Cardiac Arrest - PROMs review – Resuscitation.

ACCEPTED 27/11/2017 – Copy for WRAP

Manuscript Title:

Assessing Health-Related Quality of Life (HRQOL) in Survivors of Out-of-Hospital Cardiac Arrest: a systematic review of patient-reported outcome measures.

Authors:

Kirstie L Haywood[1], Nathan Pearson[1], Laurie J Morrison[2], Maaret Castren[3], Gisela Lilja[4,5], Gavin D Perkins[6.]

Affiliation:

1. Warwick Research in Nursing, Department of Health Sciences, Warwick Medical School, The University of Warwick, Gibbet Hill, Coventry. CV4 7AL.
2. Rescu, Li Ka Shing Knowledge Institute, St Michael's Hospital, Division of Emergency Medicine, Department of Medicine, University of Toronto, Ontario Canada.
3. Department of Emergency Medicine and Services, Helsinki University Hospital, Helsinki, Uusimaa, Finland.
4. Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Sweden.
5. Department of Clinical Sciences, Division of Neurology, Lund University, Lund, Sweden.
6. Warwick Clinical Trials Unit, Warwick Medical School, The University of Warwick, Gibbet Hill, Coventry. CV4 7AL.

Corresponding author:

Dr Kirstie L Haywood. Senior Research Fellow. Warwick Research in Nursing, Department of Health Sciences, Warwick Medical School, The University of Warwick, Gibbet Hill, Coventry. CV4 7AL.

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

Telephone: 024761 59616

Abstract 246/250; Main text 4110.

Refs 65; Tables 5; Figures 1; Appendices 3.

Key words: Cardiac Arrest; psychometrics; health-related quality of life; patient-reported outcome;

Abstract (246/250)

Aim: High quality evidence of out-of-hospital cardiac arrest (OHCA) survivors' health-related quality of life (HRQoL) can measure the long-term impact of CA. The aim of this study was to critically appraise the evidence of psychometric quality and acceptability of measures used in the assessment of HRQoL in cardiac arrest survivors.

Methods: Systematic literature searches (2004-2017) and named author searches to identify articles pertaining to the measurement of HRQoL. Data on study quality, measurement and practical properties were extracted and assessed against international standards.

Results: From 356 reviewed abstracts, 69 articles were assessed in full. 25 provided evidence for 10 measures of HRQoL: one condition-specific; three generic profile measures; two generic index; and four utility measures. Although limited, evidence for measurement validity was strongest for the HUI3 and SF-36. However, evidence for reliability, content validity, responsiveness and interpretability and acceptability was generally limited or not available in the CA population for all measures.

Conclusions: This review has demonstrated that a measure of quality of life specific to OHCA survivors is not available. Limited evidence of validity exists for one utility measure – the HUI3 - and a generic profile – the SF-36. Robust evidence of the quality and acceptability of HRQoL measures in OHCA was limited or not available. Future collaborative research must seek to urgently establish the relevance and acceptability of these measures to OHCA survivors, to establish robust evidence of essential measurement and practical properties over the short and long-term, and to inform future HRQoL assessment in the OHCA population.

Background

The importance of seeking to better understand and assess the long-term impact of cardiac arrest on survivors is evidenced by the recent inclusion of quality of life (QoL) and patient-reported outcomes (PRO) as supplementary outcomes in the standardised reporting frameworks for observational studies drawn from resuscitation registries.[1] However, historically, PROs and QoL have rarely been reported in resuscitation research,[2] and assessment guidance for this population is lacking. Moreover, the concept of survival to a good QoL has been poorly explored from the perspective of survivors of out-of-hospital cardiac arrest (OHCA) and their care givers'.

Well-developed patient-reported outcome measures (PROMs) are questionnaires containing one or more items, designed to provide a structured assessment of an individual's health-related quality of life (HRQoL). PROMs may be generic – containing items reflective of the broad concepts of HRQoL and therefore applicable to the general population – or specific – to a condition (for example, traumatic brain injury), aspect or domain of health (for example, fatigue), or population (for example, children). Patient-derived, specific measures are expected to be both more relevant and responsive to important changes in health than their generic counterparts, with their combined use therefore recommended.[3] An earlier review of quality of life after cardiac arrest described significant heterogeneity in HRQoL reporting, listing more than 40 measures of either general HRQoL or specific health domains.[4] Although the quality and acceptability of these measures was not assessed – and hence assessment recommendations not made – recommendations included a need for greater standardisation in HRQoL measurement with which to support study comparison.

In this systematic review, we aim to critically appraise and summarise published evidence of the quality and acceptability of clearly defined, multi-item patient-reported measures of HRQoL following completion by OHCA survivors or appropriate proxy. The evidence synthesis will provide a transparent evaluation with which to inform measurement selection for future application in clinical research, cardiac arrest registries and audit, and routine practice.

Methods

Identification of studies and measures: search strategy

Medical subject headings (MeSH terms) and free text searching was used to develop terms reflective of: 1) population – cardiac arrest; 2) assessment type – including patient-reported outcome measures; HRQoL; and 3) measurement and practical properties (Appendix 1).[5-7]

Two databases were searched (MEDLINE, EMBASE ((OVID)); 2004 to March 2017) (figure 1). A named author search was also conducted. Citation lists of included articles and measurement reviews in resuscitation research were also reviewed.[2,4,8]

Study inclusion / exclusion

All study designs were included if they provided evidence of measurement and/or practical properties (summarised below; detailed Table X) for clearly defined and reproducible multi-item, patient or proxy completed measures of HRQoL, following completion in the target population of OHCA survivors.

Titles and abstracts of potential articles were assessed for eligibility by one experienced reviewer (KH). Full-text articles were retrieved and selected based on English Language, and publication in peer reviewed journal. Abstracts, conference proceedings and studies pertaining to domain specific, diagnostic and screening measures were excluded. The list of included studies was checked for completeness by the review co-authors.

HRQoL measures were categorised as: specific (condition or population) or generic (profile; utility).

Data extraction

Data extractions were informed by established guidance for measurement evaluation,[6,9,10] published reviews,[7,11,12] and the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist.[13,14] Both study and PROM-specific information was extracted. Evidence on the following measurement properties was sought (Table 1): validity (content; construct – structural; convergent / divergent; discriminant); reliability (internal consistency; test-retest; measurement error); responsiveness; interpretation (including data quality (end-effects; change scores); smallest detectable difference (SDD), within-person (minimal important change (MIC)) and between-group (minimal important difference (MID)).[6,9,15] Although not included within the COSMIN framework, addition evidence of practical properties - acceptability (respondent burden; relevance) and feasibility (time,

cost)[11,15,16] - and the involvement of patients as active research partners[11,16,17] was sought. Data were extracted by a single, experienced reviewer (KH).

To inform a comparative analysis of PROM content, individual questions (items) were considered as per the revised Wilson and Cleary HRQoL Model: that is, symptoms, functional status (physical, cognitive, psychological, and social/role) and general health perceptions.[18,19]

Assessment of study methodological quality

Study methodological quality was evaluated per measurement property on the 4-point COSMIN rating scale (excellent, good, fair, poor) and determined by the lowest rating in each assessment section.[14]

Assessment of PROM quality

Each measurement property was judged against an existing checklist (Table 1).[10-12] Evidence was graded as: adequate (+) [reaches accepted standards]; conflicting (+/-); inadequate (-) [does not research accepted standards]; or indeterminate (?) [the results are difficult to define]. Whilst the 'quality' of evidence detailing practical properties was not determined, the extent of active patient engagement was graded (Table 1).[11,16,17]

Qualitative data synthesis

The data synthesis combines four factors: 1) number of studies reporting evidence per measure; 2) the degree of consistency between evaluations; 3) the quality of the reported measurement property (Table 1); and 4) study methodological quality (COSMIN scores).[11,20] The final quality score has two elements: first, the overall quality of a measurement property is reported as: adequate (+), conflicting (+/-), inadequate (-), or indeterminate (?). Second, evidence is categorized: 'strong', 'moderate', 'limited', 'conflicting', or 'unknown'.[20]

Results

Identification of studies and measures

Study and PROM identification is summarised as per PRISMA guidance (figure 1); www.prisma-statement.org). Twenty-five articles (Appendix 2) provided limited evidence for ten measures (Tables 2 [21-31] and 3 [32-56]): one condition-specific – the Quality of Life after Brain Injury questionnaire (QOLIBRI)[21] - and nine generic: three profile– the Short-Form 36-item Health Survey (SF-36 version 1 (v1))[22], it's modification (SF-36 version 2 (v2))[23], and the Short-Form 12-item version (SF-12v2)[24]; two index – the Life Satisfaction 11-item Checklist (LiSat-11)[25] and the Quality of Life Scale (QOLS)[26,27]; and four utility – the 15-dimension health-related quality of life instrument (15D)[28], EuroQol EQ-5D-3L[29], Health Utility Index version 3 (HUI3)[30], and the SF-6D[31] (Table 2). The SF-36 versions differ in the number of response options for seven role limitation items (Table 2); evidence suggests that v2 has improved reliability, validity and responsiveness.[23] Evidence for the two versions will be presented separately, but collectively appraised for the SF-36. The longest measure was the QOLIBRI (37-items); the shortest the EQ-5D-3L (5-items).

Patient and study characteristics (Appendix 1)

Sample sizes ranged from 20 to 1188; age ranged from 18 to more than 75 years. Studies were from international settings: mostly North America (n 7/25) and the Netherlands (n 7/25). Most were cross-sectional evaluations – trial sub-studies or cohort studies. Three were specific to PROM psychometric evaluations.[46,52,53]

Measurement properties and methodological quality

Study methodological quality is reported per reported measurement property per PROM (Table 3) (data extraction in Appendix 2); the evidence synthesis is reported in Table 4 (detailed in Appendix 3).

Validity - Content validity

The comparative analysis highlights similarities and discrepancies in PROM content and assessment focus (Table 2).

Symptoms – Except for the two generic index measures, all measures assess symptoms, including pain and/or discomfort. Most (6/10) assess fatigue – the SF-36 provides the most detailed assessment (2 items); most include just single items.

Functional status – physical: All measures assess physical function – the SF-36 and SF-12 include the greater number of items with this focus. Two of the five EQ-5D-3L items assess limitations in self-care and mobility, whilst 2/8 HUI3 items assess ambulation (walking distances) and dexterity. Several measures assess vision (QOLIBRI, 15D, HUI3), hearing (QOLIBRI, 15D, HUI3) and speech (15D, HUI3).

Functional status – cognition: Just 3/10 measures specifically assess cognitive impairment – the QOLIBRI includes 7 items, whilst the HUI3 and 15D include just single items.

Functional status – psychological: All measures assess mental or emotional well-being. The QOLIBRI includes several items that explore an individual's 'view of self' and concerns about anxiety, anger, depression and loneliness. The SF-36 and SF-12 also include multiple items to assess mental health concerns. The 15-D includes two items; the LiSat-11, QOLS, EQ-5D-3L and HUI3 include single items.

Functional status – social/role: Apart from the HUI3, all measures assess limitations in, or satisfaction with, social or role functioning. The QOLIBRI provides a detailed assessment of the impact of brain injury on daily life and autonomy. The SF-36 includes items relating to social function. The remaining measures provide a more limited assessment, including items that reflect changes in self-care and usual activities.

General health perception: Just five measures – the QOLIBRI, SF-36, SF-12, LiSat-11 and QOLS - include items pertaining to perceived well-being.

Validity – Construct validity (structural; construct)

Evidence of construct validity was limited (Tables 3 and 4). Five studies reported acceptable evidence of convergent or known-groups validity for the HUI3 (Table 3). One good quality study demonstrated the ability of the HUI3 to discriminate between survivors per duration of resuscitation.[50] Two smaller studies reported the ability of the HUI3 to distinguish between survivors grouped per CPC scores at 1-year post-arrest,[53] or when identified as 'fully recovered' versus 'dependent'.[54] Limited evidence also suggests that the 15D can discriminate between survivors grouped per CPC or mRS scores at 6-months.[49] Small levels of association between the HUI3 and measures of global disability such as the CPC (pre/at hospital discharge),[50,52] the GOS-E and mRS[52] have been reported. However, this association appears to increase when CPC is assessed post-discharge.[52] Similarly, small to moderate levels of association have also been reported between the 15D and the CPS and mRS at 6-months.[49] Despite the inclusion of a cognition-specific item, only small levels of association have been reported between the HUI3 and a cognition-

specific measure (Telephone Interview of Cognitive Status (r= 0.37)); and small to moderate levels (range r=0.23 to r=-0.45) with measures of psychological well-being.[55] More moderate associations have been reported with measures of self-care and reintegration into 'normal living',[55] suggesting a stronger focus on physical function.

Seven studies report acceptable (judged to be of at least 'fair' quality) evidence of validity for the SF-36, SF-12 and SF-6D (Table 3). As hypothesised, associations between the SF-12 summary scores and global measures of disability (GOSE with MCS r= 0.31; with PCS r= 0.44) were smaller than with other generic measures of health status (PCS with EQ-5D-3L r= 0.63; with SF-6D r= 0.69; MCS with SF-6D r= 0.56).[46] Similar levels of association were reported between both SF-12 summary scores and the SF-6D – suggesting that the SF-6D has an equal focus on both physical and mental well-being. The significantly stronger association between the SF-12 PCS and EQ-5D-3L (r= 0.63) than with the MCS (r= 0.29) highlights the greater focus of the EQ-5D-3L on physical limitations. However, a moderate to strong association between the EQ-5D-3L and the SF-6D was reported (r= 0.65) suggesting that they assess similar, although not identical, aspects of health status. Similarly, the moderate association between the SF-6D and GOSE at 12-months following the arrest (r= 0.52), suggests that whilst there is assessment overlap, there is substantial divergence in measurement focus.

Moderate associations between both the SF-36 PCS and MCS and the Fatigue Severity Scale (range r= 0.57 to r= 0.61) highlight the multi-faceted nature of fatigue.[36] Strong associations between the SF-36 MCS and the Hospital Anxiety and Depression Scale (range r= -0.77 to -0.84) supports the ability of the MCS to capture aspects of emotional well-being in this population.[36,42] The moderate association between the SF-36 PCS and the Cognitive Failures Questionnaire (r= 0.55) suggests that, whilst not specifically including items about cognition, the PCS is influenced by aspects of cognitive impairment.

Acceptable evidence details the ability of the SF-12 MCS and PCS, the SF-6D, and EQ-5D-3L to discriminate between survivors of cardiac arrest who return to work at 12-months, and on gender (lower health state for females).[46] Scores on the SF-12 PCS and SF-6D discriminate between survivors discharged to home or an alternative at 12-months. Scores across 6/8 SF-36 domains (not bodily pain or general health) discriminate between survivors with and without cognitive impairment, with unimpaired survivors reporting better levels of well-being.[40] However, scores on the QOLS were unable to discriminate between survivors with or without anoxic brain injury.[48]

Several studies report the ability of measures to discriminate between the health status of OHCA survivors and that of the general population. Both SF-36 summary scales demonstrate worse health state in OHCA survivors at 3-years when compared to that of the Dutch general population,[36] and at 5-years when compared to an age- and gender-matched German population.[34] At 12-months, evidence suggests that the SF-12 PCS for OHCA survivors is lower, but the MCS equivalent, to both the Australian[45] and Dutch[44] population norm. However, evidence suggests that the MCS is lower for younger Australian survivors at 12-months than age-matched members of the general population.[43] Similarly, a deteriorating state of mental well-being has been reported at 12-months (when compared to that reported at 3-months), with scores significantly lower than that of an age- and gender-matched 'norm' population.[42] However, non-statistically significant between group differences for OHCA survivors and the general population have also been reported on the SF-36.[41] There was no difference in 15D utility scores for OHCA survivors at 6-months and the general Finnish population,[49] although statistically significant differences in two single domains - usual activities and sexual activities – were reported.

Limited evidence of convergent validity was reported for the QOLIBRI following completion by a small group of survivors with neurological impairment referred for rehabilitation at 2-years,[32] and a larger group at 6-months post arrest.[33] Although a small study, the strong association between the QOLIBRI and cognition (Cognitive Failures Questionnaire (range r= 0.77 to r= 0.86) and the moderate to strong association with a measure of participation and autonomy (range r=-0.44 to r=-0.86) supports the ability of the QOLIBRI to capture issues associated with cognitive impairment and participation in OHCA survivors with neurological impairment.[32]

Limited evidence suggests that there are gender differences for psychological health when assessed with the LiSat-11, with more men reporting greater satisfaction with their health.[47]

The structural validity of the reviewed measures was not reported.

Reliability (internal consistency; test-retest; measurement error):

Limited evidence of acceptable inter-rater reliability was reported for the HUI3 following completion by just ten patients.[55] Limited evidence of acceptable internal consistency reliability was reported for all domains of the SF-36 following completion by 81 patients.[34] Additionally, there is limited evidence of internal consistency for the LiSat-11,[47] and QOLS.[48] Evidence of measurement reliability and measurement error was not identified for the remaining measures (Tables 3 and 4).

Responsiveness

Statistically significant between group differences were reported for several SF-36 domains - role emotional, mental health and general health - at 12-months in favour of survivors who had participated in the active arm of a rehabilitation trial; of note, there was no between group difference at 3-months.[39] Responsiveness statistics were not reported. A statistically significant reduction in mental well-being (SF-36 MCS) was reported between 3 and 12-months, accompanied by a non-statistically significant improvement in physical well-being (PCS).[42] Small effect size statistics, comparable to those reported for the Hospital Anxiety and Depression Scale (HADS), were also reported for both component scores. Small effect size statistics were reported for the HUI3 up to 12-months post-arrest in a small prospective cohort.[55] However, robust evidence of measurement responsiveness following completion with OHCA survivors was not reviewed for any measure.

Interpretation (completion rates, data quality and meaning)

For OHCA survivors who agreed to complete the reviewed measures, acceptable completion rates have been reported for all generic measures completed following interview-administration between 6 and 12-months post-arrest (range 73% to 92%) (Table 2). Most non-responders were reportedly 'lost to follow-up'. However, evidence suggests that non-responders are more likely to be those with significant limitations, and hence scores may overestimate quality of life in survivors.[54]

Acceptable postal self-completion rates (72%) have also been reported for the SF-36(v1) at 36-months post-arrest.[36] A missing item rate of greater than 15% was reported – a rate comparable to other self-completed measures in this population.

Good data quality, with no evidence of end-effects, and low levels of missing data (3.4%) was reported for the SF-12 and SF-6D.[46] However, although missing data was low (total 1.0%; patients 0.5%; proxy 2.7%), large ceiling effects were reported for the EQ-5D-3L – with more than 46% of patients (and 23% of proxy) achieving a maximum score of 1.00 (perfect health) at 12-months. However, for this group, scores on the SF-6D and GOSE suggested substantial variability in health, with concerns related to mental health and vitality (issues not addressed in detail by the EQ-5D-3L). Data quality was not evidenced for the remaining measures.

No studies have attempted to define the SDD, MIC or MID following completion with survivors of cardiac arrest. However, Orbo et al,[42] calculated the Reliable Change Index (RCI) as a measure of the reliability of change scores on the SF-36 between 3 and 12-months post-OHCA. A reliable deterioration in mental well-being summary score (MCS) and improvement in physical well-being summary scores (PCS) was

reported for fifteen percent and twelve percent of survivors, respectively, between 3 and 12-months. However, for most survivors, at the domain level, change in mental or physical well-being was not statistically significant, but remained below that of an age- and gender-matched population.[42]

Feasibility

The feasibility of PROM completion has not been reported.

Where patients are very poorly or experience significant neurological deficit, proxy completion by a close relative or health professional is an option. However, there are few evaluations of proxy completion in this population. Acceptable SF-36 interview-administered, proxy-completion rates have been reported.[41] A stronger association between proxy completed EQ-5D-3L and the GOSE (0.67), than with survivor self-report (0.47), has been reported at 12-months post-arrest.[46] Missing data was greater following EQ-5D-3L proxy-completion (2.7% versus 0.5% in patients).[46] The associated cost, and concerns over loss to follow-up and associated reporting bias, have not been reported for PROM self- or proxy completion.

Acceptability and Patient Involvement

There is no evidence that OHCA survivors have been involved in the development or appraisal of PROM content for relevance or acceptability.

