohort study. Bone Joint J. 2015 Mar;97-B(3):372-82. doi: 10.1302/0301-620X.97B3.35738.
40. Jones, C. A., Feeney, D.H. (2005). Agreement Between Patient and Proxy Responses of Health-Related Quality of Life After Hip Fracture. Journal of American Geriatrics Society, 53, 1227-1233.
41. Jones, C. A., Pohar, S. L., Feeny, D. H., & Eng, K. (2014). Longitudinal construct validity of the Health Utilities Indices Mark 2 and Mark 3 in hip fracture. Quality Of Life Research: An International Journal Of Quality Of Life Aspects Of Treatment, Care And Rehabilitation, 23(3), 805-813.
42. Mouzopoulos, G., Stamatakos, M., Arabatzi, H., Vasiliadis, G., Batanis, G., Tsembeli, A., Tzurbakis, M., Safioleas, M. (2008). The four-year functional result after a displaced subcapital hip fracture treated with three different surgical options. Int Orthop. , 32(3), 367-373.
43. Zielinski, S. M., Keijsers, N.L., Praet, S.F., Heetveld, M.J., Bhandari, M., Wilssens, J.P., Patka, P., Van Lieshout, E.M., FAITH Trial Investigators. (2014). Functional outcome after successful internal fixation versus salvage arthroplasty of patients with a femoral neck fracture. J Orthop Trauma. 2014 Dec;28(12):e273-80. .
44. Shepherd, S. M., Prescott, R.J. (1996). Use of standardised assessment scales in elderly hip fracture patients. J R Coll Physicians Lond, 30(4), 335-343.
45. Bellelli, G., Frisoni, G.B., Turco, R., Trabucchi, M. (2008). Depressive symptoms combined with dementia affect 12-months survival in elderly patients after rehabilitation post-hip fracture surgery. Int J Geriatr Psychiatry, 23(10), 1073-1077.
46. Shyu, Y., Liang, J., Wu, C.C., Su, J.Y., Cheng, H.S., Chou, S.W., Chen, M.C., Yang, C.T. (2008). Interdisciplinary intervention for hip fracture in older Taiwanese: benefits last for 1 year. J Gerontol 63(1), 92-97.
47. Shyu, Y., Cheng, H.S., Teng, H.C., Chen, M.C., Wu, C.C., Tsai, W.C. (2009). Older people with hip fracture: depression in the postoperative first year. J Adv Nurs, 65(12), 2514-2522.
48. Lenze, E. J., Munin, M.C, Skidmore, E.R., Dew, M.A., Rogers J.C., Whyte, E.M., Quear, T., Begley, A., Reynolds, C.F. . (2007). Onset of depression in elderly persons after hip fracture: implications for prevention and early intervention of late-life depression. J Am Geriatr Soc, 55(1), 81-86.
49. Arinzon, Z., Gepstein, R., Shabat, S., Berner, Y. (2007). Pain perception during the rehabilitation phase following traumatic hip fracture in the elderly is an important prognostic factor and treatment tool. Disabil Rehabil, 29(8), 651-658.
50. Shyu, Y. I., Chen, M.C., Liang, J.,Wu, C.C.,Su, J.Y. (2004b). Predictors of functional recovery for hip fractured elders during12 months following hospital discharge: a prospective study on a Taiwanese sample. Osteoporos Int., 15, 475-482.
51. Beaupre, L. A., Cinats, J.G., Senthilselvan, A., Scharfenberger, A., Johnston, D.W., Saunders, L.D. (2005). Does standardized rehabilitation and discharge planning improve functional recovery in elderly patients with hip fracture? Arch Phys Med Rehabil, 86(12), 2231-2239.
52. Goldstein, F. C., Strasser, D.C., Woodard, J.L., Roberts, V.J. (1997). Functional outcome of cognitively impaired hip fracture patients on a geriatric rehabilitation unit. J Am Geriatr Soc, 45(1), 35-42.
53. Adunsky, A., Levenkrohn, S., Fleissig, Y., Arad, M., Heruti, R.J. (2001). Rehabilitation outcomes in patients with full weight-bearing hip fractures. Arch Gerontol Geriatr. , 33(2), 123-131.
54. Adunsky, A., Lusky, A., Arad, M., Heruti, R.J. (2002). A Comparative Study of Rehabilitation Outcomes of Elderly Hip Fracture Patients: The Advantage of a Comprehensive Orthogeriatric Approach Journal of Gerontology, 58(6), 542-547.
55. Beloosesky, Y., Grinblat, J., Epelboym, B., Weiss, A., Grosman, B., Hendel, D. (2002). Functional gain of hip fracture patients in different cognitive and functional groups. Clin Rehabil, 16(3), 321-328.
56. Jones, G. R., Miller, T.A., Petrella, R.J. (2002). Evaluation of rehabilitation outcomes in older patients with hip fractures. Am J Phys Med Rehabil, 81(7), 489-497.
57. Mendelsohn, M. E., Leidl, D.S., Overend, T.J., Petrella, R.J. (2003). Specificity of functional mobility measures in older adults after hip fracture: a pilot study. Am J Phys Med Rehabil, 82(10), 766-774.
58. Beloosesky, Y., Weiss, A., Grinblat, J., Brill, S., Hershkovitz, A. (2004). Can functional status, after rehabilitation, independently predict long-term mortality of hip-fractured elderly patients? Aging Clin Exp Res, 16(1), 44-48.
59. Cornwall, R., Gilbert, M.S., Koval, K.J., Strauss, E., Siu, A.L. (2004). Functional outcomes and mortality vary among different types of hip fractures: a function of patient characteristics. Clin Orthop Relat Res. , 425, 64-71.
60. Rolland, Y., Pillard, F., Lauwers-Cances, V., Busquère, F., Vellas, B., Lafont, C. (2004). Rehabilitation outcome of elderly patients with hip fracture and cognitive impairment. Disabil Rehabil, 26(7), 425-431.
61. Jones, C. A., Feeny, D.H. (2006). Agreement between patient and proxy responses during recovery after hip fracture: evidence for the FIM instrument. Arch Phys Med Rehabil, 87(10), 1382-1387.

