Orthopedic Trauma and Recovery of Quality of Life: An Overview of the Literature

Khalid Hamid, Pramin Raut, Bessam Ahmed and William Eardley

Department of Trauma & Orthopaedics, James Cook University Hospital, Middlesbrough, United Kingdom.

ABSTRACT: Assessment of clinical success by radiographic evidence of fracture union and surgeon-rated performance following recovery are the outcome tools of the past. Patients are now involved in the assessment of both surgeon performance and the capacity of the institutions in which they are treated to provide rehabilitation following injury. This population is increasingly involved in trials to guide most appropriate and cost-effective care. With healthcare resources globally under pressure, research focus on patient-rated outcome per unit expenditure is central to orthopedic evidence-based practice. In this era of patient-focused assessment and healthcare economics, quality of life and alterations in this status are central as outcome measures. In order to quantify the return of quality of life following injury, we present a review of the literature pertaining to this fundamental aspect of orthopedic trauma patient care.

KEYWORDS: quality, outcome, orthopedic, trauma, evaluation

Background
Traditionally a marker for healthcare performance and certainly a tool of previous decades, survival has been the primary outcome when assessing providers of care for orthopedic trauma. However, in addition to the large number of deaths, major trauma also results in large numbers of musculoskeletal disabilities requiring long-term care. Compared to the number of fatalities, the number of survivors of major trauma with serious or permanent injury is doubled. Hence, in recent years, there has been a shift toward a more comprehensive assessment of delivery of trauma care to encompass the overall quality of life (QOL) of trauma patients rather than just their survival.

Such a shift in gear is reflected in the long-term plan set for the United Kingdom National Health Service (NHS), where patient-centered care, a focus on outcomes, accountability, and efficiency are championed. With focus on outcomes in mind, the NHS Outcomes Framework was developed, highlighting the need for an indicator for improving recovery from injuries and trauma.¹

The need for such an indicator for improving recovery from injuries and trauma may be met by Patient Reported Outcome Measures (PROMs). PROMs are "standardized validated instruments (question sets) to measure patients’ perceptions of their health status (impairment), their functional status (disability), and their health-related quality of life (well-being)".² Such tools are useful to monitor the effects of a treatment or service over a period of time. They can be used to assess specific disease conditions or for a generic measure of health outcome. The use of PROMs presents a valuable way of shifting the focus away from the use of mortality and toward a more well-rounded outcome measure, in order to assess the delivery of major trauma care.

Measurement of QOL, of which PROMs are an element, aims to provide an accurate assessment of individuals’ or populations’ health. More importantly, it is central to identifying potential benefits and harm that may result from a clinical intervention. Bhandari and Giannoudis³ introduce the concept in describing that: “Central to practicing evidence based orthopedics involves integration of our clinical expertise and judgment with patients’ perceptions and societal values, and with the best available research evidence.” Early enquiry of the utility of QOL instruments focused little on patients’ perceptions and revealed that disparity exists in clinicians and patients’ judgments on the impact of interventions on healthcare outcomes.⁴ Such work highlights the importance of asking patients views on healthcare interventions and lays down the fundamental tenet that outcomes affecting patients should probably be reported by patients themselves and not clinicians. Looking at the issue of whether or not a health professional can make a valid assessment of a patient’s QOL was investigated in a series of cancer patients by Slevin et al.⁵ Doctors and patients filled out the same forms regarding anxiety and depression scales among others, finding that correlations between the two sets of scores were poor, suggesting
that the doctors could not accurately determine the patients’ feelings. In addition, the reproducibility of these scales demonstrated considerable variability in results between doctors. Slevin et al concluded from this that if a reliable and consistent method of measuring QOL in their patients is required, it must come from the patients themselves and not from their doctors and nurses.5

In addition to the variability in outcome assessment between the clinician and the patient, significant variation is seen within patients themselves when reporting the same health intervention. Psychological factors including health anxiety are the strongest determinants of patient satisfaction, and when patients have a role in medical decisions they are more satisfied with their health care.6