Discussion

High quality and relevant HRQoL assessment provides essential survivor-derived evidence of the often profound, long-term impact of cardiac arrest. However, the survivors' perspective is not widely reflected in current resuscitation research, and a measure specific to cardiac-arrest does not exist. Apart from one utility measure – the HUI3 – and a generic profile – the SF-36 – for which limited, but acceptable evidence of measurement validity was reviewed, robust evidence of the quality and acceptability of HRQoL measures in this population was largely limited or not available.

This is the first systematic review of patient-reported measures of HRQoL following completion by OHCA survivors with which to inform measurement selection. The review is strengthened by a transparent assessment of both study and measurement methodological quality according to consensus-derived standards,[7,13,14] and by reference to an established HRQoL framework with which to underpin a comparative evaluation of PROM content.[18,19] Although undertaken by one, experienced reviewer (KH), the study inclusion, data extraction and synthesis was discussed with an established working group of clinical academics, researchers and patients (as part of the Core Outcome Set for Cardiac Arrest (COSCA) initiative),[2,57] thus enhancing the transparency of the process and final recommendations. However, unlike other reviews of PROM quality,[11,12,16] the primary purpose of most reviewed studies was often not PROM evaluation. Therefore, studies were often judged to be of limited quality per the COSMIN criteria.

The review highlighted a lack of conceptual and empirical research regarding HRQoL assessment in this population – and evidence of measurement data quality, interpretability, reliability, construct validity and responsiveness was mostly unavailable. Moreover, PROM content validity and relevance to OHCA survivors was not specifically evaluated. Application of the HRQoL framework, suggests that the QOLIBRI – a traumatic brain-injury specific measure – was most reflective of the multi-dimensional nature of HRQoL, including many concerns identified as important by OHCA survivors.[57-59] The shorter utility measures – HUI3 and EQ-5D-3L – have a narrower focus and, as reported by others, may fail to include health concerns of relevance to specific patient populations.[60-62] A consequence of using measures with limited content validity is that the impact of OHCA and associated healthcare is sub-optimally assessed - clinical trials may overestimate good outcome and fail to identify important differences between groups on the outcomes that really matter to patients.

Except for the two generic index measures, all assess pain or discomfort. Fatigue, another important consideration for CA survivors, was assessed by six measures – the QOLIBRI, SF-36, SF-12, 15D and SF-

6D. Both the HUI3 and 15D include items about vision, hearing and speech; the 15D additionally includes items about eating and excretion. It is possible that items relating to vision and hearing are not relevant to CA survivors, and hence may result in an overestimation of utility scores and hinder the measures ability to detect meaningful change in this population.[62] However, speech, eating and excretion may be particularly relevant to survivors with more severe impairment and hence more likely to be completed by proxy respondents.

Just two measures (QOLIBRI, HUI3) assess cognition, but limited evidence suggests just a small association between the HUI3 and a cognition-specific measure.[55] In contrast, the QOLIBRI includes several items to explore a wide-range of issues associated with cognitive impairment, and a strong association with a cognition-specific measure has been reported.[32] Recognition of the importance of cognitive impairment in this population,[8,42] suggests a more detailed assessment of cognitive impairment than can be afforded by a single item is recommended.

Despite the small number of items, two of the utility measures - the EQ-5D-3L and SF-6D - cover a comparable range of HRQoL concepts. However, whilst evidence of good data quality and discriminative ability has been reported for the SF-6D at 12-months post-arrest, large ceiling effects and poor discriminatory ability have been reported for the EQ-5D-3L, suggesting that it underestimates the significant impact of cardiac arrest and may not detect important change.[46] A revised version – the EQ-5D-5L – has improved response options which may improve the data quality, but has yet to be evaluated in this population.

Proxy-completion – that is, by a significant other – has been evaluated following completion of the EQ-5D-3L[46] and the SF-36(v2)[41] in OHCA survivors. Whilst higher levels of missing data were reported following proxy completion of the EQ-5D-3L, there were significantly lower ceiling effects (23% versus 46%), possibly reflecting the inclusion of survivors with greater limitations.

HRQoL data are often missing or incomplete for patients with the poorest outcomes, which may result in systematic bias.[63] To enhance HRQoL data capture, standardised administration and routine screening for avoidable missing data is recommended,[64] which should be detailed in study protocols and standard operating procedures (SOPs).[63,64]

Although all measures were developed to be self-completed, all were interview-administered – in person, via the telephone or both – in the OHCA studies reviewed, with acceptable interview completion rates reported up to 12-months post-arrest. Proxy-completion, by appropriate assessors, is advised to

ensure that the perspective of survivors with the poorest outcomes are included in research and audit, and that HRQoL assessment does not underestimate the impact of OHCA survival. However, the logistical challenges of HRQoL assessment are not insignificant. Earlier guidance suggests that HRQoL should, as a minimum, be assessed up to 3-months post-arrest.[65] However, for many survivors, their HRQoL may continue to change[42] and a longer-term assessment of HRQoL is recommended.[39,57] Issues associated with the feasibility and acceptability of long-term PROM completion required urgent attention.

Most of the reviewed studies included HRQoL evaluations up to 12-months post-arrest – but no study commented on the ability of measures to detect real change. Demonstrating the ability of the measures to detect meaningful change – both within individual and between groups – over time is essential to enhancing confidence in data quality and interpretation.

Given the importance of HRQoL for future clinical trials, registries and cohort studies, and the paucity of conceptualisation and evidence of essential measurement and practical properties, further research is essential. Collaborative research with survivors of cardiac arrest and their partners or patient advocates is strongly recommended to improve the quality and acceptability of HRQoL assessment and to co-produce guidance informing the way in which HRQoL assessment can be applied in future research, registries and healthcare quality assessments.

Just a small number of mostly generic HRQoL measures have been evaluated with OHCA survivors. However, study methodological quality was poor and critical evidence of measurement properties and relevance to survivors was largely unavailable or at best limited. These significant limitations hinder clear assessment recommendations, whilst also limiting data interpretation when such measures are applied in research and healthcare quality assessment. A comparative evaluation of measurement content validity highlighted that few measures captured the multi-faceted nature of HRQoL and the outcomes that matter to OHCA survivors. Although providing a narrow, impairment-based assessment, limited evidence suggests that the HUI3 may be an acceptable, short, generic measure of HRQoL. Alternatively, the SF-36(v2) – for which evidence was also limited - may provide a more detailed assessment. Further comparative evaluations of widely-used generic measures – to include essential evidence of reliability, validity, responsiveness, interpretation and feasibility – is urgently required. Moreover, exploration of the relevance and acceptability of such measures with representative members of the OHCA survivors' community is urgently required to determine the need for a survivor-derived measure.

References:

1. Perkins GD, Jacobs IG, Nadkarni VM, et al; for the Utstein Collaborators. Cardiac arrest and cardiopulmonary resuscitation outcome reports: update of the Utstein resuscitation registry templates for out-of-hospital cardiac arrest. Resuscitation 2015;96:328–340

2. Whitehead L, Perkins GD, Clarey A and Haywood KL. A systematic review of the outcomes reported in cardiac arrest clinical trials: the need for a core outcome set. Resuscitation 2015;88:150–157.

3. Devlin NJ, Appleby J (2010) Getting the Most Out of PROMs: Putting health outcomes at the heart of NHS decision-making. The King's Fund. London.

4. Elliott VJ, Rodgers DL and Brett SJ. Systematic review of quality of life and other patient-centred outcomes after cardiac arrest survival. Resuscitation 2011;82:247-56.

5. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009 Oct;18(8):1115-1123.

6. de Vet H, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. A Practical Guide. Cambridge University Press, England. 2011.

7. Terwee CB, Prinsen CA, Ricci Garotti MG, et al. The quality of systematic reviews of health-related outcome measurement instruments. Qual Life Res 2016 Apr;25(4):767-79.

8. Moulaert VR, Verbunt JA, van Heugten CM, Wade DT. Cognitive impairments in survivors of out-of-hospital cardiac arrest: a systematic review. Resuscitation 2009 Mar;80(3):297-305

9. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A practical guide to their development and use. Oxford University Press. USA. 5th edition. 2014

10. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42.

11. Haywood KL, Mars TS, Potter R, Patel S, Matharu M, Underwood M. Assessing the impact of headaches and the outcomes of treatment: a systematic review of patient-reported outcome measures (PROMs). Cephalalgia. 2017. *In press.*

12. Conijn AP, Jens S, Terwee CB, Breek JC, Koelemay MJ. Assessing the quality of available patient reported outcome measures for intermittent claudication: a systematic review using the COSMIN checklist. Eur J Vasc Endovasc Surg 2015 Mar;49(3):316-34

13. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-45.

14. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012 May;21(4):651-7.

15. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998 2(14):i-iv.1-74

16. Haywood KL, Staniszewska S, Chapman S. Quality and acceptability of patient-reported outcome measures used in chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. Qual Life Res 2012 Feb;21(1):35-52

17. Staniszewska S, Haywood KL, Brett J, Tutton L. Patient and public involvement in patient-reported outcome measures: evolution not revolution. Patient 2012;5(2):79-87

18. Ferrans CE, Zerwic JJ, Wilbur JE, Larson JL. Conceptual model of health-related quality of life. J Nurs Scholarsh 2005;37(4):336-42

19. Bakas T, McLennon SM, Carpenter JS, et al. Systematic review of health-related quality of life models. Health Qual Life Outcomes 2012 Nov 16;10:134

20. Elbers RG, Rietberg MB, van Wegen EE, et al. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. Qual Life Res 2012 Aug;21(6):925-44.

21. Bullinger, M., & The TBI Consensus Group. Quality of life in patients with traumatic brain injury - basic issues, assessment and recommendations. Restorative Neurology and Neuroscience 2002 20:111-124.

22. Ware JE, Sherbourne CD. The MOS 36-item Short Form Health Survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30:473-83.

23. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. J Epidemiol Community Health. 1999 Jan; 53(1): 46–50.

24. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3): 220-33.

25. Fugl-Meyer AR1, Melin R, Fugl-Meyer KS. Life satisfaction in 18- to 64-year-old Swedes: in relation to gender, age, partner and immigrant status. J Rehabil Med 2002 Sep;34(5):239-46.

26. Flanagan JC: A research approach to improving our quality of life. American Psychologist 1978; 33: 138–147.

27. Burckhardt CS, Anderson KL, Archenholtz B, Hagg O. The Flanagan Quality of Life Scale: evidence of construct validity. Health Qual Life Outcomes 2003 1;59.

28. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med 2001 Jul;33(5):328-36.

29. The EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. Health Policy 1990;16: 199-208.

30. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care 2002 Feb;40(2):113-28.

31. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002 Mar;21(2):271-92.

32. Middelkamp W, Moulaert VR, Verbunt JA et al. Life after survival: long-term daily life functioning and quality of life of patients with hypoxic brain injury as a result of a cardiac arrest. Clin Rehabil 2007;21(5):425-31.

33. Mak M, Moulaert VR, Pijls RW and Verbunt JA. Measuring outcome after cardiac arrest: construct validity of Cerebral Performance Category. Resuscitation 2016; 100:6-10.

34. Graf J, Muhlohoff C, Doig GS et al. Healthcare costs, long-term survival, and quality of life following intensive care unit admission after cardiac arrest. Crit Care 2008;12(4):R92.

35. Bro-Jeppesen J, Kjaergaard J, Horsted TI, et al. The impact of therapeutic hypothermia on neurological function and quality of life after cardiac arrest. Resuscitation 2009; 80(2):171-176.

36. Moulaert VR, Wachelder EM, Verbunt JA, Wade DT, van Heugten CM. Determinants of quality of life in survivors of cardiac arrest. J Rehabil Med 2010; 42(6):553-8.

37. Reinhard V, Parna K, Lang K, Pisarev H, Sipria A, Starkopf J. Long-term outcome of bystander-witnessed out-of-hospital cardiac arrest in Estonia from 1999 to 2002. Resuscitation 2009; 80(1):73-8.

38. Kowalik R, Szczerba E, Koltowski L, et al. Cardiac arrest survivors treated with or without mild therapeutic hypothermia: performance status and quality of life assessment. Scand J Trauma Resusc Emerg Med 2014 Dec 12; 22: 76.

39. Moulaert VR, van Heugten CM, Winkens B, et al. Early neurologically-focused follow-up after cardiac arrest improves quality of life at one year: A randomised controlled trial. Int J Cardiol 2015; 193:8-16.

40. Boyce-van der Wal LW, Volker WG, Vliet Vlieland TP, van den Heuvel DM, van Exel HJ, Goossens PH. Cognitive problems in patients in a cardiac rehabilitation program after an OHCA. Resuscitation 2015 Aug;93:63-8.

41. Cronberg T, Lilja G, Horn J, et al. for the TTM Trial Investigators. Neurologic function and health-related quality of life in patients following targeted temperature management at 33°C vs 36°C after out-of-hospital cardiac arrest: a randomized clinical trial. JAMA Neurol 2015;72:634–41.

42. Orbo M, Aslaksen PM, Larsby K, Schafer C, Tande PM, Anke A. Alterations in cognitive outcome between 3 and 12 months in survivors of out-of-hospital cardiac arrest. Resuscitation 2016;105: 92-99.

43. Deasey C, Bray J, Smith K et al. Functional outcomes and quality of life of young adults who survive out-of-hospital cardiac arrest. Emerg Med J 2013;30(7):532-7.

44. Beesems SG, Wittebrood KM, de Haan RJ, Koster RW. Cognitive function and quality of life after successful resuscitation from cardiac arrest. Resuscitation 2014;85:1269-74.

45. Smith K, Andrew E, Lijovic M, Nehme Z, Bernard S. Quality of life and functional outcomes 12 months after out-of-hospital cardiac arrest. Circulation 2015;131:174-81.

46. Andrew E, Nehme Z, Bernard S, Smith K. Comparison of health-related quality of life and functional recovery measurement tools in out-of-hospital cardiac arrest survivors. Resuscitation 2016;107:57-64.

47. Wallin E, Larsson IM, Rubertsson S, Kristofferzon ML. Cardiac arrest and hypothermia treatment--function and life satisfaction among survivors in the first 6 months. Resuscitation 2014;85(4):538-43.

48. Wilson M, Staniforth A, Till R, das Nair R, Vesey P. The psychosocial outcomes of anoxic brain injury following cardiac arrest. Resuscitation 2014 Jun;85(6):795-800.

49. Tiainen M, Poutiainen E, Oksanen T et al. Functional outcome, cognition and quality of life after out-of-hospital cardiac arrest and therapeutic hypothermia: data from a randomized controlled trial. Scand J Trauma Resusc Emerg Med 2015 Feb 6;23:12.

50. Nichol G, Stiell IG, Hebert P et al. What is the quality of life for survivors of cardiac arrest? A prospective study. Academic Emergency Medicine 1999;6(2):95-102.

51. Stiell I, Nichol G, Wells G, De Maio V, Nesbitt L, Blackburn J, Spaite D; for the OPALS Study Group. Health-related quality of life is better for cardiac arrest survivors who received citizen cardiopulmonary resuscitation. Circulation 2003;108:1939–44.

52. Raina KD, Callaway C, Rittenberger JC, Holm MB. Neurological and functional status following cardiac arrest: method and tool utility. Resuscitation 2008;79:249-56.

53. Stiell IG, Nesbitt LP, Nichol G, Maloney J, Dreyer J, Beaudoin T, Blackburn J and Wells GA; the OPALS Study Group. Comparison of the Cerebral Performance Category score and the Health Utilities Index for survivors of cardiac arrest. Ann Emerg Med 2009;53:241–248.

54. Longstreth WT Jr, Nichol G, Van Ottingham L and Hallstrom AP. Two simple questions to assess neurologic outcomes at 3 months after out-of-hospital cardiac arrest: experience from the public access defibrillation trial. Resuscitation 2010;81:530–3.

55. Raina KD, Rittenberger JC, Holm MB, Callaway CW. Functional Outcomes: One Year after a Cardiac Arrest. Biomed Res Int 2015;2015:283608.

56. Nichol G, Guffey D, Stiell IG, et al; the Resuscitation Outcomes Consortium Investigators. Post-discharge outcomes after resuscitation from out-of-hospital cardiac arrest: a ROC PRIMED substudy. Resuscitation 2015;93:74–81.

57. Haywood KL, Whitehead L, Nadkarni V et al. Core Outcome Set for Cardiac Arrest (COSCA) in adults: An Advisory Statement From the International Liaison Committee on Resuscitation. Circulation 2017: *in press*

58. Whitehead LL (2017) Identifying a Core Outcome Set for Cardiac Arrest Effectiveness Trials. Doctoral Thesis. University of Warwick.

59. Haydon G, van der Riet P, Inder K. A systematic review and meta-synthesis of the qualitative literature exploring the experiences and quality of life of survivors of cardiac arrest. Eur J Cardiovasc Nurs 2017 Aug;16(6):475-483.

60. Lim WC, Black N, Lamping D, Rowan K, Mays N. Conceptualizing and measuring health-related quality of life in critical care. J Crit Care. 2016 Feb;31(1):183-93.

61. Pietersma S, de Vries M, van den Akker-van Marle ME. Domains of quality of life: results of a three-stage Delphi consensus procedure among patients, family of patients, clinicians, scientists and the general public. Qual Life Res 2014 Jun;23(5):1543-56

62. Kuspinar A, Mayo NE. A review of the psychometric properties of generic utility measures in multiple sclerosis. Pharmacoeconomics. 2014 Aug;32(8):759-73.

63. Fairclough DL, Peterson HF and Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? Stat Med 1998;17:667-77

64. Kyte D, Ives J, Draper H, Calvert M. Current practices in patient-reported outcome (PRO) data collection in clinical trials: a cross-sectional survey of UK trial staff and management. BMJ Open. 2016 Oct 3;6(19):e012281.

65. Becker LB, Aufderheide TP, Geocadin RG, et al; American Heart Association Emergency Cardiovascular Care Committee; Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation. Primary outcomes for resuscitation science studies: a consensus statement from the American Heart Association. Circulation. 2011 Nov 8;124(19):2158-77

Tables:

Table 1 Assessment criteria for the quality of reported measurement properties,[10,12,13] evidence of practical properties,[15] and patient involvement in PROM development/ evaluation.[11,17]

Table 2 Summary and item content of the reviewed HRQoL measures (n=10).

Table 3: Methodological quality (COSMIN) of each study (n=25) per HRQoL measure (n=10) and quality of investigated measurement properties.

Table 4: Data synthesis, levels of evidence and overall quality of reviewed measures of HRQoL (n=10)[a]

Figure 1: Review of measures for Cardiac Arrest Clinical Trials - PRISMA flow-chart for article inclusion

Appendices:

Appendix 1: Search Strategy

Appendix 2: Details of the included studies (n=25)

Appendix 3: Data extraction - measurement and practical properties for HRQOL measures (n=10) evaluated following completion by OHCA survivors.