Table 2 Meth Qual per Study and per PROM TC R! 130616

62. Nguyen-Oghalai, T. U., Ottenbacher, K.J., Granger, C.V., Smith, S.T., Goodwin, J.S. (2006). Impact of osteoarthritis on rehabilitation for persons with hip fracture. Arthritis Rheum, 55(6), 920-924.
63. Hershkovitz, A., , Kalandariov, Z., Hermush, V., Weiss, R., Brill, S. (2007). Factors affecting short-term rehabilitation outcomes of disabled elderly patients with proximal hip fracture. Arch Phys Med Rehabil, 88(7), 916-921.
64. Mizrahi, E. H., Fleissig, Y., Arad, M., Adunsky, A. (2007). The impact of previous strokes on the rehabilitation of elderly patients sustaining a hip fracture. Arch Phys Med Rehabil, 88(9), 1136-1139.
65. Graham, J. E., Chang, P.F., Bergés, I.M., Granger, C.V., Ottenbacher, K.J. (2008). Race / Ethnicity and Outcomes Following Inpatient Rehabilitation for Hip Fracture. J Gerontol 63(8), 860-866.
66. Heruti, R. J., Lusky, A., Barell, V., Ohry, A., Adunsky, A. (1999). Cognitive status at admission: does it affect the rehabilitation outcome of elderly patients with hip fracture? Arch Phys Med Rehabil, 80(4), 432-436.
67. Deutsch, A., Granger, C.V., Fiedler, R.C., DeJong, G., Kane, R.L., Ottenbacher, K.J., Heinemann, A.W., Naughton, J.P., Trevisan, M. (2005). Outcomes and reimbursement of inpatient rehabilitation facilities and subacute rehabilitation programs for Medicare beneficiaries with hip fracture. Med Care, 43(9), 892-901.
68. McGilton, K. S., Mahomed, N., Davis, A.M., Flannery, J., Calabrese, S. (2009). Outcomes for older adults in an inpatient rehabilitation facility following hip fracture (HF) surgery. Arch Gerontol Geriatr, 49(1), 23-31.
69. Dorra, H. H., Lenze, E.J., Kim, Y., Mulsant, B.H., Munin M.C., Dew, M.A., Reynolds, C.F. (2002). Clinically relevant behaviors in elderly hip fracture inpatients. Int J Psychiatry Med, 32(3), 249-259.
70. Kirke, P. N., Sutton, M., Burke, H., Daly, L. (2002). Outcome of hip fracture in older Irish women: a 2-year follow-up of subjects in a case-control study. Injury, 33(5), 387-391.
71. Vidán, M., Serra, J.A., Moreno, C., Riquelme, G., Ortiz, J. (2005). Efficacy of a comprehensive geriatric intervention in older patients hospitalized for hip fracture: a randomized, controlled trial. J Am Geriatr Soc, 53(9), 1476-1482.
72. de Vet, H., Terwee, C.B., Mokkink, L.B., Knol, D.L. (2011). Measurement in Medicine. A Practical Guide. Cambridge University Press, Sept, http://dx.doi.org/10.1017/CBO9780511996214.

Table 3: Measurement properties and methodological quality: data synthesis [a], levels of evidence and overall quality of measurement and practical properties per reviewed PROM (n=28)

PROM (n)[b]	Eval (n)	Reliability			Validity				Responsiveness
		Test-retest	Internal consistency	Measurement error	Content	Convergent/ divergent	Known groups	Structural	Responsiveness [d]
Generic – health status (12/28)									
Profile measures (6/12)									
COOP/WONCA	1	-	-	-	-	+ Moderate	? Unknown	-	? Unknown
NHP	4	-	+ Limited	-	-	+ Moderate	+ Limited	-	+ Moderate
Quality of Life Scale	1	-	+ Limited	-	-	-	+ Unknown	-	? Unknown
SF-36 (v1)	17	? Unknown [c]	+ Moderate	-	-	+ Moderate	+ Moderate	-	+ Moderate
SF-12	3	-	-	-	-	+ Limited	-	-	-
WHOQoL-BREF	1	-	-	-	-	-	? Unknown	-	-
Single item VAS (1/12)									
EuroQoL VAS	5	-	n/a	-	-	? Unknown	? Unknown	n/a	? Unknown
Preference-based Utility Measures (4/12)									
EuroQoL EQ-5D	16	-	n/a	-	-	+ Limited	+ Moderate	n/a	+ Moderate
HUI 2	3	+ Moderate [c]	n/a	-	-	? Unknown	+ Limited	n/a	? Unknown
HUI 3	2	+ Moderate [c]	n/a	-	-	-	-	n/a	? Unknown
SF-6D	1	-	n/a	-	-	? Unknown	+ Limited	n/a	-
Capability measure (1/12)									
ICECAP-O	1	-	n/a	-	-	? Unknown	-	n/a	? Unknown
Specific measures (16/28)									
Hip-specific (3/16): Surgeon-based assessment (2/3)									

Table 3 Synthesis of reviewed evidence (TC R1 130616)

Charnley Hip Score*	4	-	-	-	-	-	+ Moderate	-	-
Harris Hip Score*	8	-	-	-	-	+ Limited	+ Moderate	-	-
Patient-completed assessment (1/3)									
Oxford Hip Score	3	-	-	-	-	+ Limited	+ Limited	-	? Unknown
Disease-specific (2/16)									
OPAQ-2	1	-	-	-	-	+ Limited	-	-	? Unknown
WOMAC	4	-	-	-	-	-	+ Moderate	-	-
Domain-specific (11/16)									
Emotional well-being (3/11)									
GDS	4	-	? Unknown	-	-	-	+ Moderate	-	? Unknown
HAM-D	1	-	+ Limited	-	-	-	-	-	-
Zung DI	1	-	-	-	-	? Unknown	-	-	-
Activities of Daily Living / Instrumental Activities of Daily Living (ADL/IADL)(8/11)									
Barthel Index	10	-	? Unknown	-	-	+ Moderate	+ Moderate	-	? Unknown
Modified BI	2	-	-	-	-	+ Limited	+ Limited	-	-
Functional Activities Index (FAI)	3	-	-	-	-	+ Moderate	+ Moderate	-	-
FIM	21	+ Moderate [c]	? Unknown	-	-	+ Moderate	+ Moderate	-	? Unknown
Functional Status Questionnaire (FSQ)	1	-	-	-	-	-	? Unknown	-	-
Lawton's IADL	5	-	- Limited	-	-	- Limited	+ Limited	-	-
Katz ADL	8	-	-	-	-	? Unknown	+ Moderate	-	-
OMFAQ	1	-	-	-	-	-	? Unknown	-	-

Table 3 Synthesis of reviewed evidence (TC R1 130616)

	n	Test-retest	Internal consistency	Measurement error	Content	Convergent / divergent	Known groups	Structural	Resp

Footnote:

[a] Data synthesis: The data were qualitatively synthesized to determine the overall quality of measurement properties and acceptability of each reviewed PROM. The synthesis took the following factors into account: 1) methodological quality of the reviewed studies (COSMIN scores); 2) the number of studies reporting evidence of measurement properties per PROM; 3) the results for each measurement property for each PROM; and 4) the consistency of results between reviewed studies.
The data synthesis score has two elements.

1) First, the overall quality of a measurement property was reported as: adequate (+), not adequate (-), conflicting (+/-), or unclear (?).
2) Second, levels of evidence for the overall quality of each measurement property were further defined to indicate 'strong' – consistent findings in multiple studies of good methodological quality OR in one study of excellent quality; 'moderate' – consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality; 'limited' – one study of fair methodological quality; 'conflicting' – conflicting findings; or 'unknown' evidence – only studies of poor methodological quality (detailed by Elbers et al, 2012 [12]; Conijn et al, 2015 [23]).