Patient engagement or “activation” is not just important to those undergoing surgery but is also an increasingly important component of strategies to reform health care. In their review of the topic, Hibbard and Greene found that policies and interventions aimed at strengthening patients’ role in managing their health care can contribute to improved outcomes and recommend that patient activation should be measured as an intermediate outcome of care that is linked to improved outcomes.7 The Patient Activation Measure involves patients answering a series of questions gauging a person’s involvement in their health care. Scored on a 0–100 scale patients can be categorized into four levels of activation in response to 13 statements about beliefs, confidence in managing health-related tasks, and self-assessed knowledge.

Furthermore and supporting engagement further, they found that patients who start at the lowest activation levels prior to activation interventions tend to increase the most in terms of satisfaction.7

Alongside engagement, little can be discussed in the modern healthcare landscape regarding outcome without discussing cost implications. Reinforcing the impact of patient involvement in their decision-making and outcome assessment, it has been demonstrated that patients with lower activation are associated with higher healthcare costs. In a study of 33,163 patients of a large healthcare delivery system in Minnesota, Hibbard et al8 found that patient activation was a significant predictor of cost. Overall, in the year following a healthcare episode, patients with the lowest activation levels had predicted average costs that were 21% higher than the costs of patients with the highest activation levels.8

Healthcare outcomes should therefore be assessed by patients and not clinicians although it needs to be remembered that patent reporting is heterogeneous: different patients will report the same episode of care dissimilarly. It is equally evident that the extent to which patients are involved in their care impacts on their experience, influencing outcome reporting. These factors must be borne in mind when considering the use, applicability, and reporting of QOL outcomes in the orthopedic trauma population. The study is an overview of PROMs that may be applied in the orthopedic trauma setting. It cannot be interpreted as a formal systematic review as the contributory literature is too diverse and heterogeneous. It intends to inform the reader of the options available when beginning to assess trauma outcomes.

Methodology
To perform this review, we carried out an electronic search of the Medline database using the PubMed search engine limited to manuscripts published in English until December 2015. The Medical Subject Headings of Quality, orthopedic, measurement and outcomes were used and retrieved citations scanned for relevance. Retrieved manuscripts were reviewed and individual bibliographies were cross-checked for further articles.

Disease- or injury-specific and generic: measuring QOL. Two basic types of instruments are available to assess the QOL: generic, which include health profiles and utility measurements based on patient’s preferences with regard to treatment and outcome; and specific instruments, which focus on problems associated with individual diseases, patient groups, or areas of function. In essence, generic tools are widely generalizable but lack specificity to a disease, injury, or anatomical region.

On the other hand, while the specific instruments are clearly not lacking in specificity, but are markedly limited in their use either by disease or injury process or by body region involved.9

Generally speaking, generic instruments can be classified into elements of physical well-being and the functional and psychological aspects of health. In keeping with previous work, physicians have utilized physical outcomes as a measure of recovery: an outdated practice that carries the disadvantage of ignoring the nonphysical aspects of disease and the impact of multiple psychological inputs into recovery.9

A part of the issue with measuring QOL lies in defining it as an entity. It is a broad term that encompasses health-related factors including the emotional, functional engagement, recovery, and physical-well being of an individual. It also captures the general non-health-related factors such as occupational and social status; hence, it has, until relatively recently, been challenging to establish a patient-centered QOL measurement tool that addresses all the areas that impact recovery and perceptions of care.10,11

The difficulty in covering the domains required for meaningful generic assessment is that they are largely subjective. While recently there has been an increase in the use of generic tools for orthopedic trials,12 including their use by proxy, it is difficult to tease out which factors are truly impacting. This decreases the applicability of such generic findings and can fail to identify true differences between populations.

Generic measures.