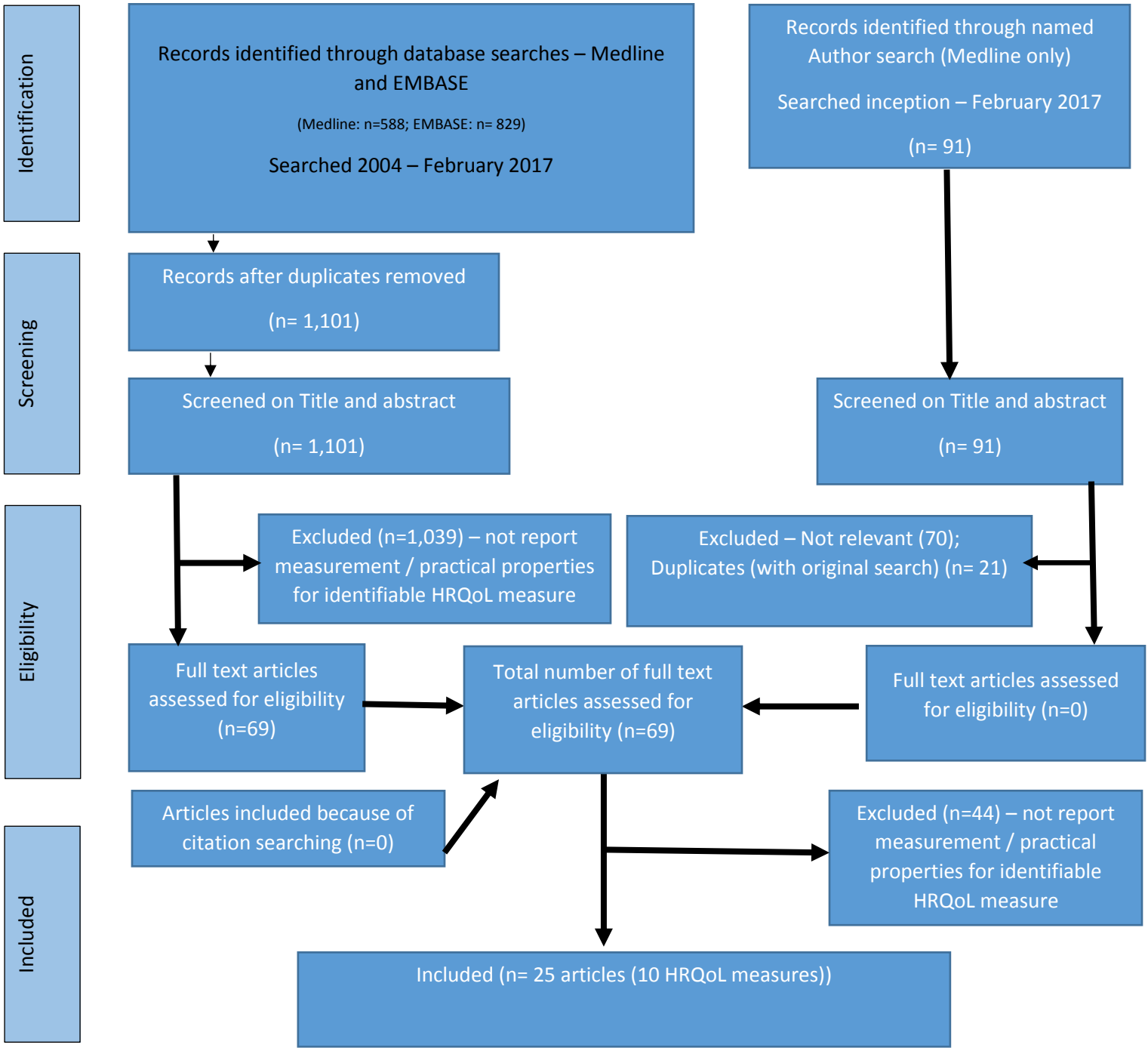
Funding source:

This research was undertaken as part of the COSCA initiative (developing a Core Outcome Set for Cardiac Arrest), but did not receive specific funding.

Authors contribution:

KH and GP conceived and designed the study; KH organized the conduct of the study; KH carried out the study (including acquisition of study data); KH and NP designed the search strategy and completed all searches; KH extracted all data; KH, GP, LM, MC, GL and NP contributed to the analysis and interpretation of study data. KH drafted the output; all authors provided a critique of the output for important intellectual content.

Figure 1: Review of measures for Cardiac Arrest Clinical Trials - PRISMA flow-chart for article inclusion



OHCA – PROM Review - Resuscitation May 2017

Table 1: Assessment criteria for the quality of reported measurement properties,[10,12,13] evidence of practical properties[15] and patient involvement in PROM development/ evaluation.[11,17]

Measurement properties [10,12,13]	**Rating**	**Assessment of quality**
Validity		
Content validity - *the extent to which the item content of a measure is an adequate reflection of the construct being measured*	+	Authors provide a clear description of the measurement aim, target population, concept(s) measured and process of item selection. Members of the target population and experts in the field were clearly identified as being involved in development. For measures applied for the first time in a new population, evidence that the views of members of the target population (and experts in the field) have been sought to determine relevance, comprehension and comprehensiveness.
	?	Insufficient evidence available
	-	No detail re measurement aim, target population, concept(s) measured, process of item selection; members of the target population or experts were not specifically involved in development. For measures applied for the first time in a new population, evidence whereby the relevance and acceptability of the measure with members of the target audience or experts was not provided.
Construct validity - Structural validity - *the extent to which PROM scores adequately reflect the dimensionality of the construct being measured.*	+	Factors should explain 50% of the variance
	?	Explained variance not reported
	-	Factors explain < 50% of the variance
Construct validity - Hypothesis testing - *convergent (the extent to which measures of related constructs are related to each other)*	+	Correlations with measures of the same construct should be >0.50 OR at least 75% of the results in accordance with hypothesized associations AND correlations with related constructs should be higher than with those reported with unrelated constructs
	?	Only report correlations with unrelated constructs OR the extent to which between group differences are expected is not described / justified
	-	Correlations with measures of the same construct are <0.50 OR < 75% of the results in accordance with hypothesized associations OR correlations with related constructs are lower than those reported with unrelated constructs
Construct validity - Hypothesis testing: Known-groups validity (not included in COSMIN checklist)	+	Hypothesized between group differences are supported (or can be assumed) AND between group differences are statistically significant
	?	Between group differences are poorly hypothesized, but between group differences are statistically significant
	-	Expected between group difference poorly defined or justified AND the statistical significance of between group differences not reported.

- *discriminant (the extent to which a measure can demonstrate differences between groups known to differ on important variables)*		
Reliability		
Internal consistency	+	Cronbach's alpha(s) > 0.70
- *the extent to which items within a measure are internally consistent*	?	Cronbach's alpha not evaluated or dimensionality unknown
	-	Cronbach's alpha(s) < 0.70
Reliability *(test-retest / inter-rater / inter-rater)*	+	Intra-class Correlation Coefficient (ICC)/weighted Kappa >0.70 OR Pearson's r >0.80
- *the extent to which a measure provides the same results on repeated completions, assuming no change in the underlying health state*	?	Neither ICC/weighted Kappa, not Pearson's r evaluated
	-	ICC/weighted Kappa <0.70 OR Pearson's r <0.80
Reliability – *measurement error*	N/A	Descriptive (not rated)
- The systematic and random error of a score that cannot be attributed to the true change in the construct being measured		
Responsiveness	+	Change-score correlations with measures of the same construct are >0.50 OR at least 75% of the results are in accordance with hypothesized associations OR the Area Under the Curve (AUC) is >0.70 AND change-score correlations with measures of related constructs are higher than those reported with unrelated constructs
- *the ability to detect important change over time in the construct being measured (criterion / construct-based assessment)*	?	Solely correlations with unrelated constructs
	-	Change-score correlations with measure of the same construct <0.50 OR < 75% of the results are in accordance with hypothesized associations OR AUC is <0.70 AND change-score correlations with related constructs are lower than those reported with unrelated constructs

Interpretability - *the degree to which qualitative meaning can be assigned to PROM scores or change in scores - that is, clinical or commonly understood implications can be linked with the score / change* - *includes consideration of completion rates and data quality (end effects; change in scores)*	N/A	Descriptive (not rated) - requires evidence that the minimal important (within-person) change (MIC) and/or minimal importance (between group) difference (MID) exceeds evidence of the smallest detectable difference (SDD). Supported by evidence of acceptable data quality (that is, score distribution, absence of end effects (Floor/ Ceiling)
Practical properties [15]		
Acceptability - is the PROM acceptable to patients?	N/A	Descriptive (not rated) – evidence of respondent burden and relevance
Feasibility - Is the PROM easy to administer and process?	N/A	Descriptive (not rated) – time and cost to administer and score
Patient Engagement [11,17]		
Patient and Public Involvement / Engagement in PROM development / evaluation[11,17] - *the extent of active involvement / collaboration in PROM development / evaluation assessed at 3-levels*	++	User-led – broadly interpreted, patients control, direct and manage the PROM development/ evaluation
	+	Collaboration – involves active engagement. Ongoing partnership between researchers and patients in PROM development/ evaluation. Patients may be members of a research team or advisory group and collaborate on design, development and/or dissemination.
	-	Consultation – patients are consulted for their views, for example, through focus group, but these views are not necessarily adopted.

Table 2 Summary and item content of the reviewed HRQoL measures (n=10)

PROM ***Developer*** ***Web-link*** ***Cost (license)*** ***Time to complete***	**Origin**	**Conceptual focus** **Response options; Recall period** **Completion format** **Language versions**	**Domains of HRQoL [Ferrans et al, 2005] [18,19][a]** ***(items per domain)***					**How to score**
				Functional Status				
			Symptoms	***Physical***	***Cognition***	***Psychological***	***Social / Role***	***General Health perception***
Condition-specific (n1)								
Quality of Life after Brain Injury **(QOLIBRI)** [Bullinger et al, 2002][21] http://www.qolibrinet.com/index.htm ***License:*** No license fee – free to use for researchers, clinicians and non-profit organizations. Developers request that users register use of the QOLIBRI on the website ***Completion time:*** *Self-complete approx. 7-10 mins* ***User guide:*** see weblink	USA	HRQOL for individuals after traumatic brain injury Designed to capture changes in HRQOL in areas commonly affected by brain injury. Development guided by a HRQoL assessment model: a person's perspective on his or her subjective health condition, functioning and wellbeing in the domains of physical, psychological (emotional and cognitive), social and daily life. 37 items across 6-domains of health – domains explore: a.'Satisfaction with key aspects of life' (4 domains): Cognition (7 items); Self (7 items); Daily life and autonomy (7); Social relationships (6 items); b.Feeling bothered with key aspects of life (2 domains); Emotions (5 items) and Physical problems (5 items). ***Response options*** Satisfaction items on a 1 to 5 scale – where 1 = npt at all satisfied ad 5 = very satisfied. Feeling bothered items on a 1 to 5 scale – where 1 = very bothered and 5 = not at all bothered. ***Recall period***: The past week	Pain – *included in 'physical problems'* Energy – *included in 'view of self'*	Physical problems (5) – includes movement problems, pain, vision/hearing,	Cognition (7) – concentrate, express self, remember, problem solve, decisions, navigate, thinking	View of self (7) – energy, motivation, self-esteem, looks, achievements, self-perception, future Emotions (5) – anxiety, depression, ager/aggression, loneliness, boredom.	Daily life & autonomy (7) – independence, getting out, domestic, finances, work, social, feeling in charge. Social relationships (6) – family, friends, partner, sex life, affection, attitudes of others	Satisfaction and view of oneself included in 'emotions and view of self'	Item responses are summed and divided by the number of responses to give a mean value. (Score cannot be calculated if > one third of responses (per domain or in total) are missing). 6 sub-scales can be used separately or combined to provide a HRQOL profile; or summed to produce an Index score. Scores converted to a 0-100 scale, where 0= worst possible HRQOL and 100 = best. http://www.qolibrinet.com/scoring.htm

		Completion: Self- or interview-administered. ***Language:*** > 10 language versions: http://www.qolibrinet.com/registration.htm							
Generic – profile measures (n3)									
Short Form 36-item Health Survey – version 1 and version 2 **(SF-36 v1 / SF-36 v2)** [Ware & Sherbourne 1992[22]; Jenkinson et al, 1999][23] https://campaign.optum.com/content/optum/en/optum-outcomes/what-we-do/health-surveys.html ***License*** *for use per project; minimum fee $USA* *Survey license request:* https://www.optum.com/campaign/ls/outcomes-survey-request.html ***Completion time:*** *Range 5 to 30 minutes (not reported in Cardiac Arrest population)* ***http://www.qolibrinet.com/scoring.htm***	USA	Functional health and well-being from the patient's perspective – underpinned by 8 health domains across both physical (4) and mental (4) aspects of health Total 35 items plus one health transition item ***Response options:*** Between 3 and 6-level categorical response options per item ***Revision of the SF-36v1 to v2:*** 5-level response options replaced dichotomous options for 7 items in the role function items. Other modifications improved content and layout.[23] ***Recall period:*** Standard recall 4-weeks; Acute recall 1-week ***Completion:*** Self, Interview (in person; telephone) or proxy supported ***Language:*** > 170 language versions: https://campaign.optum.com/optum-outcomes/what-we-do/health-survey-translation/surveys-translation-tables.html The IQOLA project supported the development of conceptually equivalent and culturally appropriate translations (see http://www.iqola.org/	Bodily Pain (BP) (2) Vitality (VT) - fatigue / tiredness (2)	Physical functioning (PF) (10) Role limitation (RP) (4)	-	Mental health (MH) (5); Role limitation (RE) (3)	Social functioning (SF) (2)	General health (GH) (5) - perceived well-being	2-ways of presenting the data: 2.1 8-domain profile 2.2 Two component summary scales: Physical Component Summary (PCS); Mental Component Summary (MCS Scoring requires SF-36 specific algorithm. Norm-based scoring: score transformed to 0-100 (mean 50 (SD 10)) Population-based norms available

		NOTE: Utility values A preference-based utility index – the SF-6D – can be calculated following completion of the SF-36 to inform economic analyses https://www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d							
Short Form 12-item Health Survey – version 2 **(SF-12 v2)** [Ware et al, 1996][24] https://campaign.optum.com/content/optum/en/optum-outcomes/what-we-do/health-surveys.html ***License*** *for use per project; minimum fee $USA* *Survey license request:* https://www.optum.com/campaign/ls/outcomes-survey-request.html	USA	Functional health and well-being from the patient's perspective – underpinned by 8 health domains across both physical (4) and mental (4) aspects of health Total 11 items plus one health transition item ***Response options:*** Between 3 and 6-level categorical response options per item ***Recall period:*** Standard recall 4-weeks; Acute recall 1-week ***Completion:*** Self, Interview (in person; telephone) or proxy supported ***Language:*** > 170 language versions: https://campaign.optum.com/optum-outcomes/what-we-do/health-survey-translation/surveys-translation-tables.html The IQOLA project supported the development of conceptually equivalent and culturally appropriate translations (see http://www.iqola.org/ ***NOTE:*** Utility values A preference-based utility index – the SF-6D – can be calculated following completion of the SF-36 to inform economic analyses	Bodily Pain (BP) (1) Vitality (VT) - fatigue / tiredness (1)	Physical functioning (PF) (2) Role limitation (RP) (2)	-	Mental health (MH) (2); Role limitation (RE) (2)	Social Functioning (SF) (1)	General health (GH) (5) - perceived well-being (1)	Scores presented as two component summary scales only: Physical Component Summary (PCS); Mental Component Summary (MCS Scoring requires SF-12 specific algorithm. Norm-based scoring: score transformed to 0-100 (mean 50 (SD 10)) Population-based norms available Summary scores: Physical (PCS), Mental (MCS) (mean 50, sd 10)

		https://www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d							
Generic - Index (n2/10)									
Life Satisfaction Checklist **LiSat-11** **[Fugl-Meyer et al, 2002][25]** ***License:*** Not clear ***Completion time:*** approx. 5 minutes (not reported in CA population) ***Users-guide****:* - ***No active web-site identified***	Sweden	Generic self-report checklist of Life Satisfaction (aspiration – achievements gap) across 10 domains of life plus life as a whole). Developed in Swedish population (aged 18-64yrs). LiSat-11 is an extension of the LiSat-9: addition of i) somatic and ii) psychological health. Factor analysis of original measure described a 4-factor solution ***Response options:*** 6-grade ordinal scale from 1 (very dissatisfied) to 6 (very satisfied) ***Recall period:*** Not clear ***Completion:*** Interview or self-administered. For use with adults (aged 18 years and older). ***Formats:*** Unclear ***Language:*** Unclear	-	(ADL) Ability to manage my self-care (1) (Somatic) Physical health (1)	-	Psychological health (1)	My vocational situation (1) My financial situation (1) My leisure situation (1) My contact with friends and acquaintances (1) My sex life (1) My family life (1) My partner relationship (1)	My life as a whole (1)	Developers describe 4 factors: 1) Closeness (3 items) 2) Health (3 items) 3) Spare time (2 items) 4) Provision (2 items) But – alpha levels were low; therefore, index score supported. Item summation. Index score reported.
Quality of Life Scale **QOLS** **[Flanagan, 1978[26]; Burckhardt & Anderson, 2003[27]]** ***License:*** QOLS is copyrighted by Carol Burckhardt. Contact burckhac@ohsu.edu for	USA	Generic self-report measure of quality of life for use across patient groups and culture – 16-items across 6 conceptual domains: material and physical well-being; relationships with other people; social, community and civic activities; personal development and fulfillment; recreation; independence (1 item). ***Response options***: 7-point 'delighted (7)-terrible (1)' scale OR	-	Health – being physically fit and vigorous (1) Participating in active recreation (1)	-	Understanding yourself (1)	Material comforts (1) Relationships (1) Having and rearing children (1) Close relationships (1) Close friends (1)	Independence – doing for yourself (1)	Item summation to give a total index score: range from 16 to 112; where higher scores indicate better QoL. Average total score for healthy population is approx. 90 [see Burckhardt & Anderson, 2003]

a free copy of the English language version. ***Completion time:*** approx. 5 minutes (not reported in CA population) ***Users-guide****:* see Burckhardt & Anderson, 2003 ***No active web-site identified***		7-point satisfaction scale (anchored 'very satisfied (7)' to 'very dissatisfied (1)' ***Recall period:*** At this time ***Completion:*** Interview or self-administered. For use with adults (aged 18 years and older). Requires approximately 5 minutes. ***Formats:*** Unclear ***Language:*** At least 16 – include English, Swedish, Norwegian, and Hebrew translations.					Helping others (1) Participation-organizations (1) Learning (1) Work – job or home (1) Expressing yourself – creatively (1) Socializing (1) Reading, music etc (1)		
Generic – preference-based utility measures (n4)									
The 15-dimension Health-Related Quality of Life Instrument **15D** [Sintonen, 2001][28] http://www.15d-instrument.net/15d/ ***License:*** ***Completion time:*** less than 10-minutes (but not reported in CA population) ***User guide: see web-site*** *Originally developed in 1981; revised in 1986 and 1993*	Finland	*Conceptual underpinning not clearly reported.* Descriptive system - 'health-related quality of life of adults' described across 15 dimensions. ***Response options:*** 5-ordinal level response options (more/less of an attribute) ***Recall period:*** present health status. ***Completion:*** Primarily self-administered. For use with adults (aged 16years and older). Interview and proxy also supported. *** Version for Adolescents – 16D (for children aged 12-15yrs) and for younger children – 17D (aged 8-11yrs) also available.* ***Formats:*** ***Language:*** Original Finnish for Finland; now >30 language versions.	Discomfort and symptoms (disco) (1) Vitality (vital) (1)	Mobility (move)(1) Sleeping (sleep)(1) Body function – Breathing (breath)(1) Eating (eat)(1) Elimination/ excretion (excret)(1) Senses – Vision (see)(1) Hearing (hear)(1) Speech (speech)(1) (8)	-	Mental function *(cognition and memory)(1)* Depression (dep)(1) Distress (distr)(1)	Usual activities (uact) (1) Sexual activity (sex)(1)	-	Utility index score (multiattribute utility theory)): algorithm to score: http://www.15d-instrument.net/valuation-system/ 15D single index on a scale 0.11 to 1.00; where 0 is dead and 1.00 is perfect health *(0.0162 is unconscious or comatose)*

		http://www.15d-instrument.net/15d/languages/							
EuroQoL EQ-5D-3L (EuroQoL Group, 1990)[29] http://www.euroqol.org/home.html ***License*** *for use per project; Free. But use must be registered on EuroQol website* http://www.euroqol.org/register-to-use-eq-5d.html ***Completion time:*** *Less than 5 minutes (not reported in Cardiac Arrest population)* ***User guide:*** *free at the following link -* http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf	Multiple	Standardized, preference-based measure of health status for use in clinical and economic appraisal *d* EQ-5D descriptive system: 5 items across '5 domains' *(2/5 reflect physical functional status)* *(EQ VAS – self-rated health on a 20cm vertical visual analogue scale (VAS))* ***Response options:*** 3-level categorical response options per item (no problems (1) to extreme problems (3) *Completion of all items will produce a 5-digit number describing the respondent's health state (but the numerals 1-5 have no inherent arithmetic properties and should not be used as a cardinal score)* ***Recall period:*** Today ***Completion:*** Self, Interview (in person; telephone) or proxy (two proxy versions) supported http://www.euroqol.org/about-eq-5d/modes-of-administration.html ***Formats:*** PDA; pen and paper; proxy paper; tablet; telephone; web. http://www.euroqol.org/eq-5d-products/eq-5d-5l.html ***Language:*** > 120 language versions (see www.euroqol.org/)	Pain / discomfort (1)	Mobility Self-care (2)	-	Anxiety / depression (1)	Usual activities *(including work, study, housework, family or leisure activities)* (1)	-	2-ways of presenting the data: 1. EQ-5D-5L Index value EuroQoL-specific coding algorithms to support calculation of Utility Score (Index): Crosswalk values sets from EQ-5D-3L support calculation of EQ-5D-5L utility score. Index range -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 Country-specific value sets and population-based norms available Report both measure of central tendency and a measure of dispersion: eg, mean and SD; median and percentiles. 2. EQ-5D-5L descriptive system as a health profile - reflects individual item scores. 2.1 Report as the frequency or proportion of reported problems for each level for each dimension 2.2 Dichotomise into 'No problems' (1) and 'Problems' (2-5) – report frequencies of reported problems.