[b] PROMs (acronyms; alphabetical order): BDI-II - Beck Depression Inventory – version II; BI – Barthel Index; Charnley HS – Charnley Hip Score; COOP/W – COOP/WONCA Charts; EQ-5D - EuroQoL EQ-5D; FAI – Functional Activities Index; FIM – Functional Independence Measure; FSQ - Functional Status Questionnaire; GDS - Geriatric Depression Scale; HAM-D - Hamilton Rating Scale for Depression (HAM-D); HHS – Harris Hip Score (self-report domains included if reported separately); HUI 2/3 – Health Utility Index; ICECAP-O - ICEpop CAPability measure for Older people; Katz ADL - Katz Index of Independence in Activities of Daily Life; Lawton IADL - Lawton-Instrumental Activities of Daily Life Scale; MBI – Modified Barthel Index; NHP – Nottingham Health Profile; OHS - Oxford Hip Score; OMFAQ – Older Americans Resources and Services (OARS) Multi-dimensional Functional Assessment Questionnaire; OPAQ2 – Osteoporosis Quality of Life Questionnaire (version 2); QLS - Quality of Life Scale; SF-36/12 – Short-Form 36/12-item Health Survey; WHOQOL – World Health Organisation Quality of Life Questionnaire; WOMAC - Western Ontario and McMaster Universities Osteoarthritis Index; ZDI – Zung Depression Inventory.

[c] Inter-rater reliability

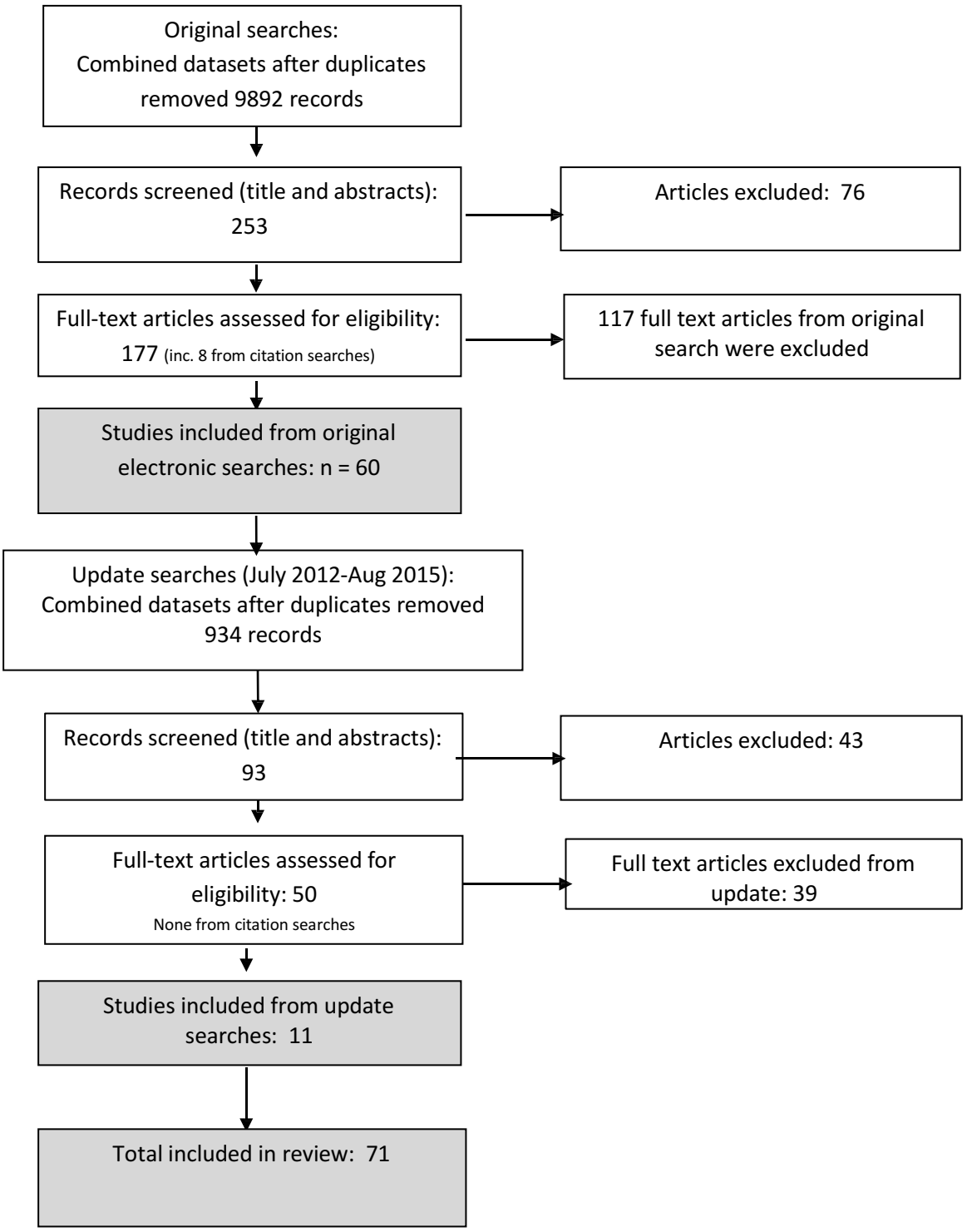
[d] Where only ES statistics reported – classified as '? Unknown' evidence of responsiveness

n/a Non-applicable.

Table 3 Synthesis of reviewed evidence (TC R1 130616)

PROHIP – Systematic review

Figure 1 Flow diagram for article selection



Appendix 1. Hip Review - Characteristics of included studies (n=71)

	Reference Country	Population, Setting, Completion	Study / Intervention	n	Mean age (years (SD); range)	Gender (% female)	PROMs	Language
1	Adunsky et al 2001 Israel	Post hip fracture rehabilitation programme Interview	Retrospective cohort: to evaluated the factors affecting recovery of function post-hip fracture Baseline: Admission to rehabilitation following hip fracture surgery Follow-up: Discharge from rehabilitation	217	79.85 (7.94) (median 81.00)	80%	FIM	Arabic/Hebrew
2	Adunsky et al 2002 Israel	Post hip fracture rehabilitation programme Interview	Retrospective cohort: an evaluation of the association between the CogFIM, MMSE, and Clock Drawing Task in the hip fracture population Baseline: Admission to rehabilitation following hip fracture surgery Follow-up: Discharge from rehabilitation	143	78.80 (8.00); range 52-98 years	73%	FIM Cog-FIM Motor-FIM	Arabic/Hebrew
3	Arinzon et al 2007 Israel	Post hip fracture rehabilitation programme Interview	Prospective cohort (no matched controls): an evaluation of the role of pain perception (pain Visual Analogue Scale) on functional recovery. Baseline: Admission to rehabilitation following hip fracture surgery Follow-up: Discharge from rehabilitation	63	78.01 (7.04)	68%	FIM Katz ADL Zung DRS	Arabic/Hebrew
4	Beaupre et al 2005 Canada	Community (Post-op hip fracture) Interview/telephone interview for follow-up	Prospective cohort: comparison of outcomes between two groups: 1) Treatment (Pathway) Group (TPG): consecutive patient cohort treated with care pathway 1999-2000; 2) Control(CG): consecutive cohort treated 1996-1997 (non-pathway) Baseline: Pre-fracture (retrospectively at time of fracture) Follow-up: Time 1: 4-6 days post-fracture; Time 2 at 3-months; Time 3 at 6-months	1) 451 2) 468	1) 81.70 (7.80) 2) 81.70 (7.60)	1) 78% 2) 77%	Modified Barthel Index (M-BI)	French Canadian