SF-36. The Medical Outcome Study Short Form Health Survey (SF-36) is a 36-item PROM, which focuses on eight
domains, each comprised of several questions. For example, the questions in the physical domain attempt to ascertain the degree of limitation caused by the individual’s health on activities like walking and climbing stairs, using three responses: “not limited at all,” “limited a little,” and “limited a lot”. For the more general elements, five responses are offered: “excellent,” “very good,” “good,” “fair,” and “poor”. The responses to all of the questions within a domain are used to generate a score (0–100) for the construct. The scores are calibrated so that 50 can be taken as the average response allowing for comparison of an individual to the “norm.”

The SF-36 has been shown to be reliable when measured using consistency and test–retest methods. The SF-36 has also been validated as being responsive to health status over time in injured adults, and has been recommended as a generic QOL tool for all trauma patients. One study examining the extent to which different tools evaluate health outcomes in major trauma patients found SF-36 to be the most frequently used outcome measure. There are variations to the survey such as the SF-12: this is a shorter derivative of the SF-36 but was found to be equally responsive to change in health status of major trauma patients.

One criticism of the SF-36 arises from the assumption that the intervals between all the graded responses are the same, such as that between

- “a little limited” and “limited a lot”; and between
- “not limited at all” and “a little limited.”

Additionally, it assumes that each item within the construct, such as walking and climbing stairs, is of equal importance to the individual. Such simplification will result in reduced sensitivity to change in the overall quality of an individual life and also within a single construct.

EQ-5D. Like the SF-36, the European Quality of Life (EQ-5D) is a multi-construct generic PROM; however, it focuses only on five constructs: mobility, self-care, usual activities, pain, and depression. Each of these five constructs has three responses: no problems, some problems, and severe problems, accumulating in a total of 245 possible health states.

In addition, the EQ-5D includes a Visual Analog Scale from 0 (worst imaginable health) to 100 (best imaginable health) to ascertain the patient’s perspective of their overall health status.

The EQ-5D is a validated PROM that is widely used in various clinical conditions including trauma. Since 2009, it has been offered to all patients in the NHS who have undergone a hip replacement, knee replacement, varicose vein surgery, or hernia repair. In addition, a consensus of multi-professional and multidisciplinary expert groups, using a systematic review carried out by the Trauma Audit and Research Network and the Cochrane injuries group, recommended the use of the EQ-5D as a long-term outcome indicator in trauma patients. The reason for such popularity of the EQ-5D in the NHS is simple. Using a random sample of 3,000 people in the UK and employing time trade-off and VAS methods, utility scores were derived for each of the 245 possible health states, which can be expressed as quality-adjusted life-years (QALYs). According to the recommendation by the National Institute for Health Care Excellence (NICE), all health effects should be expressed in QALYs; the EQ-5D is furthermore the preferred measure of health-related QOL in adults. Hence, in addition to monitoring long-term health outcome, the EQ-5D proves to be very useful in health economics.

A common criticism of this measure is the lack of sensitivity to change afforded only by three levels of responses available within each construct. In order to address this issue, a version of the measure with five responses has been developed, called the EQ-5D-5L. A set of utility values for the 3125 EQ-5D-5L states has been assigned using the existing EQ-5D (3L) value sets; however, such a process has its limitations and does not accurately reflect the general public’s preferences. In addition, the extra levels of response add a dimension of complexity to the interpretation of the levels, such as “severe” and “extreme” problems, further adding to the complexity of interpretation.