Health Utility Index – 3 **(HUI-3)** [Feeney et al, 2002][30] http://www.healthutilities.com/ ***License*** *for use per project; minimum fee $USA 3k [Horsman, 2003]* ***Completion time:*** *Approx. 8mins self-completion* *Approx 3mins interview completion* *(not reported in Cardiac Arrest population)* ***User guide:*** *available once HUI3 is purchased*	Canada	*Preference-based, comprehensive system for measuring health status and HRQoL and for producing utility scores. Applicable for all persons aged 5years and older.* *HUI3 classification system: describes the comprehensive health state of an individual across 8 attributes of general health (6/8 items reflect physical functional status)* ***Response options:*** *Between 4 and 6 descriptive response options (ability/disability)* ***Recall periodl:*** *'Current' or 'Usual' – 'Usual' recommended for clinical studies. Choice of 1-week, 2-week or 4-week recall available.* ***Completion:*** *Self, Interview (in person; telephone) or proxy (proxy version available) supported* ***Language:*** *16 versions – including English, Chinese, Dutch, French, German, Italian, Japanese, Portuguese, Russian, Spanish, Swedish*	Pain - severity (1)	Ambulation – *ability to walk (distances)* Dexterity – ability to use hands and fingers Senses – Vision Senses - Hearing Speech – *ability to be understood* (5)	Cognition – *ability to solve day to day problems* (1)	Emotion – *happiness and interest in life* (1)	-	-	2-ways of presenting the data: 1.HUI3 utility index: scored using single and multi-attribute utility functions. HUI-specific coding algorithms to support calculation of single-attribute Utility Score (Index); Index range -0.36 to 1.00; Where 1.00 is perfect health, 0 is dead, <0 is a health state worse than death Population-based norms available 2. Multi-attribute descriptive system – 'Classification system' - reflects individual item scores.
Short Form 6-dimensions **SF-6D** *NOTE: Scoring of the SF-6D requires completion of the SF-36or SF-12* [Brazier et al, 2002][31] https://www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d ***License*** *– as per SF-36/12*	UK	Generic, preference-based single index measure. Produces 6-dimension scores (6D), a 6-digit health state and a utility value. If from SF-36: 11 (of 36) items;	Bodily Pain (BP) (1) Vitality (VT) - fatigue / tiredness (1)	Physical functioning (PF) (1) Role limitation (RP) (1)	-	Mental health (1)	Social functioning (1)	-	0.46 to 1.00; where 0 is dead and 1.00 is perfect health

Completion time – as per SF-36/12									

Footnote:

a Item content distribution according to the Ferrans et al revision to the Wilson and Cleary HRQOL model[18,19]

Table 3: Methodological quality (COSMIN[a]) of each study (n=25) per HRQoL measure (n=10) and quality of investigated measurement properties.[b]

Measure Study (n)	Country (language)	Subjects (n)	Reliability		Validity			Responsiveness	Interpretability - Include score distribution – item/scale level
			Internal consistency	Temporal stability	Content	Construct			
						Convergent/ divergent	Known-groups		
Health-related Quality of Life (n=10)									
Condition-specific (n=1/10)									
QOLIBRI (2 studies)									
Middlekamp et al (2007)[32]	NL Dutch	20				+ Poor			
Mak et al (2016)[33]	NL Dutch	59				+ Good			
Generic Profile (n=3/10)									
SF-36v1 (8 studies)									
Graf et al (2008)[34]	Germany German	81	+ Poor				+ Poor		
Bro-Jeppesen et al (2009)[35]	Denmark Danish	156	+ Poor				+ Poor		
Mouleart et al (2010)[36]	NL Dutch	63				PCS + MCS + Fair	PCS + MCS + Fair		2/63 patients had >15% missing items for SF-36 (same as FAI, CIQ, CFQ, FSS)
Reinhard et al (2013)[37]	Estonia	44					+ Poor		
Kowalik et al (2014)[38]	Poland Polish	65				+ Poor	+ Poor		
Moulaert et al (2015)[39]	NL Dutch	143					+ Fair		
Boyce-van der Wal et al (2015)[40]	NL Dutch	77				+ Fair	+ Fair		
SF-36v2 (2 studies)									
Cronberg et al (2015)[41]	Sweden Swedish	946					? Fair		In-person (or telephone) interview (n455)(proxy 8%) completion equivalent to 'Two Simple Questions' rate – and higher than several other measures (MMSE, IQCODE)
Orbo et al (2016)[42]	Norway Norwegian	33				PCS + MCS + Fair		ES: MCS -0.38; PCS 0.21	*Reliable Change Index:* Deterioration: MCS 5 (15%) of patients; PCS 0. Improvement: MCS 0; PCS 4 (12%). No change: MCS 28 (86%); PCS 29 (88%)
SF-12v2 (4 studies)									
Deasy et al (2012)[43]	Australia English	56					+ Fair		Item/scale level data – no end effects 'Good' telephone interview completion rates; but 'challenges with loss to follow-up'
Beesems et al (2014)[44]	NL	220					+		

	Danish						Fair		
Smith et al (2015)[45]	Australia English	687					+ Fair		'Good' telephone completion rates
Andrews et al (2016)[46]	Australia English	1188				PCS + MCS + Fair	PCS + MCS + Fair		79% of those known to be alive at 12/12) responded to interviews Missing items at scale level: 3.4% (34/1188) No end effects
Generic Index (n=2/10))									
LiSat-11 (1 study)									
Wallin et al (2014)[47]	Sweden Swedish	45	+ Poor				- Poor		
QOLS (1 study)									
Wilson et al (2014)[48]	English UK	56					- Poor		
Preference-based utility (n=4/10)									
15D (1 study)									
Tainen et al (2015)[49]	Finland Finish	49				? Poor	+ Fair		Interviews 1yr: 85.6% response rate
EQ-5D-3L (3 studies)									
Deasy et al (2012)[43]	Australia English	56							Item level: At 1-year: >70% report 'No problems' with Mobility, Self-care, Pain
Smith et al (2015)[45]	Australia English	687							Item level: At 1-year: >30% No problems in 5/5 domains: Self-care (87.6%); Pain (71.7%); Usual activities (67.8%); Mobility (66.4%); Anxiety (66.2%)
Andrews et al (2016)[46]	Australia English	1188				+/- Fair	+ Fair		Large ceiling effects and poor discrimination at higher levels of QoL/function 75th percentile is equal to a score of 1.00 for all GOSE categories above lower moderate disability
HUI3 (7 studies)									
Nichol et al (1999)[50]	USA US-English	86				? Fair	+ Good		
Stiell et al (2003)[51]	USA US-English	268							Score range -0.30 to 1.0 >0.80 134/316 (42%); <0.80 134/316 (42%); 0.60-0.80 56/316 (18%) <0.60 78/316 (25%) (Unknown n=48 (Known survivors 316)) *(HUI3 >25% have scores <0.60 at 12/12 (severe impairment))* (CPC at 12/12: Good 88%; Mod 9%; Severe 6%)
Raina et al (2008)[52]	USA	21		+		+/-			

	US-English			Poor		Poor			
Steill et al (2009)[53]	USA US-English	305				+ Fair	+ Fair		HUI3 median 0.84 (IQR 0.61 to 0.97) *Approximate range -0.25 to 1.00. Skewed distribution - End effects? – unclear how many scored 1.00*
Longstretch et al (2010)[54]	USA US-English	32					+ Fair		At 3/12: authors report that missing most often in those not reporting full recovery (so scores may overestimate QoL)
Raina et al (2015)[55]	USA US-English	29				+/- Poor		ES 0.43 (1-6mth); 0.19 (6-12mth); 0.26 (1-12mth)	Mean scores at HD,1/6/12mths: HUI3 severe disability all time-points (all <0.68); vs. GOS-E (low mod disability) vs. CPC and mRS suggest 'good outcome'
Nichol et al (2015)[56]	USA US-English	644				? Fair	+ Fair		Mean HUI3 (*scores <0.70 = severe impairment*): 3/12 0.75 (0.33); 6/12 0.74 (0.35); Worst 0.71 (0.35); Best 0.77 (0.33) *Range and end effects NR. ?Skewed distribution towards better health.* Predictive model: Suggests that for every unit change in mRS, this predicts a change of -0.06 in the HUI3
SF-6D (1)									
Andrews et al (2016)[46]	Australia English	1188				+ Fair	+ Fair		No end effects*: strong evidence of variability in SF-6D score across GOSE categories (in patients reporting full health on EQ-5D-3L).* *SF-6D scores improved as GOSE scores improve*

FOOTNOTE:

[a] COSMIN – Consensus on Standards for Measurement Instruments. Four-grade rating for study methodological quality: Excellent, Good, Fair Poor.[13,14]

[b] Quality of Measurement property: Evidence is graded as: adequate (+) [reaches accepted standards]; conflicting (+/-); inadequate (-) [does not research accepted standards]; or indeterminate (?) [the results are difficult to define]. (Adapted from 10,12)(Table 1 for detail).

Table 4: Data synthesis, levels of evidence and overall quality of reviewed measures of HRQoL (n=10)[a]

PROM[b]	Number of evaluations	Reliability			Validity		Construct Validity		Responsiveness	Interpretation
		Internal consistency	Temporal stability	Measurement error	Content validity	Structural validity	Hypothesis testing	Known-groups	Responsiveness	
Condition-specific (1)										
QOLIBRI	2						+ Moderate			
Generic measures (9)										
Profile measures (3/9)										
SF-36v1	8	+ Unknown					+ Limited	+ Moderate		
SF-36v2	2						+ Moderate	+ Moderate	*ES only (small)*	+ Unknown
SF-12	4						+ Limited	+ Limited		+ Moderate
Generic Index (2/9)										
LiSat-11	1	+ Unknown						+ Unknown		
QOLS	1							+ Unknown		
Preference-based Utility measures (4/9)										
15D	1						? Unknown	+ Limited		
EQ-5D-3L	3						+/- Limited	+/- Limited		- Moderate
HUI-3	7		_ Unknown				+/- Conflicting	+ Moderate	*ES only (small)*	
SF-6D	1						+ Limited	+ Limited		+ Moderate

Footnote: [a] Data synthesis: The data were qualitatively synthesized to determine the overall quality of measurement properties and acceptability of each reviewed HRQoL measure. 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 measure; 3) the results for each measurement property for each measure; and 4) the consistency of results between reviewed studies.

The data synthesis score has two elements[11,12,20]:

First, the overall quality of a measurement property was reported as: adequate (+), not adequate (-), conflicting (+/-), or unclear/indeterminate (?).

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

Where the data entry box is left blank, this signifies no available evidence.

[b] PROM acronyms (detailed in text and Tables 1 and 2)

OHCA – PROM Review – Appendices: Resuscitation August 2017

Appendix 1: Search strategy: review of HRQoL measures in OHCA

1.1 Database: Embase Classic+Embase <1947 to 2017 Week 09>

Search Strategy:

---

1 ((cardiac or heart or circulatory or cardiorespiratory or cardiopulmonary or postcardiac or post-cardiac or post cardiac) adj1 (arrest or stop or resuscitation)).mp. (86618)

2 (HR-PRO or HRPRO or HRQL or HRQoL or QL or QoL or (PRO or PROs or PROM or PROMs)).ti,ab. or quality of life.mp. or (health index* or health indices or health profile*).ti,ab. or health status.mp. or ((patient or self or proxy) adj (appraisal* or appraised or report or reported or reporting or rated or rating* or based or assessed or assessment*)).ti,ab. or ((disability or function or functional or functions or subjective or utility or utilities or wellbeing or well being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire* or profile or profiles or scale or scales or score or scores or status or survey or surveys)).ti,ab. or ((neurological or functional) adj2 (measurement or outcome or status)).ti,ab. or neurological functional outcome.mp. or functional neurological outcome.mp. (1053239)

3 (instrumentation or methods).sh. or (Validation Studies or Comparative Study).pt. or exp Psychometrics/ or psychometr*.ti,ab. or (clinimetr* or clinometr*).tw. or exp "Outcome Assessment (Health Care)"/ or outcome assessment.ti,ab. or outcome measure*.tw. or exp Observer Variation/ or observer variation.ti,ab. or exp Health Status Indicators/ or exp "Reproducibility of Results"/ or reproducib*.ti,ab. or exp Discriminant Analysis/ or (reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or "internal consistency").ti,ab. or (cronbach* and (alpha or alphas)).ti,ab. or (item and (correlation* or selection* or reduction*)).ti,ab. or (agreement or precision or imprecision or "precise values" or test-retest).ti,ab. or (test and retest).ti,ab. or (reliab* and (test or retest)).ti,ab. or (stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or interassay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappa's or kappas or repeatab*).ti,ab. or ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab. or (generaliza* or generalisa* or concordance).ti,ab. or (intraclass and correlation*).ti,ab. or (discriminative or "known group" or factor analysis or factor analyses or dimension* or subscale*).ti,ab. or (multitrait and scaling and (analysis or analyses)).ti,ab. or (item discriminant or interscale correlation* or error or errors or "individual variability").ti,ab. or (variability and (analysis or values)).ti,ab. or (uncertainty and (measurement or measuring)).ti,ab. or ("standard error of measurement" or sensitiv* or responsive*).ti,ab. or ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab. or (small* and (real or detectable) and (change or

difference)).ti,ab. or (meaningful change or "ceiling effect" or "floor effect" or "Item response model" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross-cultural equivalence").ti,ab. (5437246)

4 (addresses or biography or case reports or comment or directory or editorial or festschrift or interview or lectures or legal cases or legislation or letter or news or newspaper article or patient education handout or popular works or congresses or consensus development conference or consensus development conference, nih or practice guideline).pt. not (*animals/ not *humans/) (1512263)

5 (1 and 2 and 3) not 4 (1715)

6 limit 5 to (human and english language and yr="2004 -Current" and (adult <18 to 64 years> or aged <65+ years>)) (829)

1.2 Database: Ovid MEDLINE(R) <1946 to February Week 3 2017>

Search Strategy:

---

1 ((cardiac or heart or circulatory or cardiorespiratory or cardiopulmonary or postcardiac or post-cardiac or post cardiac) adj1 (arrest or stop or resuscitation)).mp. (55708)

2 (HR-PRO or HRPRO or HRQL or HRQoL or QL or QoL or (PRO or PROs or PROM or PROMs)).ti,ab. or quality of life.mp. or (health index* or health indices or health profile*).ti,ab. or health status.mp. or ((patient or self or proxy) adj (appraisal* or appraised or report or reported or reporting or rated or rating* or based or assessed or assessment*)).ti,ab. or ((disability or function or functional or functions or subjective or utility or utilities or wellbeing or well being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire* or profile or profiles or scale or scales or score or scores or status or survey or surveys)).ti,ab. or ((neurological or functional) adj2 (measurement or outcome or status)).ti,ab. or neurological functional outcome.mp. or functional neurological outcome.mp. (640349)

3 (instrumentation or methods).sh. or (Validation Studies or Comparative Study).pt. or exp Psychometrics/ or psychometr*.ti,ab. or (clinimetr* or clinometr*).tw. or exp "Outcome Assessment (Health Care)"/ or outcome assessment.ti,ab. or outcome measure*.tw. or exp Observer Variation/ or observer variation.ti,ab. or exp Health Status Indicators/ or exp "Reproducibility of Results"/ or reproducib*.ti,ab. or exp Discriminant Analysis/ or (reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or "internal consistency").ti,ab. or (cronbach* and (alpha or alphas)).ti,ab. or (item and (correlation* or selection* or reduction*)).ti,ab. or (agreement or precision or imprecision or "precise values" or test-retest).ti,ab. or (test and retest).ti,ab. or (reliab* and (test or retest)).ti,ab. or (stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or

inter-observer or intraobserver or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or interassay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappa's or kappas or repeatab*).ti,ab. or ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab. or (generaliza* or generalisa* or concordance).ti,ab. or (intraclass and correlation*).ti,ab. or (discriminative or "known group" or factor analysis or factor analyses or dimension* or subscale*).ti,ab. or (multitrait and scaling and (analysis or analyses)).ti,ab. or (item discriminant or interscale correlation* or error or errors or "individual variability").ti,ab. or (variability and (analysis or values)).ti,ab. or (uncertainty and (measurement or measuring)).ti,ab. or ("standard error of measurement" or sensitiv* or responsive*).ti,ab. or ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab. or (small* and (real or detectable) and (change or difference)).ti,ab. or (meaningful change or "ceiling effect" or "floor effect" or "Item response model" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross-cultural equivalence").ti,ab. (5652335)

4 (addresses or biography or case reports or comment or directory or editorial or festschrift or interview or lectures or legal cases or legislation or letter or news or newspaper article or patient education handout or popular works or congresses or consensus development conference or consensus development conference, nih or practice guideline).pt. not (*animals/ not *humans/) (3567543)

5 (1 and 2 and 3) not 4 (1157)

6 limit 5 to (english language and humans and yr="2004 -Current" and "all adult (19 plus years)") (588)

1.3 Named author searches:

1 ((cardiac or heart or circulatory or cardiorespiratory or cardiopulmonary or postcardiac or post-cardiac or post cardiac) adj1 (arrest or stop or resuscitation)).mp. (55150)

2 (HR-PRO or HRPRO or HRQL or HRQoL or QL or QoL or (PRO or PROs or PROM or PROMs)).ti,ab. or quality of life.mp. or (health index* or health indices or health profile*).ti,ab. or health status.mp. or ((patient or self or proxy) adj (appraisal* or appraised or report or reported or reporting or rated or rating* or based or assessed or assessment*)).ti,ab. or ((disability or function or functional or functions or subjective or utility or utilities or wellbeing or well being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire* or profile or profiles or scale or scales or score or scores or status or survey or surveys)).ti,ab. or ((neurological or functional) adj2 (measurement or outcome or status)).ti,ab. or neurological functional outcome.mp. or functional neurological outcome.mp. (628867)

3 NAME AUTHOR

Clifton Callaway

Tobias Cronberg

Veronique Moulaert

Graham Nichol

Karen Smith

4 1 and 3

5 2 and 4

Results:

Clifton Callaway (29); Tobias Cronberg (24); Veronique Moulaert (7); Graham Nichol (23); Karen Smith (8)

Appendix 2: HRQoL review - Details of the included studies (n25)

	Authors (Country)	Study design and setting *(Primary or secondary focus PROM evaluation)*	Participants (n, age and gender)	Measures (Mean (SD))	Study detail	Measurement/ Practical property assessed	Study quality *(COSMIN where applicable)*
46	Andrew et al (2016) Australia	Prospective cohort study – cross-section al assessment at 12mth follow up of adults who survive OHCA. Identified from CA Registry *Primary focus PROM evaluation*	1188 687 (80.7%) 50 interview 157 proxy Mean 60.0 (15.0) Range 18-75+ Male 79.0%	SF-12v1 EQ-5D-3L GOS-E SF-6D GOS-E 55.6% had GOS-E scores >7 (41.1% if those who died post-discharge included)	Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1486/1621 (92%) alive at 12/12 1188/1486 (79% of those known to be alive at 12/12) responded to interviews (265 lost; 39 unsuitable for interview): response rate 82% Patient responders vs proxy: patients younger, CA due to cardiac aetiology, discharged to home, working prior to CA, return to work post CA ($p<0.05$) (80.7% interviews at 1yr; majority of non-responders were 'lost to follow-up')	Interpretability Hypothesis testing – convergent validity: *hypothesized associations not stated a priori – but can be assumed* (but results in lower COSMIN score)	COSMIN: Hypothesis testing; Fair Interpretability: authors do not determine MIC or MID (COSMIN = POOR)
				SF-12 v1 (Median (IQR)) MCS 56 (I52 – 59) (full range 11-70) PCS 50 (40 – 56) (full range 11-70)	Patient only completion (no proxy – evidence that not a valid reflection of score)	Missing items at scale level: 3.4% (34/1188) No end effects Convergent validity: PCS with: SF-6D 0.69 EQ-5D-3L 0.63 GOSE 0.440 MCS 0.036 MCS with: SF-6D 0.558 EQ-5D-3L 0.29 GOSE 0.314 PCS 0.036 KGV: MCS Gender: Male 57 (52-59); Female 56 (48-59) (p=0.051) Discharged home: Yes 57 (52-59); No 54 (48-60) (p=0.14) Return to Work: Yes 56 (53-59); No 53 (44-58) ($p<0.001$) KGV: PCS	Interpretability: *(assumed how missing items were dealt with) BUT – No MIC or MID: so COSMIN = POOR)* Hypothesis testing: Fair *(no hypothesis; limited evidence re comparators)* KGV: Fair