5	Bellelli et al 2008 Italy	Community (Post-op hip fracture) Interview/telephone interview for follow-up	Consecutive cohort – surgical admissions: patients screened for depression (GDS positive if score> 7/15)). Categorized into 4 groups: Dementia (Dem) (n=40) Depressed (Dep) (n= 54) Dementia and Depression (DD) (n= 27) Neither (NDD) (n= 90 = Reference group) Baseline: Admission to rehabilitation Follow-up: Time 1 at Discharge from Rehabilitation; Time 2 at 12-months	1) 40 2) 54 3) 27 4) 90	1. 83.50 (7.20) 2. 79.50 (7.00) 3. 85.10 (6.30) 4. 79.20 (7.00)	1. 92% 2. 87%, 3. 96% 4. 83%	BI GDS Lawton IADL	Italian
6	Beloosesky et al 2002 Israel	Community (Post-op hip fracture) Interview/telephone interview for follow-up	Prospective cohort: Categorised at baseline by: 1) Degree of independence pre-fracture (recall completion of Katz ADL) and 2) cognitive impairment (completion of MMSE). Baseline: Pre-fracture (recall). Follow-up: 6-months	153	81.30 (7.60) (median 82.0); range 65-102 years	75.8%	FIM	Arabic/Hebrew
7	Beloosesky et al 2004 Israel	Community (Post-op hip fracture) Interview/telephone interview for follow-up	Prospective cohort study: 86.5% surgical repair; 13.5% treated conservatively. Follow-up: at 4yrs.	123	77.40 (7.30); range 60-96 years	67% f	NHP	Arabic/Hebrew
8	Binder et al 1994 USA	Community (post-op) Interview	Randomized Clinical Trial (RCT) of rehabilitation post hip fracture: 1) Extended physiotherapy Vs 2) Low-impact home exercise Baseline; Follow-up 6-months	1)46 2)44	1) 80.00 (7.00) 2) 81.00 (8.00)	1) 72% 2) 77%	SF-36 FSQ OMFAQ	US English
9	Borgquist et al 1992 Sweden	Community (post-op hip fracture) Self-completed (Postal)	Prospective cohort: exploring quality of life and activities of daily living (ADL) in hip fracture population & population reference group: 1)Hip fracture after surgery: Groups defined as: a Complications vs no complications; b Able to walk outside vs Not able to walk outside);	100	1) 74.00	80%	NHP	Swedish

		To compare self-assessment of PROMs with objective status outcome assessments (ADL)	2)Population reference group Baseline: not reported Follow-up: Time 1 at 4mths (ADL only); Time 2 at 6mths (NHP only); Time 3 at 12-months (NHP only)					
10	Blomfeldt et al 2006 Sweden	Community (post op hip fracture) Interview	Prospective cohort: post-surgical repair of hip-fracture. Two groups: 1) Primary total hip replacement; 2) Secondary total hip replacement Baseline: pre-fracture (retrospective) Follow-up: Time 1 at 4-months; Time 2 at 12-months; Time 3 at 24-months.	1) 43 2) 41	1) 79.00 (5.00) 2) 80.00 (5.30)	1) 86% 2) 93%	EQ-5D Charnley HS	Swedish
11	Boonen et al 2004 Belgium	Community (post-op hip fracture) Self-completed (Postal)	Prospective matched control cohort: 1) Unselected consecutive post-op hip fracture; 2) Convenience sample, local population (no fracture) *Comparator measure = objective assessment of disability (Rapid Disability Rating Scale 2).* Baseline: before hospital discharge Follow-up: at 12-mths **Only 51% able to self-complete SF-36 at discharge and 12-mths (significantly younger, less Cognitive impairment, and better function)*	134	1) 78.00; range 50-95 years 2) 78.00; range 50-97 years	100%	SF-36	Flemmish
12	Buecking et al 2014 Germany	Hospital (admission & discharge post hip fracture) Interview	Prospective cohort: to identify independent factors (depression and mental status) correlated with quality of life in hip fracture population. Data collected at admission and discharge	227	81.00 (8.00) (median 82.00); range 60-99 years	27%	EQ-5D 3L EQ-VAS	German
13	Cornwall et al 2004 USA	Community (post op hip fracture) Interview	Prospective cohort: groups defined by fracture type: 1) Displaced Femoral neck fracture (FN); 2) Non-displaced FN fracture; 3) Unstable intertrochanteric fracture (IT)	1)181 2)70	1) 81.80 (8.80) 2) 78.40 (10.30)	1) 80% 2) 87%	FIM total	US English

		Patient or proxy (caregiver)	4) Stable intertrochanteric fracture. Baseline: Pre-fracture function (retrospective at admission). Follow-up: 2-months and 6-months	3)178 4)108	3) 82.90 (7.90) 4)82.50 (8.70)	3) 81% 4) 81%		
14	Cranney et al 2005 Canada *Comparative evaluation*	Community (post op hip fracture) Interview	Prospective matched control cohort: 1) Unselected consecutive post-op hip fracture; 2) Convenience sample, local population (no fracture) Baseline; Follow-up – 3 and 9 mths.	1)20 2)20	1) 80.00 2 79.00	100%	SF-36 HUI2 SF-6D	Canadian English
15	Deutch et al 2005 USA	Post hip fracture rehabilitation programme *Data collected from Medicare database*	Retrospective cohort: 1) Inpatient rehabilitation facilities programme Vs 2)Skilled nursing sub-acute rehabilitation programme Baseline: admission to rehabilitation programme Follow-up: Time 1 at discharge from rehabilitation programme	1)24,714 2) 5,079	1) 80.40 (8.20) 2) 82.10 (8.10)	1) 78% 2) 81%	FIM: Motor – FIM Cog - FIM	US English
16	Dorra et al 2002 USA	Post hip fracture rehabilitation programme Interview	Prospective cohort: two groups categorised according to FIM scores (re-calculated as Rehabilitation Effectiveness Score (change in FIM motor subscale (discharge-admission score) divided by max possible improvement (max possible score – admission score) x 100)): 1) Poor rehabilitation outcome (bottom quartile) 2) Good rehabilitation outcome (top quartile) Baseline: Admission to Rehab. Follow-up: Time 1 at discharge from rehabilitation programme	1) 69 2) 68	Total 79.30 (8.00); range 60-98 years 1) 81.90 (7.80); range 60-98 years 2) 75.70 (8.20); range 60-94 years	Total 79.4% 1) 84% 2) 75%	FIM: Motor-FIM	US English
17	Dupplis & Wikblad 2004 Sweden	Community (post op hip fracture) Interview	Prospective cohort: categorised as: 1) Hip fracture with Delirium post-surgery (D); 2) Hip fracture without delirium at baseline (No D). Reference group (general Swedish population ≥75 years)	1) 32 2) 83	1) 85.40 (5.30) 2) 82.60 (5.50)	NR	SF-36	Swedish