Trauma outcome profile. The Trauma Outcome Profile (TOP) was initially developed by the Working Group of the German Society of Traumatology as a first PROM to assess the overall QOL, specifically in patients with multiple injuries. Although designed specifically for trauma patients, the TOP is a 57-item generic questionnaire focusing on 10 different constructs: depression, anxiousness, posttraumatic stress disorder, social interaction, daily activities, mental functioning, body image, satisfaction, pain, and body function, the latter assessed in both post-injury and pre-injury states. Each construct is given a rating of 0–100 with 0 being the worst state and 100 the best. Additionally, an outcome score of less than 80 was marked as “conspicuous” to allow for a simpler comparison between individuals or over time. The aim of developing the TOP was to assess the QOL, aimed specifically at severely injured patients and their problems. The TOP was developed and pre-tested using 70 polytrauma patients and 70 controls with minor injuries, from which the author concluded that the TOP can be used as a standardized stand-alone screening measure in follow-up investigations for individual trauma patients. A further evaluation of the TOP in a well-defined cohort of polytrauma patients found it to be a reliable and well-discriminating score, covering both general and trauma-specific aspects of QOL, and exhibiting clear correlations with already existing scores such as the SF-36.

A study on the ability of outcome measures to capture the full range of patient-important health impacts in major trauma using the International Classification of Function, Disability and Health as a framework found that the TOP captured a wider breadth of health concepts compared to its PROM counterparts such as the SF-36 and EQ-5D. However, the TOP has not been validated in large cohorts of patients,
and this seems to be its biggest drawback alongside the considerable burden on the patient of such a lengthy assessment. Additionally, one could argue against the assessment of pre-injury pain and functional status retrospectively, as it is vulnerable to subjectivity and can vary widely. More importantly, the original publication is not available in the English language at present and as per the recommendations of its use in a Swiss population; further work with larger study groups is required.

**World Health Organization Disability Assessment Schedule 2.0.** The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) is a generic tool that assesses general health and disability based on a metric called the International Classification of Functioning, Disability and Health (ICF). There are two versions of the WHODAS 2.0: one with 36 domains and another with 12 domains, including assessment of cognition, mobility, self-care, life activities, and participation. WHODAS 2.0 has been shown to be both reliable and valid and uniquely, has been field tested in populations with varying degrees of both physical and cognitive illness in locations worldwide. However, it has certain limitations; for instance, bodily impairments and environmental factors are not included in the standard version. Similarly, it is designed only for adults; however, its 12-item version was found to be reliable and valid in a youth population as young as 15 years.

These tools as described are available to assess the generic health status. They have limited use in the orthopedic trauma domain as a sole assessment, due in part to their broad enquiry. Used in isolation, generalizability may be limited and ability to assess impact of interventions threatened by a lack of specificity. In order to afford focus on specific injury impact on outcome, more anatomical and injury- or disease-focused tools are available.

**Injury- and disease-specific tools.** Assessing outcome from this perspective, the upper limb, in particular the shoulder, has established patient outcome scoring on QOL after surgical interventions. Using seven assessment metrics, Oh et al found that no single outcome scale was superior and those tested did not reflect health-related QOL. In addition, they found that the scores correlated poorly with each other. Of note for research, this means that the comparison of a given result using different outcome instruments is of little practical utility.

Slobogean and Slobogeanc shed further light on issues specific to shoulder trauma scoring, identifying that part of the problem is the sheer number of instruments that are used and the fact that these are intended for a disparate patient group. Over thirty assessment tools have been described, although the bulk of these are designed to score chronic conditions such as rotator cuff pathology or the spectrum of shoulder instability. Concentrating on the four commonest scoring systems, Slobogean and Slobogeanc found that most instruments have not been properly evaluated in shoulder fracture subjects and that there are substantial limitations to selecting the “best” instrument to measure outcomes following shoulder trauma.

Moving down the arm, Hoang-Kim et al found that this issue with multiple instruments is not confined to the shoulder. In the hand and wrist, like the shoulder, a wide range of instruments for use across a broad spectrum of pathology has led to lack of standardization. The authors emphasize that this heterogeneous population, assessed using tools unintended for trauma, leads to implications for the generalizability of results and interpretation of trial findings. Goldhahn et al reiterated the historical difficulties in assessing different tools in distal radius fractures. Looking at identifying a core set of domains for standardized reporting in clinical practice and research, the authors highlighted the key discrepancy: tools for research do not always correlate with tools for clinical decision making and evaluation. In their review, they found that many clinicians still see patient-reported measures as research tools and not for everyday use. The need for tools that combine both was seen as an area for further development.