						Gender: Male 51 (41-56); Female 45 (33-53) (p<0.001) Discharged home: Yes 51 (40-56); No 44 (37-55) (p=0.003) Return to Work: Yes 54 (48-57); No 45 (36-54) (p<0.001)	
				EQ-5D-3L Patients median 0.85 (IQR 0.76 – 1.00) (full range -0.595 – 1.00) Proxy median 0.75 (IQR 0.52 – 0.88) (full range -0.595 – 1.00)	Comparison of score distribution for EQ-5D-3L and SF-6D against GOSE scores (fig 2): demonstrate a statistically significant difference in score distribution for patients reporting a lower or upper moderate GOSE and lower or upper good GOSE (p<0.001) For EQ-5D-3L: 75th percentile is equal to a score of 1.00 for all GOSE categories above lower moderate disability *(strong evidence of significant ceiling effect)* (Supplementary evidence – moderate correlation between similar domains on SF-6D and EQ-5D-3L)	Missing items at scale level: Total - 1.0% (12/1188) Patients – 0.5% (5/928) Proxy – 2.7% (7/260) Large ceiling effects: Patients 46% (n421); Proxy 23% For those patients reporting full health (n421), substantial variability on SF-6D and GOSE – patients more likely to report problems with mental health and vitality Convergent validity: Patients: EQ-5D-3L with: SF-6D 0.65 SF-12 PCS 0.62 SF-12 MCS 0.29 GOSE 0.47 Proxy: EQ-5D-3L with: GOS-E 0.67 KGV: Statistically significant difference between patients (0.85) and proxy (0.75)(p<0.001) KGV: Gender: Male 0.88 (0.78-1.0); Female 0..85 (0.73-1.0) (p<0.001) Discharged home: Yes 0.88 (0.7-1.0); No 0.85 (0.71-1.0) (p=0.09) Return to Work: Yes 1.00 (0.85-1.0); No 0.85 (0.69-1.0) (p<0.001)	Interpretability: (assumed how missing items were dealt with) BUT – No MIC or MID: so COSMIN = POOR) Hypothesis testing: Fair *(no hypothesis; limited evidence re comparators)* *Fair* KGV: Fair
				GOSE Total population (patients and proxy combined): majority		Missing items at item level: 0.5% (6/1188) Patients: 0.5% (5/928) Proxy: 0.4% (1/260)	Interpretability: (assumed how missing items were dealt with) BUT – No MIC or

				lower good recovery (388/1188; 32.8%) Patients: majority lower (36.8%) or upper good (35.3%) recovery (total 72%) (full range: lower severe (92/928 (7.8%) to upper good (381/928 (32.2%)) Proxy: majority lower severe (27.4%) or upper severe (11.2%) recovery (total 39%) (full range: vegetative state (4/260 (1.5%) to upper good 55/260 (21.2%))		Total population – full range of score. >30% patient responders lower/upper good at 12/12 Patients: majority lower (36.8%) or upper good (35.3%) recovery (total 72%) (full range: lower severe (92/928 (7.8%) to upper good (381/928 (32.2%)) Proxy: majority lower severe (27.4%) or upper severe (11.2%) recovery (total 39%) (full range: vegetative state (4/260 (1.5%) to upper good 55/260 (21.2%)) Convergent validity: Patients / Proxy: GOSE with: SF-6D 0.52 / NA EQ-5D-3L 0.47 / 0.67 SF-12 PCS 0.440 / NA SF-12 MCS 0.314 / NA KGV: cut-point >7.0 Return to work: Yes n296 (77.3%); No n60 (48.0) (p<0.001) Discharged home: Yes n604 (73.7%); No n57 (59.4%) (p<0.001)	MID: so COSMIN = POOR) Hypothesis testing: Fair *(no hypothesis; limited evidence re comparators)* KGV: Fair
				SF-6D Median 0.86 (IQR 0.72 – 0.92)	Comparison of score distribution for EQ-5D-3L and SF-6D against GOSE scores (fig 2): demonstrate a statistically significant difference in score distribution for patients reporting a lower or upper moderate GOSE and lower or upper good GOSE (p<0.001) For EQ-5D-3L: 75th percentile is equal to a score of 1.00 for all GOSE categories above lower moderate disability *(strong evidence of significant ceiling effect) – but there is strong evidence of variability in SF-6D score across these categories (in patients reporting full health on the EQ-5D-3L).* *SF-6D scores improved as GOSE scores improved (fig 2)* (Supplementary evidence – moderate correlation between similar domains on SF-6D and EQ-5D-3L)	Missing items at scale level: 3.4% (34/1188) 7% full health (full range 0.345 to 1.00) No end effects Convergent validity: SF-6D with: SF-12 PCS 0.69 EQ-5D-3L 0.65 MCS 0.56 GOSE 0.52 SF-6D items with EQ-5D-3L items *(smaller than assumed hypothesized association -– hence negative ratings (<0.50):* PF range: Anx/Dep 0.22 to Usual activity 0.52 and Mobility 0.51 (+) Role limit range: Self-care 0.20 to UA 0.42 (-)	Interpretability: (assumed how missing items were dealt with) BUT – No MIC or MID: so COSMIN = POOR) Hypothesis testing: Fair *(no hypothesis; limited evidence re comparators)*

						Social function range: Self-care 0.12 to UA 0.33 (-) Pain range: Anx/Dep 0.21 to Pain 0.57 (+) MH range: Self-care 0.12 to Anx/Dep 0.45 (-) Vit range: Self-care 0.2 to Mobility 0.35 (-) KGV: Gender: Male 0.86 (0.72-0.92); Female 0.79 (0.66-0.86) (p<0.001) Discharged home: Yes 0.86 (0.72-0.92); No 0.79 (0.67-0.86) (p<0.001) Return to Work: Yes 0.86 (0.80-0.92); No 0.74 (0.63-0.86) (p<0.001)	KGV: Fair
44	Beesems et al (2014) NL	Cross-sectional cohort study – survivors of OHCA at 6-12mtns. *Secondary focus PROM evaluation*	220	SF-12 PCS <60yrs 45.8 (9.0) 60-80 46.7 (9.6) >80 40.5 (9.6) SF-12 MCS <60yrs 52.2 (8.6) 60-80 54.3 (8.0) >80yrs 53.2 (10.4)	Telephone interview administration of a battery of questionnaires to evaluate the QoL, neuro-cognitive function and independency in ADL of patients; and caregiver strain. Focus SF-12, CPC, MRS; also include Telephone Interview for cognitive status (TICS) Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) Caregiver Strain Index	Known-groups validity (no a priori hypothesis (but can be assumed)): Statistically significant difference between younger (<80yrs) and older (>80yrs) on SF-12 PCS (40.5 (9.6) Authors suggest that telephone completion of measures is 'simple, feasible and cheap and can be recommended as standard for studies after OHCA'. *However, they don't report on the timing or associated cost.*	More descriptive application than evaluative of MP/PP *COSMIN: Fair*
				OPC and CPC MRS	Evidence suggests the the CPC at discharge over-estimates functional ability: where 94% (n234) were classified as CPC 1 at discharge; at 6/12 just 84% were independent in daily life (MRS)		
40	Boyce-van der Wal et al (2015) NL	Prospective cohort study – consecutive OHCA survivors referred for cardiac rehabilitation (2011-2013) *Secondary focus PROM evaluation*	77 Mean age 57.2yrs (13.8) Male 82%	SF-36 *(unclear if version 1 or 2)* Also: Mini-Mental State Exam (MMSE) Cognitive Failure Questionnaire (CFQ) Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) Impact on Participation and	Interview administration of selected measures of cognitive functioning, participation and autonomy, and HRQoL -within 4-weeks of OHCA: Focus of evaluation on cognitive function	Known-groups validity (no a priori hypothesis (but can be assumed)): Statistically significant difference between cognitively impaired and non-impaired groups across 6/8 domains of the SF-36 (not BP or GH). Convergent validity (no a priori hypothesis; all analyses not illustrated; missing data not reported): SF-36 SF with MMSE r=0.317 SF-36 RE with CFQ r=-0.400	*COSMIN: KGV: Fair (missing data /data quality not reported)* *Hypothesis testing: Fair (no hypothesis; limited evidence re comparators; missing data /*

				Autonomy Questionnaire (IPAQ) Hospital Anxiety and Depression Scale (HADS)		SF-36 BP with CFQ r= -0.366 SF-36 MS with CFQ r=-0.351 SF-36 Vt with CFQ r= -0.250 SF-36 SF with IQCODE r=-0.412 SF-36 Vt with IQCODE r=-0.332	*data quality note reported)*
35	Bro-Jeppesen et al (2009) Denmark	Prospective cohort study of comatose OHCA survivors consecutively admitted over a 4-year period: first 2-years no therapeutic hypothermia (TH); second 2-years all received TH. *Secondary focus PROM evaluation*	Total: 156 Control No TH (77) Intervention TH (79)	SF-36 *(unclear if version 1 or 2)* Also: Cerebral Performance Checklist (CPC) (retrospective assessment) MMSE	Aim to assess the long-term outcome of OHCA survivors – assessment of cognitive function and QoL before and after implementation of TH. Follow-up interviews at 6-mths: completion of SF-36 and MMSE.	Known-groups validity (no a priori hypothesis (but can be assumed)): Non-statistically significant difference between Control (No TH) and Intervention (TH) groups at 6/12. Sub-domains RP and RE were lower in the control period (not NSS). RP and RE lower than Dutch population norms (p<0.05). Reports 'Cronbach's alpha' for SF-36 domains: range 0.88 to 0.91. Data not presented.	*COSMIN:* *KGV: Poor (missing data /data quality not reported)* *Internal consistency reliability: Poor (missing data /data quality not reported)*
43	Deasy et al (2013) Australia	Prospective cohort study – long-term follow up of young adults who survive OHCA Median follow-up 5 years (range 2.7 to 8.6yrs)	56/106 survivors participated at long-term follow-up (53%) Age range 18-39 yrs Male 74%		Telephone administration of selected measures: to explore the functional and QOL outcomes of OHCA survivors. Initial cohort (106) age 29 (range 23-35) At Follow-up (53) age 35 (range 29-41)	*Telephone administration 'good completion rates' but reports challenges of loss to follow-up.*	
		Secondary focus PROM evaluation		EQ-5D-3L *Item level scores only – no index*	Reports item level (frequency endorsement) only: >70% report no problems with mobility, personal care, pain; *Possible end effects* >30% report problems with usual activities >60% with anxiety/depression *No comment on MP etc.*		*COSMIN: possibly* *Interpretation – data quality*
				SF-12v1 PCS 46.5 (10.0) MCS 38.0 (14.5)	Low MH scores (38.0) when compared to Aus pop; PCS 45.6.(scores <40 = mod to severe disability). Reports item level (freq endorsement) and score data *– possible end effects for social items and mental health items (see table 3)– but no comment on Psych etc.*		*COSMIN:* *KGV (no a priori hypothesis (could be assumed)): Fair*
				GOS-E	Item level frequency endorsement illustrated	Reports item level (freq endorsement) - no comment on MP etc.	

				(CPC)	Obtained from hospital discharge records for uncontactable survivors only.	*Critique: not discriminate between survivors with high cerebral function or quality of life.* *Criticises CPC for its bluntness and inability to detect major QoL and long-term issues. But does not provide any psych evidence.*	
41	Cronberg et al (2015) Sweden	RCT (TTM Trial) *Secondary focus PROM evaluation*	939 Follow-up at 6/12: 455/491 (92%) attended structured interview	SF-36 v2 MCS: 48.4 (12.7) vs 48.3 (13.4) (P =.79) PCS: 45.8 (10.5) vs 46.3 (11.3) (P =.45). 'Two Simple Questions' (daily function and mental recovery – see [18] CPC MRS	Face to face (92%) or telephone interviews at 6/12. Self-complete approx. 92%; proxy approx. 8% No between group difference on SF-36 for survivors in either arm of the trial (33vs36): MCS: 48.4 (12.7) vs 48.3 (13.4) (P =.79) PCS: 45.8 (10.5) vs 46.3 (11.3) (P =.45). Authors report that the values are comparable to the population norm (+/> 47). Two Simple Questions – no between group difference.	Descriptive stats for SF-36: limited evidence of KGV (but no hypothesis and limited data) Limited evidence re CPC and MRS (ref to 39): *suggests (p638) 'Require tests that …improve the discrimination of the degree of neurological recovery. The mRS has higher resolution than the CPC, and is evident [see 39] … that many patients with the highest CPC still have degrees of disability and dependence by their MRS scores'.*	Largely descriptive *COSMIN:* *KGV: Fair (limited)*
34	Graf et al (2008) Germany	Cross-sectional evaluation at 5-years post-ICU discharge (patient level assessment of health status and fully costed economic evaluation) *Secondary focus PROM evaluation*	110 / 354 (31%) alive at 5-years N=81 (74% of all 5-year survivors) completed 5-year questionnaire. Age of surviving cohort not reported. Age of original cohort of survivors (n354) 66 (13)	SF-36 (v1)	Postal self-completion of questionnaire including SF-36 (German (V1)) at 5-years	Known-groups validity (no a priori hypothesis (but can be assumed)): SF-36 domain scores compared to gender and age-matched German control (norm population): 6/8 SF-domains lower than population norm values (Not Pain or Emotional Role). Statistical significance not reported. Reports 'Cronbach's alpha' for SF-36 domains: range 0.88 to 0.91. Data not presented.	*COSMIN:* *KGV: Poor (missing data /data quality not reported)* *Internal consistency reliability: Poor (missing data /data quality not reported)*
38	Kowalik et al (2014) Poland	Retrospective, observational study: comparison between survivors treated with mild therapeutic hypothermia (MTH)(28) versus historical controls (did not receive MTH)(37)	Total n=31 OHCA survivors completed follow-up interviews 16/28 Intervention - MTH Versus 15/37 Control (No MTH)	SF-36 *(v12)* Also: Disability Rating Scale (DRS) Bathel Index	Telephone administration at between 1 and 4 1/2years post-OHCA survival. Postal administration by those non-contactable by telephone.	Known-groups validity (no a priori hypothesis): Statistically significance between group difference for survivors in receipt of MTH versus No MTH: 2/8 domains only: Role Limitations due to Emotional problems and Vitality were better in survivors who received MTH.	*COSMIN:* *KGV: Poor (missing data /data quality not reported)*

		Secondary focus PROM evaluation	MTH 55.6 yrs (2.8) No MTH 59.4 (2.9) Male 11 (10%)			Convergent validity (no a priori hypothesis; all analyses not illustrated; missing data not reported): SF-36 Vt with DRS r=0.405 SF-36 SF with DRS r= 0.391 SF-36 GH with DRS r= 0.351	*Hypothesis testing: Poor (no hypothesis; limited evidence re comparators; missing data / data quality note reported)*
54	Longstreth et al (2012) Canada	Prospective cohort sub-study of an RCT (x-section assessment at 3/12): Initial 'evaluation' of 'Two Simple Questions' in CA survivors: 1) In the last 2-weeks, did you require help from another person for your everyday activities? 2) Do you feel that you have made a complete mental recovery from your heart arrest? *Response:* Yes or No. *Interpretation (classify as):* Dependent –'Yes' to Q1. Independent –'Yes' to Q1 and 'No' to Q2 Full recovery – 'No' to Q1 and 'Yes' to Q2. *Primary / Secondary focus PROM evaluation: apply COSMIN*	32 CA survivors 55.6 (16.1) Male 87.5%	'Two simple questions': Respondents classified as: 1).Full recovery (n=24 (75%)); 2).Independent (n=3 (9%)) 3).Dependent (n=5 (16%)) MMSE (ALFI) 19.1 (5.1) HUI3 0.76 (0.28)	3-mth post hospital discharge - Interview (telephone) administration: n=32 survivors *Missing data for 8 survivors (3 full recovery; 3 independent; 2 dependent – suggests that missing most often in those not reporting full recovery so HUI3 scores in this sample may overestimate QoL)* *Authors suggest the two items can be used to compliment scores from GOS or CPC (see figure 1).(but evidence is limited / not explored)*	When categorised per 'Two Simple Questions' (KGV) – statistically significant between group difference on HUI3 and MMSE: HUI3: 0.82 (full recovery) vs 0.30 (dependent) (n=32) MMSE: 20.4 (full recovery)/ 18.7 (independent) / 11.2 (dependent)(n=32) CPC: Authors suggest that CPC at hospital discharge overestimates recovery compared to results at 3mths.	Small numbers COSMIN: Hypothesis testing: 1.*Known groups validity for HUI3: grouping per 2Q – but this is a 'new approach' with no reporting of evidence of MP. Size small (<30 = poor) = Fair. Missing data – assumed (Fair)*
33	Mak et al (2016) NL	Cross-sectional evaluation of the 'Structured' CPC in CA survivors (more than 6/12 survival) *Format and completion of the CPC changed to enhance the content validity of the measure – in effect, this is a different measure to the original CPC. But the detail provided is insufficient to allow for reproduction.*	59 Median age 62 (range 20-88) Male 84%	'Structured CPC' Construct variables: Cognitive Failure Q (CFQ) Barthel Index (BI) Frenchay Activities Index (FAI) Community Integration Questionnaire (CIQ)	'Structured CPC' to ensure uniform, interview-based administration. Structured CPC administered by telephone interview in those who responded to (and completed) a postal questionnaire containing other listed measures. Data quality: non-normal distribution for all measures: median (range mix to max): CFQ 73.00 (11-98) (possible 0-100, where higher scores = more cognitive complaints)	*A priori hypothesized association between variables stated (p7): low correlation with body structure and function; moderate with activities and participation; low with QoL:* Association between Structured CPC and: CFQ -0.40; BI -0.57; FAI -0.65; CIQ -0.53; QOLIBRI -0.67	Focus on validity of CPC COSMIN: 1.Hypotheses testing: Good