			Baseline: post-op Follow-up: 6-months					
18	Frihagen et al 2007 Sweden	Community (post op hip fracture) Interview	Randomized Controlled Trial (RCT) for acute displaced hip fracture: 1) Internal fixation (IF) Vs. 2) Hemiarthroplasty (HA) Baseline: not reported. Follow-up: 4/ 12/ 24-months	N: 137	1) 83.20 (7.65) 2) 82.50 (7.32)	1)78% 2)71%	EQ-5D 3L EQ VAS BI HHS	Swedish
19	Frihagen et al 2008 Sweden	Community (post op hip fracture) Interview	Prospective cohort (sample taken from RCT (Frihagen et al 2007)): 1) Complications group Vs. 2) No complications groups Baseline Follow-ups: 4/ 12-months	1) 23 2) 56	82.80 (7.48)	74%	EQ-5D 3L EQ VAS HHS	Swedish
20	Gjertsen et al 2011 Norway	Community (post op hip fracture) Interview	Retrospective cohort (data from Norwegian Hip Fracture Database): 1) Displaced fracture with internal fixation; 2) Displaced fracture with hemiarthroplasty; 3) Un-displaced fracture with internal fixation. Baseline: Pre-fracture (retrospective) Follow-up: 4 and 12-months.	Total: 1948 1)550 2)778 3)670	1) 81.00 (8.90) 2) 83.00 (7.00) 3) 81.00 (8.40)	1)68% 2)76% 3)80%	EQ-5D 3L	Norwegian
21	Goldstein et al 1997 USA	Post hip fracture rehabilitation programme Interview Patient or proxy	Prospective cohort: 1) Cognitively impaired Vs. 2) Not cognitively impaired Baseline: admission to rehabilitation Follow-up: Discharge from rehabilitation (average 3 weeks)	Total: 58 1)35 2)23	Total: 84.00 (6.70); range 71 to 99 years	83%	FIM	US English
22	Graham et al 2008 USA	Post hip fracture rehabilitation programme	Retrospective cohort: patients receiving in-patient rehabilitation post hip fracture repair. Comparison between ethnic groups: 1) White, 2) Black, 3) Hispanic, 4) Asian	6970	80.20 (8.00)	74%	FIM	US English

		Interview/telephone interview for FU	Baseline: admission to rehabilitation Follow-up: Discharge from rehabilitation; 3 and 6-months					
23	Graham et al 2014 USA	Post hip fracture Community Telephone interview	Prospective cohort: comparison between: 1) Patient-centred Care management model and 2) Matched controls Follow-up: 6 and 12-months	194 1) 97 2)97	1) 82.00 (9.00) 2) 82.00 (9.00)	1)72% 2)74%	EQ-5D 3L EQ VAS HHS	US English
24	Griffin et al 2015 UK	Post hip fracture Community Interview/telephone interview for FU	Prospective Cohort: comparison of quality of life post-surgical repair of hip fracture Baseline: retrospective assessment Follow-up: 4-weeks; 4-months; 12-months	403	83.1 (8.7)	73%	EQ-5D 3L OHS	English
25	Hajbaghery & Abbasinia 2013 Iran	Community dwelling elders- post surgical repair of hip fracture Self-completed (Postal)	Case-control study: 1) Hip fracture – minimum of 3-months before study; 2) Community sample of matched cases (no hip fracture). Comparison of quality of life by gender, age, living arrangements, income, marital status, education level, current job, number of chronic disorders.	1)70 2)70	1) 73.50 (8.07) 2) 72.80 (7.48)	56%	EQ-5D 3L	Persian
26	Hall et al 2000 Australia	Community dwelling elders – post-surgical repair of hip fracture Interview	Case control: 1) Post hip fracture now living in community; 2) Community sample (no hip fracture) matched by age and gender Investigation of functional independence and quality of life at 6 and 12-months post-surgical repair of hip fracture	1)92 2)92	1) 75.88 (9.12); range 54-93 years 2) 75.73 (9.03); range 52-94 years	65%	SF-36 M-BI FAI	Australian English
27	Hallberg et al 2004 Sweden	Community (post-op hip fracture) Interview	Prospective matched control cohort: 1) Post-op hip fracture (Hip F); 2) Vertebral Fracture (Vert F); 3) Humerus Fracture (Hum F);	40	Total patient population mean: 69.30 (5.20)	100%	SF-36	Swedish

			4) Forearm Fracture (FF); 5) Convenience sample (matched cohort) - local population (no fracture)(CS). Baseline: assessed median 89 days post fracture. Follow-up: (Hip F) 2years post fracture					
28	Hallberg et al 2009 Sweden	Community-based population (post-op hip fracture). Self-report	Prospective matched control cohort (7-year follow-up of earlier cohort study): 1) Post-op hip fracture (Hip F); 2) Vertebral Fracture (Vert F); 3) Convenience sample (matched cohort) - local population (no fracture)(CS)). Baseline: assessed median 89 days post fracture. Follow-up: (Hip F) 7-years post fracture	25	Total patient population mean: 75.00 (4.70); range 64-82 years	100%	SF-36	Swedish
29	Hershkovitz et al 2007 Israel	Post hip fracture rehabilitation programme Interview	Prospective cohort: assessment of factors affecting functional recovery at discharge Baseline: admission to rehabilitation Follow-up: discharge from rehabilitation programme	133	80.0 (6.60)	74%	FIM FIM Cog FIM motor	
30	Heruti et al1999 Israel	Post hip fracture rehabilitation programme Interview	Prospective cohort: evaluation of factors affecting functional gain at discharge from rehabilitation. Focus on cognitive status at admission. Baseline: admission to rehabilitation Follow-up: discharge from rehabilitation programme (mean 23.4 days).	204	80.00 (7.10); range 64-84 years	76.5%	FIM: Motor - FIM Cog-FIM	
31	Inngul et al 2013 Sweden	Community (post-op hip fracture) Interview	Randomized clinical trial (RCT): an evaluation of hip function, health related quality of life, surgical outcome and acetabular erosion. Randomised by surgery type: hemi-arthroplasty with either 1) unipolar or 2) bipolar head Baseline: pre-surgery – according to recall principle Follow-up: 12/ 24/ 48-months post-op.	59	86.10; range 79-100 years	1) 82% 2) 70%	EQ-5D HHS	Swedish

32	Jones, Miller & Petrella 2002 Canada	Community (post-op hip fracture) Interview/telephone interview for follow-up	Prospective cohort Baseline: admission to rehab. Follow-up: discharge from rehabilitation and 6-weeks (n=44 only)	100	82.40 (7.30)	83%	FIM	Canadian English
33	Jones & Feeny 2005 Canada	Community (post-op hip fracture) Interview Patient and Proxy	Prospective cohort: an evaluation of the agreement between 1) patient and 2) proxy responses. Baseline: 3-5 days after surgery. Follow-up: 1, 3, 6mths Proxy: Spouse (23%); Offspring (55%); Other (22%)	245	1) Patients: 80.50 (7.50) 2) Proxy: mean not reported: range <40 yrs (5%) to > 75 yrs (13%).	1) 73% 2) 72%	HUI2 HUI3	Canadian English
34	Jones & Feeny 2006 Canada	Community (post-op hip fracture) Interview/telephone interview for FU Patient or proxy	Prospective cohort: an evaluation of the agreement between 1) patient and 2) proxy responses. Baseline: 3-5 days after surgery. Follow-up: 1, 3, 6mths Proxy: Spouse (27%); Offspring (33%); Other (20%).	137	1)Patients 79.40 (7.40) 2) Range <40 years to 75+ years	1) 73% 2) 93%	FIM total	Canadian English
35	Jones et al 2014 UK	Community (post-op hip fracture) Interview/telephone interview for FU	Prospective cohort: an evaluation of the longitudinal validity of HU12 & HU13 in patients recovering from hip fracture Baseline: 3-5 days after surgery Follow-up: 1 and 6-months	278	80.20 (7.50)	72%	HU12 HU13	English
36	Jongjit et al 2003 Thailand	Community (post-op hip fracture) Interview	Cross-sectional case-control study: 1) Community dwelling hip fracture patients (6 to 10-months post hip fracture repair); 2) Community dwelling age and gender matched controls (without hip fracture)	60	1) 75.88 (9.12) 2) 75.73 (9.03)	60%	FIM total FAI SF-36	Thai
37	Kirke et al 2002 Ireland	Community (post-op hip fracture) Interview	Prospective cohort: 1) Hip fracture patients; 2) Matched community controls	106	1) 79.95 2) 77.80	100%	Katz ADL	Ire English

			Baseline: Pre-fracture (collected retrospectively) Follow-up: 2yrs					
38	Latham et al 2009 Eight countries (Norway, UK, Sweden, Israel, Germany, USA, Denmark, Spain). *several translations	In-patient rehabilitation and out-patient rehabilitation follow-up – post hip fracture repair. Interview	Prospective cohort: an evaluation of self-report and performance-based measures of physical function in the hip fracture population. *3 self-report measures:* SF-36 Activity Measure Post-Acute Care Physical Mobility (AM PAC PM) AM PAC Personal Care (AM PAC PC) *4 Performance-based measures:* Physical Function Performance (PFP-10) Short Physical Performance Battery (SPPB) Four-meter gait speed (GS) Six-minute walk test (6MWT). Baseline: within 17-days post-op Follow-up: 12-weeks post-op	108	78.90 (8.10)	73.2%	SF-36	*several translations
39	Lenze et al 2007 USA	Community (post-op hip fracture) Interview	Prospective cohort: Hip fracture population with: 1) Major depressive disorder (MDD) Vs 2) No major depressive disorder (No MDD) Baseline: discharge from hospital Follow-up: 2 /6 / 10/ 14/ 18/ 22/ 26 weeks post discharge.	1) 18 2)108	1) 78.30 (10.80) 2) 81.80 (8.70)	1) 83% 2) 79%	HAM-D	US English
40	MaCauley et al 2008 USA	Community (post-op hip fracture) Interview	Randomized Controlled Trial (RCT) following traumatic hip fracture: 1)Total hip replacement (THR) Vs 2) Hemi-arthroplasty Baseline – not reported Follow-up: 6/ 12/ 24 months.	1)17 2) 23	1) 82.00 (7.00) 2) 77.00 (9.00)	1) 41% 2) 61%	SF-36 WOMAC HHS	US English
41	McGilton et al 2009	Rehabilitation	Cohort study: post-surgical repair of hip fracture:				FIM Motor	Canadian

	Canada	programme post hip fracture repair Interview	Groups defined by level of cognitive impairment (MMSE): 1) Cognitively impaired Vs 2) Not cognitively impaired Baseline: Admission to rehab. Follow-up: Discharge from rehab	1)17 2)14	1) 88.60 (5.70); range 71-100 years 2) 85.30 (7.80); range 77-100 years	1) 50% 2) 65%	FIM Cog	English
42	Mattsson et al 2005 Sweden	Community (post-op hip fracture) Interview	Randomized Controlled Trial (RCT) following traumatic hip fracture: 1) Sliding-screw + re-absorbable cement Vs. 2) Sliding-screw only. Baseline: 1-week post-op Follow-up: 6-weeks, 6-months	1)55 2)57	1) 81.20 (7.00) 2) 82.00 (6.30)	1) 80% 2) 82%	SF-36	Swedish
43	Mendelsohn et al 2003 Canada	Rehabilitation programme post hip fracture Interview	Prospective cohort: evaluate the relationship between functional mobility (Timed Up and Go [TUG], Self-Paced Walking[SPW], Berg Balance Scale [BBS]) and global functional status(FIM). Baseline: Admission to rehab. Follow-up: Discharge from rehab.	N=40	79.3 (4.5)	85%	FIM	Canadian English
44	Mishra et al 2004 UK	Community (post-op hip fracture) Telephone interview	Retrospective cohort (post-op): consecutive patients (socially independent and mentally alert) receiving a total hip replacement (THR) for displaced sub-capital fracture. Follow-up: mean 30-months post-op (range 20-54 months)	51	74.0 (7.2)	88%	Oxford HS SF-12 PCS	UK English
45	Mizrahi et al 2007 Israel	Post hip fracture rehabilitation programme Interview	Retrospective cohort: Impact of Stroke (ICD classification) on rehab / functional gain of patients with hip fracture. Groups: 1) Previous stroke; 2) No previous stroke Baseline: Admission to rehabilitation. Follow-up: Discharge from rehabilitation	460 1)51 2)409	1) 81.84 (6.27) 2) 82.23 (6.96)	1) 45% 2) 78%	FIM total FIM Motor	