		Primary focus construct validity of Structured CPC: apply COSMIN		Quality of Life after Brain Injury (QOLIBRI)	BI 20.00 (13-20) (possible 0-20, where higher scores = greater independence)) – *scores suggest basic ADL not a problem (??ceiling effect)* FAI 27.00 (4-42) (possible 0-45, where higher scores = better functioning) – *scores suggest greater difficulties with more complex activities* CIQ 18.00 (5-26)(possible 0-29, where higher scores = higher levels of participation) – *scores suggest some limitation with integration / participation* QOLIBRI 76.35 (13.51-100.00)(possible 0-100, where higher scores = better QOL) – *scores suggest some QOL limitations*	*Correlations greater than hypothesized: authors suggest that the 'Structured CPC' is better able to assess the 'way people value their own lives',*	
32	Middlekamp et al (2007) NL	Retrospective cohort study (x-section completion) *Secondary focus PROM evaluation: apply COSMIN*	20/32 50.0 (12)(range 17-64yrs)	Cognitive Failure Q (CFQ)(0-100) Impact on participation and autonomy questionnaire (IPAQ)(4 domains; range 0-4) Frenchay Activities Index (FAI)(0-45) Quality of Life after Brain Injury (QOLIBRI)	Patients with hypoxic injury post CA (2-7yrs earlier) who had been admitted to rehabilitation facility. Invited to self-complete a postal questionnaire Results: suggest problems with cognitive and daily function and participation remain: *Median (25-75%)* CFQ (possible 0-100): 52.0 (29.5-68.8) *scores suggest some cognition difficulties* IPAQ (4 domains; possible range 0-4) Domain 1: 1.0 (0.0-1.8); Domain 2: 1.5 (1.0-3.0) Domain 3: 2.0 (1.0-3.0); Domain 4: 1.0 (1.0-2.0) *scores suggest some limitation with integration / participation* FAI (possible 0-45): 23.0 (12.0-32.0) *scores suggest some limitation with ADL* QOLIBRI: 2 domains: 1. Satisfaction (possible 42-210): 126.0 (90-165.0) 2. Problems and complaints (possible 14-70): 29.0 (20.5-33.8) *scores suggest some QOL limitations*	Association between variables *(no a priori hypothesized association):* Coma duration and: CFQ r=0.57 IPAQ domains range r= 0.57 to 0.67 QOLIBRI part 1 r=-0.70 *(suggesting a worse QOL for those with longer duration of coma)* Post-traumatic amnesia and: FAI r=-0.70 *(indicating lower levels of ADL for those with longer duration of PT amnesia)* QOLIBRI r=-0.70 QOLIBRI parts I and II with: CFQ r=-0.86 (part I); r=-0.77 (part II) IPAQ domains range (part I) -0.44 to -0.86; (part II) 0.53 to 0.73 FAI r=0.34 (part I) and -0.41 (part II)	COSMIN: 1.Hypotheses testing: Poor (Small sample)
36	Moulaert et al (2010) NL	Retrospective cohort study of CA survivors (n63) – to study factors related to QoL	63 OHCA survivors	SF-36v1 Median score: PCS 71.8; MCS 73.0	Postal self-completion of questionnaire: SF-36 primary outcome measure (many measures included).	Reports missing data for all measures – 2/63 patients had >15% missing items for SF-36 (same as FAI, CIQ, CFQ, FSS).	COSMIN:

		after hypoxic period due to CA *Secondary focus possibly PROM evaluation??: apply COSMIN*	57.3 (12.5); range 18-81 (time of arrest). 60.2 (12.7); range 20-85 (at follow-up)	(Dutch norm 76 and 78). *Community Integration Questionnaire (CIQ)* *Barthel Index (BI)* *Frenchay Activities Index (FAI)* *HADS* *Cognitive Failures Questionnaire (CFQ)* *Impact of Event Scale (IES)* *Fatigue Severity Scale (FSS) 4.3 (1.6)*		Median score: PCS 71.8; MCS 73.0 (Dutch norm 76 and 78). *Both lower than general pop* Convergent validity *(no hypotheses / hypotheses assumed*) Correlation between SF-36 PCS and included measures range: 0.22 (BI) to 0.55 (CFQ) and 0.61 (FSS); MCS range 0.10 (BI) and 0.57 (FSS) and 0.77 (HADS).	Hypothesis testing: 1.Convergent V: Fair *(hypotheses not formulated by possible to deduce) – but could be as low as Poor (no information on MP of comparator instruments)* *2. Known groups validity: Fair*
39	Moulaert et al (2015) NL	RCT of brief nursing intervention for CA survivors *Secondary focus PROM evaluation: apply COSMIN*	143/ 185 CA survivors completed the trial Mean 60.0 (12.0) Male 83% 155 caregivers	Primary – societal participation and QoL *Community Integration Questionnaire (CIQ)* *SF-36v2 – for domains only (Table 4) (not PCS or MCS)* *EuroQoL EQ-VAS* *Secondary – include* *HADS* *Frenchay Activities Index* *Cognitive Failures Questionnaire* *Impact of Event Scale*	Assessments at 2-weeks, 3/12 and 12/12 after CA. SF-36 profile scores (Baseline; 3 and 12-mths): Statistically significant between group difference at 12-months (in favour of the intervention) for: SF-36 role emotional, mental health and general health; plus for HADS total and Anxiety domain. *No between group difference at 3/12. Does this provide support for longer-term follow-up? The results suggest there is potential for improvement at 12/12 (however, does not include an interim of 6 or 9-mths)*	Reports missing data for all measures – 2/63 patients had >15% missing items for SF-36 (same as FAI, CIQ, CFQ, FSS). Correlation between SF-36 PCS and included measures range: 0.22 (BI) to 0.55 (CFQ) and 0.61 (FSS); MCS range 0.10 (BI) and 0.57 (FSS) and 0.77 (HADS).	Application of measures in RCT – limited evidence COSMIN: Hypothesis testing: 1.Convergent V: Fair *(hypotheses not formulated by possible to deduce) – but could be as low as Poor (no information on MP of comparator instruments)* *2. Known groups validity: Fair*
50	Nichol et al (1999) Canada	Prospective cohort study (sub-study of RCT) – survivors followed for up to 6-mths. *Secondary focus PROM evaluation:*	86/96 survivors completed interviews (90% of available patients; 68% survivors to discharge) *6 'lost to follow-up'; 4 unable to*	HUI3 0.72 (0.22) Range 0 to 1.00 Positive skew towards higher scores – fig 3. *Mean utility significantly worse*	Telephone admin at 6-mths; proxy if patient unable to communicate (2.3%). HUI3 index: Score 1.0 6% Score +/> 0.9 33% *Score <0.9 67%*	HUI3: Construct validity *(no a priori hypothesis (?assumed)):* Known-groups – by duration of resuscitation: Shorter (<2mins): 0.81 Moderate (3-10 mins) 0.76 Longer (>10mins) 0.65 (p=0.05)	Smaller pop – but 'good' for COSMIN (n=86) Focus on HUI3 COSMIN: 1.Hypotheses testing: KGV:

			complete the interview 65.0 (14.0) Male 55%	*than Gen pop (0.86 (0.16)) and those whose activities were not limited by chronic disease (0.91 (0.08))(p<0.01)* CPC	*Scores on descriptive system:* Attributes where greater impairment observed (higher score, more impairment): emotion (>50% score between 2 and 4) pain (>40% score between 2 and 5) ambulation (40% score between 2 and 6) cognition (40% score between 2 and 6)	KGV *(assumed hyp?)* Mean utility significantly worse than Gen pop (0.86 (0.16)) and those whose activities were not limited by chronic disease (0.91 (0.08))(p<0.01) Association between HUI3 and*(no a priori hypothesis)* : CPC (pre-discharge) -0.29 MMSE (pre-discharge) 0.37	1.Gen pop: Good 2.Resus: Good Construct: No evidence MP or item level (assumed): Fair/Poor
56	Nichol et al (2015) Canada	Prospective cohort sub-study of the ROC PRIMED RCT. *Secondary focus PROM evaluation: apply COSMIN*	729 (56%) of survivors consented; 644 of respondents completed 1 or more assessment(s) (88%) 59.1 (14.1) *(Consented and interviewed n644)* Female n158 (24.5%)	HUI3 (*scores <0.70 = severe impairment*): 3/12 0.75 (0.33) 6/12 0.74 (0.35) Worst 0.71 (0.35); Best 0.77 (0.33) *Range and end effects NR. ?Skewed distribution* mRS GDS	HUI3 Telephone administration at 3 and 6-mths post discharge mRS – assessed from clinical record prior to discharge Predictive model: Suggests that for every unit change in MRS, this predicts a change of -0.06 in the HUI3. *If the MCID is 0.03, then the change is greater than the MCID. If the MCID is 1.0, the change is less than the MCID>*	MRS: 'MRS at discharge independently associated with post-discharge variables: ALFI -1.30 (-1.64, -0.96) HUI3 -0.07 (-0.09, -0.05) T-GDS 0.67 (0.41,-0.93) HUI3 Greater post-discharge HRQoL significantly associated with less neuro impairment, cog impairment or depression (after adjustment for baseline characteristics and EMS processes)(p78). mRS, HUI3 and depression did not change 'significantly' over time from discharge (Table 2): HUI3 @ 3/12 0.75 (0.33); @ 6/12 0.74 (0.35). *Reduction of 0.10 (what's a MIC for the HUI3?)* KGV: ?? statistical significance of between group difference not reported. But 'better' mean HUI scores if: Arrest in public location vs private (difference 0.07) Initial shockable rhythm vs not (difference 0.13) *Hypothesis – not clear; but can be assumed.*	Large cohort *Difficult to apply COSMIN.* *Convergent Validity; hypotheses assumed. Fair.* *KG: Hypotheses assumed. Missing data – detail re how dealt with. Statistical significant not reported (? Stat analysis). Suggest Fair.*
43	Orbo et al (2016) Norway	Prospective longitudinal study of OHCA survivors – investigate cognitive	33 OHCA survivors completed both exams	Neuropsychological tests (range)	Survivors invited to participate in individual, out-patient, face-to-face interviews at 3-mths and 12-mths post resuscitation.	Missing data not reported for SF-36 (but was reported for HADS and other variables).	

		recovery from 3 to 12-mths post resuscitation. *Secondary focus PROM evaluation*	31 Males; mean age 58.6 years (SD13) *Note: n129 discharged alive; n79 eligible for the study (n8/129 severe anoxic brain injury); included in 3mth assessment n45; n33 completed 3 and 12-mths assessment.*	SF-36 (version 2) Also: HADS	Association between HRQl , psychological distress and work status after 12-mths	Effect size statistics for mean change from 3 to 12-mths (no external anchors or variables to determine direction of change; assumed improvement?): SF-36 MCS -0.38; PCS 0.21 HADS A 0.27; D 0.35 Statistically significant (paired t-test) reduction in MCS between 3 and 12/12mth post arrest (p=0.02); NS improvement for PCS T1 (3/12): MCS 51.85 (9.57); PCS 44.10 (8.55) T2 (12/12): MCS 47.71 (11.82); PCS 45.82 (8.04). NS change at domain level (data not illustrated). Reliable change index: (as a measure of the clinical significance of change scores): cut-off point of 40 used for the SF-36 summary scores: between 3 and 12/12: Deterioration: MCS 5 (15%) of patients; PCS 0. Improvement: MCS 0; PCS 4 (12%). No change: MCS 28 (86%); PCS 29 (88%) Correlation between variables (no a priori hypothesis (but can be assumed)): SF-36 MCS with: HADS A -0.82; HADS D -0.84 SF-36 PCS 0.49 Return to work 0.40 Cognitive composites range: 0.03 (Visual memory) to 0.38 (Executive composite) SF-36 PCS with: HADS a -0.35; HADS D -0.57 SF-36 MCS 0.49 Return to work 0.35 Cognitive composites range: 0.28 (Visual memory) to 0.43 (Visual Memory composite)	*COSMIN not consider ES or t-tests a measure of Longitudinal validity / responsiveness* *Interpretation: Poor (score distribution not reported; no MIC)* *Hypothesis testing: Fair (sample size; missing data)*

52	Raina et al (2008) USA	Prospective longitudinal study of OHCA survivors – examine the relationship between the CPC, MRS and HUI (but x-sectional evaluation) *Primary focus PROM evaluation -HUI3, CPC, mRS*	21 OHCA survivors 56.0 (18.0)	CPC (focus of paper) MRS HUI3	Convenience sample of 21 survivors. Medical chart review at discharge – to inform CPC and MRS. In-person interview at 1/12 to collect MRS and HUI3 (and to inform calculation of CPC at 1/12) *CPC – limited ability to discriminate between mild and moderate brain injury; CPC scores at discharge (record review) overestimate cognitive and disability status.*	***Inter-rater reliability:*** At discharge from medical notes (n=13; 2 raters): CPC 92.31% (kappa 0.87) MRS 61.54% (0.51) For interviews (n=10; 2 raters): CPC 70.0% (kappa 0.60) MRS 80% (kappa 0.67) HUI3 97.3% (kappa 0.96)(*?? Intra-rater or test-retest???)* ***Construct validity*** *(no a priori hypothesis* (assumed)): Discharge values: CPC with MRS = 0.79 1/12 scores with CPC discharge scores: CPC = 0.72 MRS = 0.47 HUI3 -0.41 1/12 scores with 1/12 CPC scores: MRS 0.70 HUI3 -0.71 ***Known groups*** – wide distribution of 1/12 interview completed CPC, MRS and HUI3 scores for each discharge CPC category (see p5) – the wider variability and overlap of scores within and between CPC categories suggests that the CPC may be insensitive to differences in impairments and disability among people with good, moderate and severe cerebral disability	Measurement focus: Association between CPC, global disability and QOL COSMIN: Reliability: Poor (small n) Hypothesis testing: Poor (small n)
55	Raina et al (2015) USA	Prospective cohort for up to one-year post OHCA *Secondary focus PROM evaluation:*	29/49 agreed to participate at 12/12 (just 5 had died) 60.8 (16.3) Male 62%	Global disability: COC, MRS, GOS-E QOL: HUI3 Depression: Geriatric Depression Scale (GDS) Activity Limitation: Performance Assessment of Self-	Hospital discharge – CPC from medical notes. Interview-administration at 1,6 and 12/12 (12/12 focus of evaluation) Mean scores at all four time points: CPC and MRS met definition for 'good outcome' (CPC 1-2; MRS 0-3) GOS-E at 1,6 and 12/12 suggest lower moderate disability. HUI3 scores indicate severe disability at all time-points: 0.59 (1/12); 0.68 (6/12); 0.66 (12/12)	Association between measures at 12/12 *(no hypo stated can be assumed):* CPC with: MRS 0.85; GOSE -0.67 HUI3 -0.35 RNLI -0.71 MRS with: CPC 0.85; GOSE -0.68 HUI3 -0.48 RNLI -0.69 GOSE with:	Focus on impairment, disability and QOL and association between measures COSMIN: Hypothesis testing – Fair (small sample size)

				care Skills (PASS) H-Habit, S-Skills Cognitiion: Telephone Interview of Cognitive Skills (TICS) Participation: Reintegration to Normal Living Index (RNLI)	*Suggest score interpretation for HUI3:* *1 Perfect health* *0.99 - 0.89 mild disability* *0.88 – 0.70 moderate disability* *<0.70 severe disability* Moderate association between the CPC, MRS and GOS-E suggests they measure similar constructs. Although the GOS-E has a better defined response scale (8-point) it did NOT demonstrate an ability to capture change over time (NS score change and very small ES). *Authors suggest that the MRS and GOSE are both preferable to the CPC because they take prior disability into consideration.*	CPC -0.67; MRS -0.68 HUI3 0.45 RNLI 0.77 HUI with: CPC -0.35 mRS -0.48 GOSE 0.45 GDS -0.45 MMSE 0.23 TICS 0.37 PASS-H 0.40 PASS-S 0.62 RNLI 0.54 Change in scores over time: Mean values reported at discharge (CPC and MRS only), 1, 6, 12/12. Statistically significant improvement in score for CPC and MRS at 12/12 (ES 1-6mths 0.73 and 0.84; 6-12 mths 0.21 and 0.00; 1-12mths 0.86 and 0.86) – suggests greatest change in first 6/12. But little change in GOS-E (NS change; ES 0.05 (1-6/12); 0.18 (6-12/12); 0.11 (1-12/12). Also ss change in RNLI at 12/12 compared to 6/12 and 1/12. NS change in HUI3 : moderate ES 1-6/12 (0.43); small 6-12/12 (0.19) and 1-12/12 (0.26).	Responsiveness: Not COSMIN standard evidence. Just reports mean and ES
37	Reinhard et al (2009) Estonia	Cross-sectional cohort of OHCA survivors *Secondary focus PROM evaluation*	44/57 (77.2% response rate) 56.0 (15.6) (range 19-80) Male 64%	SF-36 (RAND) v1.0	Postal questionnaire - to assess the long-term survival and QoL; and support comparison with general population and patients with MI and angina. Questionnaire sent 16-62 months post OHCA.	SF-36 profile (*No a priori hypothesis; can be assumed):* OHCA survivors rated health significantly worse than general population for 5/8 domains: general health, social functioning, emotional role functioning, physical role functioning and physical health. NSS difference for vitality, mental health or body pain. OHCA survivors with known cardiovascular disease (n30) rated their health as similar to patients with MI or	Descriptive. Comparison with general population. *COSMIN: Poor*

						angina (not requiring resuscitation): NSS for all SF-36 domains.	
51	Stiell et al (2003) Canada	Prospective cohort sub-study of the OPALS study – Adult OHCA patients from 20 cities (x-sectional evaluation) *Secondary focus PROM evaluation: but data quality only? Unable to apply COSMIN ******** *Evaluated the pre-hospital factors associated with HRQOL of survivors*	268/316 known-1-year survivors (84.8%) All cases: 69.0 (14.1) Males 66% HUI3: 64.4 (13.0) Male 81%	HUI3: median 0.80 (IQR 0.50 to 0.97) CPC	Telephone admin at 1yr (proxy for those unable to self-complete (n=??): HUI3 scores (0.80) comparable to age-adjusted pop (0.83). *BUT some evidence that MIC can be as small as 0.03 (not discussed)* Authors suggest that the HUI3 has 'face validity for cardiac arrest because the attributes measured are those that may be affected by patients with neurological or functional impairment' – *but this is not evaluated in the paper*	HUI3: No specific MP/PP Data quality Compares to general population – 'similar' – but statistical significance (or MCID) not explored. Score distribution: range -0.30 to 1.0 Score >0.80 = 134/316 (42% of known survivors) Score <0.80 = 134/316 (42%) 0.60-0.80 = 56/316 (18%) Score <0.60 = 78 (25%) (Unknown n=48) *So population with scores >0.60 = n190 = 75%* *Scores less than 0.60 = 25%* *But HUI3 scores <0.70 suggest severe impairment – so more than 25% of population are severely impaired* CPC: score distribution only Good 88%; Mod 9%; Severe 6%.	COSMIN: N/A
53	Stiell et al (2009) Canada	Prospective cohort sub-study of the OPALS study – Adult OHCA patients from 20 cities. Comparative evaluation of CPC and HUI3 *Primary focus PROM evaluation: apply COSMIN*	At 12-mths (n305) 63.9yrs (SD?) Male 78%	CPC (categorical): 1 (n)/ 2 (n)/ 3 (n) 267 / 26/ 12 (total 305) HUI3 median 0.84 (IQR 0.61 to 0.97) *?range -0.25 to 1.00 (difficult to read Fig3).* *Skewed distribution - End effects? – unclear how many scored 1.00 ??94/305)* *Gen pop: 0.85 (or 0.91 for those whose activities were not*	From 8196 eligible OHCA survivors – 418 survived to discharge (5.1%) and 305 (3.1%) were interviewed at 12-mths post arrest: CPC and HUI3 *Completion rates: 305/418 of those who survived to discharge (72.9%)* Interview completion time: HUI3 approx 10mins (actual from this study?) CPC not reported *Interpretation:* *Authors suggest that CPC can be used as a gross indicator of functional outcome – noting that it does not discriminate well at the high ends of neuro function. It differentiates better between those with no/mild impairment and those with mod/severe.*	CPC and HUI3: Agreement (ICC) between HUI3 (when categorised as <0.34/ 0.34to0.66/ >0.66) and CPC (1/2/3) = 0.51 *Moderate – suggests measure related but different constructs* ROC analysis: ability of HUI3 (scores >/<0.80) to discriminate between patients categorised by CPC scores (1 vs 2/3): HUI>0.80: sensitivity 100%; specificity 27.1% *(when CPC 2/3 HUI3 unlikely to be high)* HUI3<0.40: sensitivity 56%; specificity 97% *(when CPC was 1, HUI3 unlikely to be low).*	Specific comparison of CPC and HUI3 COSMIN: Hypothesis testing: 'Fair' – hypotheses assumed. KGV: 'Fair'

				limited by chronic disease)	*CPC is NOT a substitute for HUI3 (which provides a more detailed assessment of health)*	*Evidence that HUI3 can discriminate between patients when grouped per CPC 1 vs 2/3*	
45	Smith et al (2015) Australia	Prospective cohort study – long-term follow up of adults who survive OHCA. Identified from CA Registry (x-sectional evaluation) *Secondary focus PROM evaluation*	687 (80.7%) 50 interview 157 proxy 59.1 (14.9) (range 18-75+) Male 78.2%	GOS-E 55.6% had GOS-E scores >7 (41.1% if those who died post-discharge included)	Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) (80.7% interviews at 1yr; majority of non-responders were 'lost to follow-up')	More descriptive. Not specific to evaluating measurement / practical properties. Supports feasibility of collecting measures at 1-year by telephone.	Not specific to MP/PP *COSMIN: NR – focus not Psych evaluation*
				SF-12v1 PCS 53.0 (10.2) MCS 46.1 (11.2)	PCS (53.0 (10.2))) and MCS (46.1 (11.2)) equivalent to Australian population norms at 12-mths (PCS 46.1 (11.2); PCS 53.1 (21.8))		
				EQ-5D-3L 0.82 (0.19) *(population nom 0.81 (0.34))*	Profile: >30% reported no problems in all five domains. *** Ceiling effect? **Data quality not detailed.* No problems: Self-care (87.6%); Pain (71.7%); usual activities (67.8%); Mobility (66.4%); Anxiety (66.2%) *(therefore > 1/3 reported some or extreme problems – but unable to unpack from co--morbidities)* Index score 0.82 (SD 0.19) – compares with age and sex matched UK values.		
49	Tiainen et al (2015) Finland	Cross-sectional cohort – sub-study of RCT (OHCA-VF survivors on strict vs moderate glucose control) All patients alive at 6/12 were contacted and invited for a follow-up visit. *Secondary focus PROM evaluation*	57 59.0 (range 24-78yrs) Male 77%	'QOL': 15D 0.883 (population 0.904) 'Functional outcome': CPC mRS median 1 (IQR 0-2) – suggests no significant disability despite symptoms BI (full possible range 0-100) median 100 (100-100) – suggests fully independent 'Good outcome': BI 52/57 (91%) MRS 50/57 (80%)	Face-to-face or telephone interview and standard neurological examination with survivors at 6-mths post CA (n=49/57): 85.6% response rate. Cites evidence of 'testing' of the 15D in general population, stroke and coronary artery disease population. But does not report this evidence in this paper. Reports MCSD as 0.03 (general population).	*No a priori hypotheses* 15D with CPC -0.425 15D with mRS -0.574 15D Utility index score: no difference between survivors and general population (0.883 vs 0.904). NSS 15D Profile score: Statistically significantly lower scores for 2/15 domains - Usual Activities and Sexual Activities – for CA survivors. When grouped per: mRS scores (0 and 1 / 0 and 2 / 0 and 1-2) or CPC (1 or 2): statistically significant between group difference on 15D scores. mRS 0 and 1: median 0.952 vs 0.851 (p=0.012) mRS 0 and 2: median 0.952 vs 0.730 (p=0.003)	Poor evidence of construct validity *COSMIN: Hypothesis testing: Poor (not clear re what was expected re association between variables).* *KGV: Fair*

				Cognitive outcome – neuropsychological exam		mRS 0 and 1-2: median 0.952 vs 0.840 (p<0.001) CPC 1 and 2: median 0.939 vs 0.824 (p=0.017) *But 15D scores did NOT differentiate between cognitively intact (0.952) and those with mild to moderate deficit (0.855)(p=0.323).*	
47	Wallin et al (2014) Sweden	Prospective cohort: survivors of OHCA (67%) and IHCA followed from ICU discharge to 1 and 6mths post arrest. *Secondary focus PROM evaluation*	45 64 (13)(range 24-85yrs) Males 29 (64%)	LiSat-11 - Life Satisfaction Questionnaire Also: Barthel Index MMSE CPC	Interview administration of questionnaires at 1 and 6-mths post OHCA. Score distribution reported for BI, MMSE and CPC; not for LiSat.	Known groups validity (no a priori hypothesis): Gender differences for 'psychological health' (p=0.007) where more men than women were satisfied with their health. All other domains NS. Cronbach's alpha: 0.86 at scale level. (No detail provided).	*COSMIN:* *KGV: Poor (missing data /data quality not reported)* *Internal Consistency Reliability: Poor (no structural validity; no missing data)*
48	Wilson et al (2014) UK	Exploratory study – to investigate psychosocial outcomes of OHCA survivors and impact of anoxic brain injury *Secondary focus PROM evaluation*	56 (27 with anoxia; 29 without) 66.13 (12.61) (range 37 to 84 yrs). Male 37 (66%)	Quality of Life Scale (QOLS) Also: Social Functioning Questionnaire (SFQ) HADS Impact of Event Scale – Revised (IES-R) Everyday Memory Questionnaire – Revised (EMQ-R)	Survivors recruited between 6/12 and 4-yrs post OHCA (mean 25 to 27 months). Postal self-completion of questionnaires. Categorization of groups (anoxia vs non-anoxia) based on documentation of clinical decisions at time of OHCA.	Known groups validity (a priori hypothesised association proposed – but rejected by findings): NS difference in QOLS between anoxia and non-anoxia group.	*COSMIN:* *KGV: Poor (missing data /data quality not reported; small numbers)*
	Authors (Country)	Study design and setting	Participants (n, age and gender)	Measures (Mean (SD))	Study detail	Measurement/ Practical property assessed	Study quality

Appendix 3: Data extraction - measurement and practical properties for HRQOL measures (n=10) evaluated following completion by survivors of CA

Table 3.1 Condition-specific Profile measure (brain injury): QOLIBRI (n=2 studies)

PROM	***Studies (n)***	***Evidence following evaluation in CA population [study](findings)***	***Quality of reported measurement property[a]***	***Methodological quality per reported measurement property (COSMIN)[b]***
Measurement properties				
Reliability	0	No published evaluations: internal consistency; test-retest; measurement error		
Validity				
Content validity	0	No published evaluations (see content comparison table)		
Structural validity	0	No published evaluations		
Construct validity – Convergent / divergent	2	Middlekamp 2007[32]: Retrospective cohort study – assessment of functioning and QOL in patients with post CA hypoxic brain injury - followed-up 2-7 yrs (Small sample n=20): *Association between variables explored (No a priori hypotheses stated – can be assumed):* QOLIBRI and: Coma duration - part I(Satisfaction) r=-0.70 *(suggesting a worse QOL for those with longer duration of coma);* part II (Problems and complaints) r=0.46 QOLIBRI and: Post-traumatic amnesia:r=-0.70 (*indicating lower levels of QOL for those with longer duration of PT amnesia)* QOLIBRI parts I and II with: Cognitive Failure Questionnaire: r=-0.86 (part I); r=-0.77 (part II) Impact on Participation and Autonomy Questionnaire: domains range (part I) -0.44 to -0.86; (part II) 0.53 to 0.73 Frenchay Activities Index: r=0.34 (part I) and -0.41 (part II)	+	Poor
		Mak 2016[33]: Cross-sectional evaluation of CA survivors (n=59 at more than 6/12 post arrest): Association between QOLIBRI and a 'Structured CPC' (enhanced content): -0.67 (larger then hypothesized – authors suggest due to the improved content of the 'structured CPC').	+	Good
Construct validity – Known-groups (Discriminant)	0	No published evaluations		
Responsiveness	0	No published evaluations		

Interpretability *Completion rates (missing data*	1	Middlekamp 2007[32]: Functioning and QOL in patients with post CA hypoxic brain injury - followed-up 2-7 yrs (Small sample n=20): Domain scores – median (25-75%) *(no item level data): No end effects* 1. Satisfaction (possible 42-210): 126.0 (90-165.0) 2. Problems and complaints (possible 14-70): 29.0 (20.5-33.8) *scores suggest some QOL limitations*	+ 	Poor (No MIC or MID – otherwise 'Good')
Practical properties				
Acceptability	0	No published evaluations		
Feasibility	0	No published evaluations		
Patient/ Public Involvement in PROM development/ evaluation				
	0	No evidence of active collaboration		

Table 2. Generic profile - SF-36 (total studies n= 10; Version 1 (v1) = 8; Version 2 (v2) = 2)

PROM	***Studies (n)***	***Evidence following evaluation in CA population [study](findings)***	***Quality of reported measurement property[a]***	***Methodological quality per reported measurement property (COSMIN)[b]***
Measurement properties				
Reliability *Internal consistency*	V1: 2	Graf 2008 [34] Cross-sectional evaluation at 5-years post-ICU discharge (patient-level assessment of health status and fully costed economic evaluation): Postal self-completion of questionnaire including SF-36 (German (V1)) at 5-years (n= 81). Cronbach's alpha >0.7 for all domains (data not presented)	+ 	Poor
		Bro-Jeppesen 2009 [35] Prospective cohort study of comatose OHCA survivors: aim to assess the long-term outcome of OHCA survivors – assessment of cognitive function and QoL before and after implementation of TH. Follow-up interviews at 6-mths: completion of SF-36 and MMSE (n=156) Cronbach's alpha for all domains: range 0.88 to 0.91 (data not presented)	+ 	Poor
Test-retest reliability or measurement error	0	No published evaluations		
Validity				
Content validity	0	No published evaluation (see content comparison table)		
Structural validity	0	No published evaluations		
Construct validity – Convergent	V1: 3	Moulaert 2010[36]: Retrospective cohort study of CA survivors – to study factors related to QoL after hypoxic period due to CA. Postal self-completion at 36-mths post CA (n=63)(72% response rate): *Association between variables explored (No a priori hypotheses stated – can be assumed):* SF-36 PCS with included measures range: 0.22 (Barthel Index) to 0.55 (Cognitive Failures Questionnaire) and 0.61 (Fatigue Severity Scale); MCS range 0.10 (Barthel Index) and 0.57 (Fatigue Severity Scale) and 0.77 (Hospital Anxiety and Depression Scale).	PCS + MCS +	Fair (to poor)
		Kowalik 2014 [38]: Retrospective, observational study: comparison between survivors treated with mild therapeutic hypothermia (MTH)(28) versus historical controls (did not receive MTH)(37). Telephone administration at between 1 and 4 1/2years post-OHCA survival (postal administration for those non-contactable by telephone). *Association between variables explored (No a priori hypotheses stated; can be assumed) (*all analyses not illustrated; missing data not reported):	+	Poor

		SF-36 Vt with Disability Rating Scale (DRS) r=0.405 SF-36 SF with DRS r= 0.391 SF-36 GH with DRS r= 0.351		
		Boyce-van der Wal (2015) [40]: Prospective cohort study – consecutive OHCA survivors referred for cardiac rehabilitation (n=77). Interview administration at 4/52 post-OHCA. *Association between variables (measures of cognitive functioning, participation and autonomy, and anxiety and depression) explored (No a priori hypotheses stated; can be assumed)(only statistically significant associations reported):* SF-36 SF with MMSE r=0.317 SF-36 RE with CFQ r=-0.400 SF-36 BP with CFQ r= -0.366 SF-36 MS with CFQ r=-0.351 SF-36 Vt with CFQ r= -0.250 SF-36 SF with IQCODE r=-0.412 SF-36 Vt with IQCODE r=-0.332	+	Fair
	V2: 1	Orbo 2016 [42]: Prospective longitudinal study of OHCA survivors – investigate cognitive recovery from 3 to 12-mths post resuscitation (n=33). Face-to-face interviews at 3 and 12-months. *Association between variables (no a priori hypothesis (but can be assumed)):* SF-36 MCS with: HADS A -0.82; HADS D -0.84 SF-36 PCS 0.49 Return to work 0.40 Cognitive composites range: 0.03 (Visual memory) to 0.38 (Executive composite) SF-36 PCS with: HADS a -0.35; HADS D -0.57 SF-36 MCS 0.49 Return to work 0.35 Cognitive composites range: 0.28 (Visual memory) to 0.43 (Visual Memory composite)	+ +	Fair Fair
Construct validity – Known-groups (Discriminant)	V1: 7	Graff 2008 [34] Cross-sectional evaluation at 5-years post-ICU discharge (patient-level assessment of health status and fully costed economic evaluation): Postal self-completion of questionnaire including SF-36 (German (V1)) at 5-years (n= 81) *Known-groups validity (no a priori hypothesis (but can be assumed)):* SF-36 domain scores compared to gender and age-matched German control (norm population): 6/8 SF-domains lower than population norm values (Not Pain or Emotional Role). Statistical significance not reported.	+	Poor

		Bro-Jeppesen 2009 [35] Prospective cohort study of comatose OHCA survivors: aim to assess the long-term outcome of OHCA survivors – assessment of cognitive function and QoL before and after implementation of TH. Follow-up interviews at 6-mths: completion of SF-36 and MMSE (n=156) *Known-groups validity (no a priori hypothesis (but can be assumed)):* Non-statistically significant difference between Control (No TH (n77)) and Intervention (TH (n79)) groups at 6/12 Sub-domains RP and RE were lower in the Control group (not NSS). RP and RE lower than Dutch population norms ($p<0.05$).	-	Poor
		Moulaert 2010[36]: Retrospective cohort study of CA survivors – to study factors related to QoL after hypoxic period due to CA. Postal self-completion at 36-mths post CA (n=63)(72% response rate): *Known-groups validity (no a priori hypotheses stated (statistical significance or MID not reported))* Median score: PCS 71.8; MCS 73.0 (Dutch norm 76.0 and 78.0) *Scores lower than population norm – but statistical significance or MID not evaluated (approx. 0.5 SD (so could equate to MID?))* Profile not reported	+	Fair
		Reinhard 2013 [37]: Postal self-completion between 16 and 62-months post OHCA (n44): *Known-groups validity (no a priori hypotheses stated) (statistical significance or MID not reported):* SF-36 profile - OHCA survivors rated health significantly worse than general population for 5/8 domains: general health, social functioning, emotional role functioning, physical role functioning and physical health. NSS difference for vitality, mental health or body pain. NSS difference across all domains for OHCA survivors with known cardiovascular disease (n30) and matched patients with MI or angina (not requiring resuscitation).	+	Poor
		Kowalik 2014 [38]: Retrospective, observational study: comparison between survivors treated with mild therapeutic hypothermia (MTH)(28) versus historical controls (did not receive MTH)(37). Telephone administration at between 1 and 4 1/2years post-OHCA survival (postal administration for those non-contactable by telephone). *Known-groups validity (no a priori hypothesis):* Statistically significance between group difference for survivors in receipt of MTH versus No MTH: 2/8 domains only: Role Limitations due to Emotional problems and Vitality were better in survivors who received MTH ($p<0.05$).	-	Poor
		Moulaert 2015 [39]: RCT rehabilitation post CA (143); assessments at 2-weeks, 3/12 and 12/12. *Known-groups validity (no a priori hypotheses stated – can be assumed):* Statistically significant between group difference at 12-months (in favour of the intervention) for: SF-36 role emotional, mental health and general health; also evidenced for HADS total and Anxiety domain. *No between group difference at 3/12. Does this provide support for longer-term follow-up? The results suggest there is potential for improvement at 12/12 (however, does not include an interim of 6 or 9-mths)*	+	Fair
		Boyce-van der Wal (2015)[40]: Prospective cohort study – consecutive OHCA survivors referred for cardiac rehabilitation (n=77). Interview administration at 4/52 post-OHCA.	+	Fair

		Known-groups validity (no a priori hypothesis (but can be assumed)): Statistically significant difference between cognitively impaired and non-impaired groups across 6/8 domains of the SF-36 (not BP or GH).		
	V2: 1	Cronberg 2015[41] TTM trial *(No specific PROM evaluation; No a priori hypotheses)* For two arms of the trial (n 455 and n 491 at 6/12 follow-up): Mean (SD) MCS 49.1 (12.5) and 49.0 (12.2) (p =0.79); Mean (SD) PCS 46.8 (13.8) and 47.5 (13.8) (p = 0.45): '*scores comparable to the population norm'*	-	Fair
Responsiveness	V2: 1	Orbo 2016 [42]: Prospective longitudinal study of OHCA survivors – investigate cognitive recovery from 3 to 12-mths post resuscitation. N=33 (31 Males; mean age 58.6 years (SD13)). Face-to-face interviews at 3 and 12-months. Effect size statistics for mean change from 3 to 12-mths (no external anchors or variables to determine direction of change; assumed improvement?): SF-36 MCS -0.38; PCS 0.21 HADS A 0.27; D 0.35 Statistically significant (paired t-test) reduction in MCS between 3 and 12/12mth post arrest (p=0.02); NS improvement for PCS T1 (3/12): MCS 51.85 (9.57); PCS 44.10 (8.55) T2 (12/12): MCS 47.71 (11.82); PCS 45.82 (8.04). NS change at domain level (data not illustrated).		N/A N/A
Interpretability *Completion rates (missing data)*	V1: 1	Moulaert 2010[36]: 2/63 patients had >15% missing items for SF-36 (same as FAI, CIQ, CFQ, FSS).		
	V2: 2	Cronberg 2015[41]: In-person (or telephone) interview (n455)(proxy 8%) completion equivalent to 'Two Simple Questions' rate – and higher than several other measures (MMSE, IQCODE)		
		Orbo 2016 [42]: Prospective longitudinal study of OHCA survivors – investigate cognitive recovery from 3 to 12-mths post resuscitation. N=33 (31 Males; mean age 58.6 years (SD13)). Face-to-face interviews at 3 and 12-months. Reliable change index: (as a measure of the clinical significance of change scores): cut-off point of 40 used for the SF-36 summary scores: between 3 and 12/12: Deterioration: MCS 5 (15%) of patients; PCS 0. Improvement: MCS 0; PCS 4 (12%). No change: MCS 28 (86%); PCS 29 (88%)		N/A
Practical properties				
Acceptability	0	No published evaluations		

Feasibility	0	No published evaluations		
Patient/ Public Involvement in PROM development/ evaluation				
	0	No evidence of active collaboration		

Table 3. Generic profile - SF-12 (Version 2) (total studies n=4)

PROM	***Studies (n)***	***Evidence following evaluation in CA population [study](findings)***	***Quality of reported measurement property[a]***	***Methodological quality per reported measurement property (COSMIN)[b]***
Measurement properties				
Reliability	0	No published evaluations: internal consistency; test-retest; measurement error		
Validity				
Content validity	0	No published evaluation (see content comparison table)		
Structural validity	0	No published evaluations		
Construct validity – Convergent	1	Andrew 2016 [46] Prospective cohort - Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1188/1486 (79% of those known to be alive at 12/12) responded to interviews. Association between variables *(No a priori hypotheses stated – can be assumed)* PCS with: SF-6D 0.69 EQ-5D-3L 0.63 GOSE 0.440 MCS 0.036 MCS with: SF-6D 0.558 EQ-5D-3L 0.29 GOSE 0.314 PCS 0.036	+	Fair
Construct validity – Discriminant	4	Deasy 2012 [43]: Telephone administration at 12-mths post CA (n56): *Known-groups validity (No a priori hypotheses stated; statistical significance or MID not reported):* Reports lower MCS scores (38.0) when compared to Australian population (NR); PCS 45.6 (scores <40 = mod to severe disability).	+	Fair
		Beesems 2014 [44]: Telephone interview with 220 survivors 6-12 mths post CA: just 45% had 'normal physical health' (fewer than 50% had 'good' function') and 90% had 'normal mental health' on SF-12. *Known-groups validity (No a priori hypotheses stated; statistical significance or MID not reported):* PCS scores lower than general population (<60yrs 45.8 (9.0); 60-80 46.7 (9.6); >80 40.5 (9.6)); and equivalent SF-12 MCS scores (<60yrs 52.2 (8.6); 60-80 54.3 (8.0); >80yrs 53.2 (10.4)). (see Table 4) SF-12 PCS scores were statistically significantly better for the younger survivors (<80yrs) than for the older survivors (>80yrs)(p=0.003)	+	Fair

		But for the older group, these physical limitations seem mainly related to advanced age and co-morbidities as opposed to the CA itself (see referent values).		
		Smith 2015 [45]: *T*elephone administration at 12-mths post CA (n=687): *Known-groups validity (No a priori hypotheses stated; statistical significance or MID not reported):* PCS (lower) and MCS (similar) equivalent to Australian population norms at 12-mths.	+	Fair
		Andrew 2016 [46] Prospective cohort - Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1188/1486 (79% of those known to be alive at 12/12) responded to interviews. Known-groups (*No a priori hypotheses stated – can be assumed):* SF-12 MCS Gender: Male 57 (52-59); Female 56 (48-59) (p=0.051) Discharged home: Yes 57 (52-59); No 54 (48-60) (p=0.14) Return to Work: Yes 56 (53-59); No 53 (44-58) (p<0.001) SF-12 PCS Gender: Male 51 (41-56); Female 45 (33-53) (p<0.001) Discharged home: Yes 51 (40-56); No 44 (37-55) (p=0.003) Return to Work: Yes 54 (48-57); No 45 (36-54) (p<0.001)	 + +	 Fair Fair
Responsiveness	0	No published evaluations		
Interpretability *Completion rates (missing data)*	2	Deasy 2012 [43]: Item level frequency endorsement and score data. No end effects. 'Good' telephone interview completion rates; but 'challenges with loss to follow-up'		
		Smith 2015 [45]: telephone administration at 1-yr in CA survivors (who survived to hospital discharge): 80.7% interviews at 1yr; majority of non-responders were 'lost to follow-up'		
		Andrew 2016 [46] Prospective cohort - Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1188/1486 (79% of those known to be alive at 12/12) responded to interviews. SF-12 v1 (Median (IQR)) MCS 56 (52 – 59) (full range 11-70) PCS 50 (40 – 56) (full range 11-70) No end effects (fig 1) Missing items at scale level: 3.4% (34/1188). No end effects		Good / (Poor - No MIC or MID – otherwise 'Good')
Practical properties				
Acceptability	0	No published evaluations		

Feasibility	0	No published evaluations		
Patient/ Public Involvement in PROM development/ evaluation				
	0	No evidence of active collaboration		