46	Mouzopoulos et al 2008 Greece	Community (post-op hip fracture) Interview	Prospective cohort: Between group comparison of functional ability. Patients 'randomly' assigned to receive: 1) Total arthroplasty (TA); 2) Hemiarthroplasty (HA) or 3) Internal Fixation (IF) Baseline: pre-fracture retrospective assessment. Follow-up: 1 and 4-years.	1)23 2)20 3)19	1)73.07 (4.93) 2)74.24 (3.77) 3)75.38 (4.62)	1) 76% 2) 71% 3) 68%	BI HHS	Greek
47	Nguyen-Oghalai et al 2006 USA	Rehabilitation programme post hip fracture Interview / telephone interview	Retrospective cohort (national registry of medical rehab in-patients): to assess the impact of osteoarthritis (OA) on the length of rehabilitation stay and functional recovery (assessed with the FIM) in patients sustaining a hip fracture: Groups: 1) Hip fracture with OA Vs. 2) Hip fracture without OA Baseline: Admission Follow-up: Discharge (interview); 80-180 days after discharge from rehabilitation (telephone)	13,394; 1)1953 2)11441	1) 80.70 (7.20) 2) 80.10 (7.10)	1) 82% 2) 76%	FIM	US English
48	Orive et al 2015 Spain	Community (post-op hip fracture) Interview/telephone interview for FU	Prospective cohort: to evaluate changes in health-related quality of life and activities of daily living in: 1) Hip fracture patients 2) Non-hip fracture patients Baseline Follow-up: 6-months	891 1)776 2)115	1) 83.16 (7.05) 2) 73.18 (6.43)	1) 82% 2) 53%	SF-12 WOMAC BI Lawton ADL	Spanish
49	Parsons et al 2014 UK	Community (post-op hip fracture) Interview/telephone interview at follow-up	Prospective cohort: comparative evaluation nof PROMs in hip fracture population. Groups defined by cognitive status: 1) Cognitively impaired 2) Cognitively intact Baseline Follow-up: 4 weeks; 4 and 12-months	225	83.1 (7.94) 1) 85.6 (6.37) 2) 82.0 (8.28)	75% 1) 79% 2) 72%	EQ-5D EQ VAS OHS ICECAP-O	UK English

50	Randell et al 2000 Australia	Community (post-op hip fracture) Interview	Prospective matched control cohort: 1) Community dwelling post-op hip fracture patients (unselected, consecutive); 2) Case-matched convenience sample, local population (no fracture) Baseline: within 1-week of fracture (recall pre-fracture status) Follow-up: 12 to 15-weeks	32	1) 82.00 (8.00); range 68-97 years 2) 86.00 (6.00); range 68-98 years	1) 69% 2) 72%	SF-36 (v1) OPAQ-2	Australian English
51	Rohde et al 2010 Norway	Community (post-op hip fracture) Interview	Prospective matched control cohort: 1) Community dwelling post-op hip fracture patients (unselected, consecutive); 2) Case-matched convenience sample, local population (no fracture) Baseline: within 4-days of fracture (recall pre-fracture status) Follow-up: 1 and 2-years	61	1) 74.0 (10.0) 2) 73.0 (8.0)	75% f	SF-36 QoLS	Norwegian
52	Rolland et al 2004 France	Post hip fracture rehabilitation programme Interview	Prospective cohort study: to assess functional gain in hip fracture patients (post-op) participating in a rehabilitation programme. Groups categorised according to cognitive status: 1) Cognitively impaired (CI) 2) Partially cognitively impaired (PCI) 3) Not cognitively impaired (No CI) *Outcome measures: FIM. Completed by whole team (Geriatrician, PT, Psychologist; Geriatric Nurse).* Baseline: Admission to rehabilitation programme. Follow-up: Discharge from rehabilitation programme.	61 1)28 2)23 3)10	1) 87.60 (7.20) 2) 83.90 (6.80) 3) 77.60 (7.40)	82%	FIM	French
53	Shepherd et al 1996 UK	Community (post-op hip fracture) Interview	Prospective cohort: to explore the feasibility and acceptability of range of assessment scales in hip fracture population	270	81 (8)	80%	GDS BI PGCMS	UK English

		Patient and proxy	Measures: GDS, Philadelphia Geriatric Centre Morale Scale (PGCMS) and Barthel Index (BI) Baseline: recall of pre-fracture status (retrospectively collected at 3-4 days after fracture). Follow-up: 1 / 6/ and 12-months *Evidence of feasibility, acceptability, data quality, responsiveness. Plus issues re proxy completion*					
54	Shyu et al 2004a Taiwan	Community (post-op hip fracture) Interview	1) Prospective cohort: to explore data quality, responsiveness, validity of selected PROMs 2) Results from a Randomized Controlled Trial (RCT) to inform an evaluation of group differences. *Primary outcome: evaluation of the SF-36 evaluation in two groups of elders post hip fracture.* Measures: SF-36 (Taiwanese), Chinese Barthel Index, Lawtons IADL, Chinese GDS(SF) Baseline: before discharge from hosp. Follow-up: 1/ 3/ and 6-months post-op.	116	1) 79.80 (7.20) 2) 78.51 (8.27)	1) 64% 2) 71%	SF-36 (v1) BI Lawtons ADL	Taiwanese/ Chinese version
55	Shyu et al 2004b Taiwan	Community (post-op hip fracture) Interview	Prospective cohort: an longitudinal evaluation of change in activities of daily living post hip fracture. Measures: Chinese Barthel Index, Lawtons IADL. Follow-up: 1/ 3/ 6/ 12-months post hip fracture.	110	79.40 (7.50)	60.9%	BI Lawtons ADL	Taiwanese/ Chinese version
56	Shyu et al 2008 Taiwan	Community (post-op hip fracture) Interview	Randomized Controlled Trial (RCT) of rehabilitation post hip fracture: 1) Intervention (interdisciplinary team) Vs 2) Control (usual care) Measures: Chinese GDS and Chinese Barthel index (primary outcome). Baseline: before hospital discharge. Follow-up: 1/ 3/ 6/ 12-months post-op	162 1) 80 2) 82	Total: 78.16 (7.76)	Total:68.5%	GDS BI	Taiwanese/ Chinese version