Table 4. Generic profile / index – Life Satisfaction Checklist (LiSat-11) (total studies n=1)

PROM	***Studies (n)***	***Evidence following evaluation in CA population [study](findings)***	***Quality of reported measurement property[a]***	***Methodological quality per reported measurement property (COSMIN)[b]***
Measurement properties				
Reliability *Internal consistency*	1	Wallin 2014 [47]: Prospective cohort: n=45 Survivors of OHCA (67%) and IHCA followed from ICU discharge to 1 and 6mths post arrest. Interview administration of questionnaires at 1 and 6-mths post OHCA: includes LiSat-11, Barthel Index, MMSE and CPC. Cronbach's alpha 0.86 (no additional detail provided). Only reported at scale/index level	+ Poor
Test-retest reliability or measurement error	0	No published evaluations		
Validity				
Content validity	0	No published evaluation (see content comparison table)		
Structural validity	0	No published evaluations		
Construct validity – Convergent	0	No published evaluation		
Construct validity – Discriminant	1	Wallin 2014 [47]: Prospective cohort: Survivors of OHCA (67%) and IHCA followed from ICU discharge to 1 and 6mths post arrest. Interview administration at 1 and 6-mths post-arrest (n=45): *Known groups validity (no a priori hypothesis; cannot be assumed)*: Gender differences for 'psychological health' (p=0.007) where more men than women were satisfied with their health. All other domains NS.	- Poor
Responsiveness	0	No published evaluation		
Interpretability *Completion rates (missing data)*	0	No published evaluation		
Practical properties				
Acceptability	0	No published evaluations		

Feasibility	0	No published evaluations		
Patient/ Public Involvement in PROM development/ evaluation				
	0	No evidence of active collaboration		

Table 5. Generic profile / index – Quality of Life Scale (total studies n=1)

PROM	*Studies (n)*	*Evidence following evaluation in CA population [study](findings)*	*Quality of reported measurement property*[a]	*Methodological quality per reported measurement property (COSMIN)*[b]
Measurement properties				
Reliability	0	No published evaluations: internal consistency; test-retest; measurement error		
Validity				
Content validity	0	No published evaluation (see content comparison table)		
Structural validity	0	No published evaluations		
Construct validity – Convergent	0	No published evaluations		
Construct validity – Discriminant	1	Wilson 2014 [48]: Exploratory study – to investigate psychosocial outcomes of OHCA survivors and impact of anoxic brain injury. Survivors recruited between 6/12 and 4-yrs post OHCA (mean 25 to 27 months). 56 (27 with anoxia; 29 without). Categorization of groups (anoxia vs non-anoxia) based on documentation of clinical decisions at time of OHCA. Postal self-completion of questionnaires. *Known groups validity (a priori hypothesised association proposed – but rejected by findings):* NS difference in QOLS between anoxia and non-anoxia group.	-	Poor
Responsiveness	0	No published evaluation		
Interpretability *Completion rates (missing data)*	0	No published evaluation		
Practical properties				
Acceptability	0	No published evaluations		
Feasibility	0	No published evaluations		
Patient/ Public Involvement in PROM development/ evaluation				
	0	No evidence of active collaboration		

Table 6. Generic Preference-based Utility measure – 15-Dimensions Quality of Life Questionnaire (15D) (total studies n=1).

PROM	***Studies (n)***	***Evidence following evaluation in CA population [study](findings)***	***[a]Quality of reported measurement property[a]***	***[b] Methodological quality per reported measurement property (COSMIN)[b]***
Measurement properties				
Reliability	0	No published evaluations: test-retest; measurement error		
Validity				
Content validity	0	No published evaluations (see content comparison table)		
Structural validity	0	Not relevant		
Construct validity – Convergent	1	Tiainen 2015 [49]: Cross-sectional cohort – sub-study of RCT: Face-to-face or telephone interview with survivors at 6-mths post CA (n=49/57): Association between variables reported *(No a priori hypotheses stated – expected association unclear):* 15D with CPC -0.425 15D with mRS -0.574	? +	Poor – *(unclear what was expected from associations; no info re MP of comparator measures etc; to Fair for KGV*)
Construct validity – Discriminant	1	Tiainen 2015 [49]: Face-to-face or telephone interview with survivors at 6-mths post CA (n=57) (86% response rate): Known-groups (*No a priori hypotheses stated – can be assumed):* Utility index score: no difference between survivors and general population (0.883 vs 0.904). NSS Profile score: Statistically significantly lower scores for 2/15 domains - Usual Activities and Sexual Activities – for CA survivors.	?	N/A (Poor)
		Tiainen 2015 [49]: Face-to-face or telephone interview with survivors at 6-mths post CA (n=57) (86% response rate): Known-groups (*No a priori hypotheses stated – can be assumed):* When grouped per: mRS scores (0 and 1 / 0 and 2 / 0 and 1-2) or CPC (1 or 2): statistically significant between group difference on 15D scores. mRS 0 and 1: median 0.952 vs 0.851 (p=0.012); mRS 0 and 2: median 0.952 vs 0.730 (p=0.003); mRS 0 and 1-2: median 0.952 vs 0.840 (p<0.001) CPC 1 and 2: median 0.939 vs 0.824 (p=0.017) *But 15D scores did NOT differentiate between cognitively intact (0.952) and those with mild to moderate deficit (0.855)(p=0.323).*	+	Fair

Responsiveness	0	No published evaluations		
Interpretability *Completion rates (missing data)*		Tiainen 2015 [49]: Face-to-face or telephone interview with survivors at 6-mths post CA (n=49/57): 85.6% response rate.		
Practical properties				
Acceptability	0	No published evaluations		
Feasibility	0	No published evaluations		
Patient/ Public Involvement in PROM development/ evaluation				
	0	No evidence of active collaboration		

Table 7. Generic preference-based Utility measure – EuroQoL EQ-5D-3L (total studies n= 3)

PROM	***Studies (n)***	***Evidence following evaluation in CA population [study](findings)***	***Quality of reported measurement property[a]***	***Methodological quality per reported measurement property (COSMIN)[b]***
Measurement properties				
Reliability	0	No published evaluations: test-retest or measurement error		
Validity				
Content validity	0	No published evaluations (see content comparison table)		
Structural validity	0	Not relevant		
Construct validity – Convergent	1	Andrew 2016 [46] Prospective cohort - Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1188/1486 (79% of those known to be alive at 12/12) responded to interviews. Association between variables *(No a priori hypotheses stated – can be assumed)* Convergent validity: Patients: EQ-5D-3L with: SF-6D 0.65 SF-12 PCS 0.62 SF-12 MCS 0.29 GOSE 0.47 Proxy: EQ-5D-3L with: GOS-E 0.67	+/-	Fair
Construct validity – Discriminant	1	Andrew 2016 [46] Prospective cohort - Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1188/1486 (79% of those known to be alive at 12/12) responded to interviews. Known-groups (*No a priori hypotheses stated – can be assumed):* Statistically significant difference between patients (0.85) and proxy (0.75)(p<0.001) Gender: Male 0.88 (0.78-1.0); Female 0..85 (0.73-1.0) (p<0.001) Discharged home: Yes 0.88 (0.7-1.0); No 0.85 (0.71-1.0) (p=0.09) Return to Work: Yes 1.00 (0.85-1.0); No 0.85 (0.69-1.0) (p<0.001)	+ +	Fair
Responsiveness	0	No published evaluation		
Interpretability	3	Smith 2015 [45]: Telephone administration at 1-yr in CA survivors (who survived to hospital discharge): n=687 (80.7%) interviews at 1yr. Profile: >30% reported no problems in all five domains. *** Ceiling effect? **Data quality not detailed.*		

Data quality – end-effects Completion rates (missing data)		No problems: Self-care (87.6%); Pain (71.7%); usual activities (67.8%); Mobility (66.4%); Anxiety (66.2%) *(therefore > 1/3 reported some or extreme problems – but unable to unpack from co--morbidities)* Index score 0.82 (SD 0.19) – compares with age and sex matched UK values.		
		Deasy 2012 [43]: Reports item level (frequency endorsement) only: >70% report NO problems with mobility, self-care, pain/discomfort; *** Ceiling effect? **Data quality not detailed.* >30% report problems with usual activities and >60% with anxiety/depression		
		Andrew 2016 [46] Prospective cohort - Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1188/1486 (79% of those known to be alive at 12/12) responded to interviews. Missing items at scale level: Total - 1.0% (12/1188); Patients – 0.5% (5/928); Proxy – 2.7% (7/260) Large ceiling effects: Patients 46% (n421); Proxy 23% For those patients reporting full health (n421), substantial variability on SF-6D and GOSE – patients more likely to report problems with mental health and vitality Comparison of score distribution for EQ-5D-3L and SF-6D against GOSE scores (fig 2): demonstrate a statistically significant difference in score distribution for patients reporting a lower or upper moderate GOSE and lower or upper good GOSE ($p<0.001$) For EQ-5D-3L: 75th percentile is equal to a score of 1.00 for all GOSE categories above lower moderate disability *(strong evidence of significant ceiling effect)*	+	Good / (Poor - No MIC or MID – otherwise 'Good')
Practical properties				
Acceptability	0	No published evaluations		
Feasibility	0	No published evaluations		
Patient/ Public Involvement in PROM development/ evaluation				
	0	No evidence of active collaboration		

Table 8. Generic preference-based utility measure - HUI3 (total studies n= 7)

PROM	*Studies (n)*	*Evidence following evaluation in CA population [study](findings)*	*Quality of reported measurement property*[a]	*Methodological quality per reported measurement property (COSMIN)*[b]
Measurement properties				
Reliability				
Inter-rater	1	Raina 2008 [52]: Prospective cohort (convenience sample) of 21 adult survivors: Inter-rater reliability: Ten patient interviews (2 raters): HUI3 97.3% (kappa 0.96) *(?? Why not test-rest for self-completion)*	+	Poor
Validity				
Content validity	0	No published evaluation (see content comparison table)		
Structural validity	0	Not relevant		
Construct validity – Convergent	5	Stiell 2009[53]: Comparative evaluation of CPC and HUI3 in prospective cohort of CA survivors (at 1-year). 418 survivors at hospital discharge; 305 (73%) interview administration of CPC and HUI3 at 12-mths. *Association between variables: (No a priori hypotheses stated – can be assumed)* 'Agreement' (ICC) between HUI3 (when categorised as <0.34/0.34to0.66/>0.66) and CPC (1/2/3) = 0.51	+	Fair
		Raina 2008 [52]: Prospective cohort (convenience sample) of 21 adult survivors: Medical chart review at discharge – to inform CPC and MRS. In-person interview at 1/12 to collect mRS and HUI3 (and to inform 1/12 CPC) Association between variables *(No a priori hypotheses stated – can be assumed)* 1/12 HUI3 scores with CPC discharge scores: -0.41 1/12 HUI3 scores with 1/12 CPC scores: -0.71	- +	Poor
		Raina 2015 [55]: Prospective cohort of 29 adult survivors; interview administration at 12/12: Association between variables *(No a priori hypotheses stated – can be assumed)* HUI3 with: Measures of 'global disability': CPC -0.35; GOS-E 0.45; MRS -0.48 Geriatric Depression Scale (GDS) -0.45; MMSE 0.23 Telephone interview of cognitive status (TICS) 0.37 Performance Assessment of Self-care Skills (PASS) – Habit 0.40; Skill 0.66; Participation: Reintegration to Normal Living Index (RNLI) 0.54	-/+	Poor (small n)
		Nichol 2015 [56]: Telephone administration at 3 and 6-mths post discharge (n644) mRS at discharge independently associated with post-discharge variables: ALFI -1.30 (-1.64, -0.96); HUI3 -0.07 (-0.09, -0.05); T-GDS 0.67 (0.41,-0.93)	?	Fair

		Predictive model: Suggests that for every unit change in mRS, this predicts a change of -0.06 in the HUI3. *(If the MCID is 0.03, then the change is greater than the MCID. If the MCID is 1.0, the change is less than the MCID?)*		
		Nichol 1999 [50]: Prospective cohort of CA survivors (n86/96 survivors): HUI3 telephone admin at 6/12. *Association between variables (No a priori hypotheses stated – can be assumed):* HUI3 with CPC (pre-discharge) -0.29; MMSE (pre-discharge) 0.37	?	Good
Construct validity – known-groups (discriminant)	3	Longstretch 2010 [54]: 3-mth post hospital discharge completion (part of RCT): Interview (telephone) administration (n=32): Known-groups (*No a priori hypotheses stated – can be assumed):* When grouped as 'Full Recovery' versus 'Dependent' ('Two Simple Questions'): statistically significant between group difference on HUI3: 0.82 (full recovery) vs 0.30 (dependent).	+	Fair
		Stiell 2009[53]: Comparative evaluation of CPC and HUI3 in prospective cohort of CA survivors (at 1-year). 418 survivors at hospital discharge; 305 (3.7%) interview administration of CPC and HUI3 at 12-mths Known-groups (*No a priori hypotheses stated – can be assumed):* CPC and HUI3: Agreement (ICC) between HUI3 (when categorised as <0.34/ 0.34to0.66/ >0.66) and CPC (1/2/3) = 0.51 ROC analysis: ability of HUI3 (scores >/<0.80) to discriminate between patients categorized by CPC scores (1 vs 2/3): HUI>0.80: sensitivity 100%; specificity 27.1% (when CPC 2/3 HUI3 unlikely to be high) HUI<0.40: sensitivity 56%; specificity 97% (when CPC was 1, HUI3 unlikely to be low). *Evidence that HUI3 can discriminate between patients when grouped per CPC 1 vs 2/3*	+ +	Fair
		Nichol 1999 [50]: Prospective cohort of CA survivors (n86/96 survivors): HUI3 telephone admin at 6/12. Known-groups (*No a priori hypotheses stated – can be assumed):* HUI3 0.72 (0.22): Mean utility significantly worse than Gen pop (0.86 (0.16)) and those whose activities were not limited by chronic disease (0.91 (0.08))(p<0.01) HUI3 score by duration of resuscitation: Shorter (<2mins): 0.81; Moderate (3-10 mins) 0.76; Longer (>10mins) 0.65 (p=0.05)	+ +	Good
Responsiveness	1	Raina 2015 [55]: Prospective cohort (n29); interview administration at follow-up: *No a priori hypothesis; no external anchors; improvement assumed* Change in scores over time: Mean values reported at discharge (CPC and MRS only), 1, 6, 12/12. Non-statistically significant improvement in score for HUI3 6 and 12/12 (1/12 0.59; 6/12 0.68; 12/12 0.66) Moderate ES 1-6/12 (0.43); small 6-12/12 (0.19) and 1-12/12 (0.26).	-	*COSMIN does not accept ES as evidence of Responsiveness*

Interpretation *End effects* *Completion rates (missing data)*		Longstretch 2010 [54]: 3-mth post hospital discharge completion (part of RCT): Interview (telephone) administration – missing data for 8 survivors (3 full recovery; 3 independent; 2 dependent – suggests that missing most often in those not reporting full recovery so HUI3 scores in this sample may overestimate QoL)		
		Stiell 2009 [53]: Prospective cohort sub-study of the OPALS study – Adult OHCA patients from 20 cities (n268/316 known survivors (84%)). Comparative evaluation of CPC and HUI3 in prospective cohort of CA survivors (at 1-year). 418 survivors at hospital discharge; 305 (73%) interview administration of CPC and HUI3 at 12-mth. Completion rates: 305/418 of those who survived to discharge (72.9%) Interview completion time: HUI3 approx 10mins (actual from this study?); CPC not reported Interpretation: Authors suggest that CPC can be used as a gross indicator of functional outcome – noting that it does not discriminate well at the high ends of neuro function. It differentiates better between those with no/mild impairment and those with mod/severe. CPC is NOT a substitute for HUI3 (which provides a more detailed assessment of health) CPC with HUI3: Agreement (ICC) between HUI3 (when categorised as <0.34/ 0.34to0.66/ >0.66) and CPC (1/2/3) = 0.51 = Moderate – suggests measure related but different constructs ROC analysis: ability of HUI3 (scores >/<0.80) to discriminate between patients categorised by CPC scores (1 vs 2/3): HUI>0.80: sensitivity 100%; specificity 27.1% (when CPC 2/3 HUI3 unlikely to be high) HUI3<0.40: sensitivity 56%; specificity 97% (when CPC was 1, HUI3 unlikely to be low). Evidence that HUI3 can discriminate between patients when grouped per CPC 1 vs 2/3	+ +	Fair Fair
		Stiell 2003 [51] Prospective cohort sub-study of the OPALS study – Adult OHCA patients from 20 cities (n268/316 known survivors (84%)). Telephone admin at 1yr. HUI3 scores (0.80) comparable to age-adjusted pop (0.83)(statistical significance not explored). Score distribution: range -0.30 to 1.0 Score >0.80 = 134/316 (42% of known survivors) Score <0.80 = 134/316 (42%) 0.60-0.80 = 56/316 (18%) Score <0.60 = 78 (25%) (Unknown n=48) *So, population with scores >0.60 = n190 = 75%. Scores less than 0.60 = 25%* *HUI3 scores <0.70 suggest severe impairment – suggesting that >25% of population are severely impaired*		COSMIN N/A
Practical properties				
Acceptability	0	No published evaluations		
Feasibility	0	No published evaluations		

Patient/ Public Involvement in PROM development/ evaluation				
	0	No evidence of active collaboration		

Table 9. Generic preference-based utility measure - SF-6D (n= 1 study)

PROM	*Studies (n)*	*Evidence following evaluation in CA population [study](findings)*	*Quality of reported measurement property*[a]	*Methodological quality per reported measurement property (COSMIN)*[b]
Measurement properties				
Reliability	0	No published evaluation: test-retest; measurement error		
Validity				
Content validity	0	No published evaluation (see content comparison table)		
Structural validity	0	Not relevant		
Construct validity – Convergent	1	Andrew 2016 [46] Prospective cohort - Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1188/1486 (79% of those known to be alive at 12/12) responded to interviews. Association between variables *(No a priori hypotheses stated – can be assumed)* SF-6D with: SF-12 PCS 0.69 EQ-5D-3L 0.65 MCS 0.56 GOSE 0.52 SF-6D items with EQ-5D-3L items *(smaller than assumed hypothesized association (negative ratings (<0.50)):* PF range: Anx/Dep 0.22 to Usual activity 0.52 and Mobility 0.51 (+) Role limit range: Self-care 0.20 to UA 0.42 (-) Social function range: Self-care 0.12 to UA 0.33 (-) Pain range: Anx/Dep 0.21 to Pain 0.57 (+) MH range: Self-care 0.12 to Anx/Dep 0.45 (-) Vit range: Self-care 0.2 to Mobility 0.35 (-)	+ +/-	Fair Fair
Construct validity – Discriminant	1	Andrew 2016 [46] Prospective cohort - Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1188/1486 (79% of those known to be alive at 12/12) responded to interviews. Known-groups (*No a priori hypotheses stated – can be assumed):* Gender: Male 0.86 (0.72-0.92); Female 0.79 (0.66-0.86) (p<0.001) Discharged home: Yes 0.86 (0.72-0.92); No 0.79 (0.67-0.86) (p<0.001) Return to Work: Yes 0.86 (0.80-0.92); No 0.74 (0.63-0.86) (p<0.001)	+	Fair
Responsiveness	0	No published evaluation		

Interpretability *Completion rates (missing data)*	1	Andrew 2016 [46] Prospective cohort - Telephone interviews at 12/12 in CA survivors (who survived to hospital discharge) 1188/1486 (79% of those known to be alive at 12/12) responded to interviews. Missing items at scale level: 3.4% (34/1188). 7% full health (full range 0.345 to 1.00). No end effects Comparison of score distribution for EQ-5D-3L and SF-6D against GOSE scores (fig 2): demonstrate a statistically significant difference in score distribution for patients reporting a lower or upper moderate GOSE and lower or upper good GOSE ($p<0.001$) For EQ-5D-3L: 75th percentile is equal to a score of 1.00 for all GOSE categories above lower moderate disability *(strong evidence of significant ceiling effect) – but there is strong evidence of variability in SF-6D score across these categories (in patients reporting full health on the EQ-5D-3L).* *SF-6D scores improved as GOSE scores improved (fig 2)*	+	Good / (Poor - No MIC or MID – otherwise 'Good')
Practical properties				
Acceptability	0	No published evaluations		
Feasibility	0	No published evaluations		
Patient/ Public Involvement in PROM development/ evaluation				
	0	No evidence of active collaboration		