57	Shyu et al 2009 Taiwan	Community (post-op hip fracture) Interview	Prospective cohort: focus on depression post hip fracture. Measures: Chinese GDS and Chinese Barthel index (primary outcome). Baseline: before hospital discharge. Follow-up: 1/ 3/ 6/ 12-months post-op.	147	77.90 (7.90)	67.3%	GDS BI	Taiwanese/ Chinese version
58	Shyu et al 2013 Taiwan	Community (post-op hip fracture) Interview	Prospective cohort: comparing the effects of interdisciplinary comprehensive care programmes: 1) Subacute care group 2) Comprehensive care group 3) Usual care Baseline: Pre-fracture Follow-up: 1/ 3/ 6/ 12-months post-op	269 1) 92 2) 92 3) 85	Total: 76.20	Range: 60 to 67.3%	SF-36	Taiwanese/ Chinese version
59	Soderqvist et al 2006 Sweden	Community (post-op hip fracture) Interview Patient and proxy	Prospective cohort: two groups defined by cognitive status: 1) Cognitively impaired (Short Portable Mental Status Questionnaire (SPMSQ) scores <3); 2) Not cognitively impaired (SPMSQ scores >3) Follow-up: T1=4mths, T2=12mths	213 1)50 2)163	Total: 84.00; range 65-99 years 1) 86.10 (5.50) 2) 82.80 (6.70)	81%	EQ-5D Katz ADL Charnley – HS (walk; pain)	Swedish
60	Tidermark et al 2002a Sweden	Community (post-op hip fracture) Interview	Prospective cohort: groups defined by hip fracture type 1) Un-displaced fracture Vs 2) Displaced fracture Baseline: Pre-op retrospectively reported Follow-up: 4/ 12/ 24-months	90 1)24 2)66	 1) 80.00 (8.00) 2) 80.10 (6.90)	 1)71% 2) 74%	EQ-5D Katz ADL	Swedish
61	Tidermark et al 2002b Sweden	Community (post-op hip fracture) Interview	Prospective matched cohort: groups defined as: 1) Healed fracture 2) Healing complications after fracture 3) Age-matched Swedish reference population Baseline: Pre-op (retrospective assessment)	90	80.00 (7.30); range 66-92 years	76%	EQ-5D Katz ADL	Swedish

			Follow-up: 1 week/ 4 and 17-months					
62	Tidermark et al 2003 Sweden	Community (post-op hip fracture) Interview	Randomized Controlled Trial (RCT) of surgical repair in hip fracture patients: 1) Total Hip Replacement Vs. 2) Internal fixation *Focus: evaluation of comparative responsiveness of EQ-5D and SF-36* Baseline: Pre-op (retrospective assessment) Follow-up: 4-months	95 1)48 2)47	 1) 79.20 (5.00) 2) 80.80 (6.60)	 1) 81% 2) 81%	EQ-5D SF-36 Katz ADL Charnley Hip Score	Swedish
63	Tidermark et al 2007 Sweden	Community (post-op hip fracture) Interview	Randomized Controlled Trial (RCT) of surgical repair in hip fracture patients: 1) Total Hip Replacement Vs. 2) Internal fixation *Focus: evaluation of comparative responsiveness of EQ-5D and NHP* Baseline: Pre-op (retrospective assessment) Follow-up: 6-months	59	83.00 (5.00); range 70-92 years	100%	EQ-5D NHP	Sweden
64	Tosteston et al 2001 USA	Community (post-op hip fracture) Interview	Cohort study: three groups of women: 1) No fragility fracture; 2) Vertebral fracture (no hip fracture); 3) Hip fracture (within last 1-5yrs). Cross-sectional assessment of current status	67	 1) 67.40 (0.60) 2) 73.40 (0.80) 3) 80.30 (1.10)	100%	SF-36	US English
65	Tsauo et al 2005 Taiwan	Community (post-op hip fracture) Interview	Randomized Controlled Trial (RCT) of rehabilitation post hip-fracture repair: 1) Home-based physiotherapy Vs 2) Control group (usual care). Baseline: discharge from hospital Follow-up: 1/ 3/ 6-months post discharge.	25 1)13 2)12	 1)74.10 (12.00) 2) 71.90 (12.50)	80%	WHOQOL-BREF HHS	Taiwanese
66	Van Balen et al	Community (post-op	Evaluative comparison of health status measures:	208	Mean, medium,	79%	BI	Dutch

	2003 Netherlands	hip fracture) Interview or Self-complete	data quality, reliability, validity, responsiveness. Interview administered: Barthel Index; Rehabilitation Activities profile. Self-administered: NHP and COOP Patient sample consisted of early discharge (106) & normal discharge (102) post-surgical repair of hip fracture. Baseline: pre- fracture (recall) Follow-up: 1 week/ 1 and 4-months		25th & 75th Percentile: 83 years, 84 years, (77-89 years)		NHP COOP/WONCA	
67	Vergara et al 2014 Spain	Community (post-op hip fracture) Postal self-report	Prospective cohort: evaluation of factors influencing functional recovery (socio-demographic data (age, gender, instruction level, living condition, received help), comorbidities, characteristics of the fracture, treatment, destination of discharge, health-related quality of life)	638	83.20 (7.20)	84%	SF-12 (v1) WOMAC BI Lawton IADL	Spanish
68	Vidan et al 2005 Spain	Community (post-op hip fracture) Interview/telephone interviews for FU	Randomized controlled trial (RCT) of rehabilitation post hip-fracture (during the acute care, in-patient period): 1) Comprehensive geriatric intervention Vs 2) Usual care Baseline: within 72 hours of admission (unclear if pre-fracture status is collected retrospectively) Follow-up: hospital discharge; 3/ 6/ 12-months	250	1) 81.10 (7.80) 2) 82.60 (7.40)	1) 85% 2) 79%	Katz ADL	Spanish
69	Ziden et al 2010 Sweden	Community (post-op hip fracture) Interview	Randomized controlled trial (RCT) of rehabilitation post hip-fracture: 1) Home rehabilitation programme (HR) Vs 2) Conventional care (CC). Comparator measures: Physical mobility measures (TUG, STS) Mood (CES-D) ADL (FIM and FAI) **Study focus on FIM, FAI, TUG and STS. No evidence of measurement or practical properties for CES-D*	102 1) 48 2) 54	1) 81.20 (5.90) 2) 82.50 (7.60)	1) 60% 2) 78%	FIM FAI (SF-36)	Swedish

			Baseline: recall and baseline data Follow-up: 1/6/12-months					
70	Zielinski et al 2014 Netherlands	Community (post-op hip fracture) Interview	Prospective cohort (follow-up from FAITH RCT study): surgical repair of hip fracture: 1) Internal Fixation Vs 2) Salvage Anthroplasty Data collected at 2 years	248 1)164 2) 68	 1) 70.00; range 62-78 years 2) 72.00; range 66-79 years	 1) 55% 2) 69%	SF-36 WOMAC	Dutch
71	Zlowodski et al 2008 4 centres: North America (USA and Canada), England, Denmark.	Community (post-op hip fracture) Telephon e interview.	Retrospective cohort study post-hip fracture repair (identified from hospital database): Three groups defined as: 1) Severe shortening of femoral neck; 2) Moderate shortening of femoral neck; 3) No/Mild shortening of femoral neck *Comparator measures: SF-36 (PF primary outcome) and EQ-5D; Radiographic assessment.* Assessment point: follow-up to surgery: mean 20-months (range 5 – 105 months).	70	71.00; range 20-90 years	74%	SF-36 EQ-5D	US English

Footnotes:
Post-op = the post-operative period following surgical repair of a hip fracture
NR = Not reported
RCT = Randomized Controlled Trial