Overall, one of the most widely used and validated wrist-specific scoring systems is the Patient-Rated Wrist Evaluation (PRWE) tool. Developed through the survey of 100 members of the international wrist investigators, patient opinions on pain and ability to do activities of daily living and work were thought to be the most important dimensions to include in subjective outcome tools. Brevity and simplicity were also seen as essential in the clinical environment.

The individual elements of the score have been subjected to systematic review and across 22 studies, the PRWE was found to be reliable, valid, and responsive across many wrist/hand conditions, critically, including trauma. Of note, with this tool and where it differs from many other assessment devices, the minimum clinically important difference and the minimal detectable difference have been studied. Finding a value of 11.5 points as the key difference to power studies in this population, Walenkamp et al have added significantly to the ease of its use, as it can be used to inform sample sizes in this population.

Looking at the hip, Ahmad et al identified the key issue of the requirement for both injury-/joint-specific scoring and disease-specific QOL outcome measures when evaluating interventions. Joint-specific hip evaluations include the objective, clinician-evaluated Merle d’Aubigne and Postel and Charnley scores, which are prone to observer bias and do not reflect patient-reported components or contain psychometric elements.

Other joint-specific scores are the Harris and Oxford Hip Scores (OHS), the former again clinician interpreted and prone to bias. The OHS is a “modern” PROM, completed by patients and is used as part of the evaluation of arthroplasty in the United Kingdom National Joint Registry. The strengths of these scores lie in their ease of use for either clinician or patient, although for all other than the OHS, patient reporting is minimal. Designed for evaluating interventions for joint
disease, none of these scores has been convincingly validated in trauma.

Apart from the joint-specific scores, Ahmad et al also reviewed the disease-specific QOL outcome measures used around the hip. Of five tools identified in their literature search, the authors found two of use in orthopedics, the McMaster University Osteoarthritis Index (WOMAC) and the Arthritis Impact Measurement Scale 2 (AIMS 2). The WOMAC is a validated, patient-completed measure for osteoarthritis of the hip and is used in evaluating outcomes following total hip replacement. As with other such scoring systems, the WOMAC groups responses into domains—pain, stiffness, and physical function. A value of the WOMAC is its reliability and sensitivity to change. Weakness does exist, however, with its specificity—the domains are heavily influenced by other conditions such as low back pain. Although commonly used in assessing trauma populations, there is no evidence for its validation across generic lower limb injury.

AIMS2 is a widely used, disease-specific measure that has a broad scope, measuring many aspects of the health status. Although very broad, it is designed primarily with a rheumatology population in mind and has not been widely used or validated in significant trauma.40

Looking at injuries around the knee, Schmidt-Rohlfing et al41 identified nine studies that report on outcome following traumatic injury to the joint, highlighting that a valid scoring system requires the involvement of varying elements, such as activity level, functional ability, symptoms, clinical findings, and self-reported satisfaction. As with other scoring systems, the authors found that the objective elements of the available instruments are easy to record. Range of knee motion, presence of swelling, and return to work and sport are objective and straightforward parameters.

The more subjective elements of all of the scores (The Lysholm score, the Tegner activity score, the Activities of Daily Living scale of the Knee Outcome Survey, the Cincinnati knee rating system, the International Knee Documentation Committee Score, and the Knee Injury and Osteoarthritis Outcome Score) may act as limitations to their use, particularly in the multiple injured patient, due to injury pattern overlay. In essence, the tools may be appropriate while scoring for a relatively benign isolated injury such as a sporting cruciate ligament disruption. They are less appropriate and therefore of limited use in significant trauma.

Completing the anatomical injury assessment with the Foot and Ankle, Farrugia et al42 emphasized the theme from across all of the body region—specific tools; not all of these outcome measures have been completely validated. This leads to challenges in applying the results of outcomes research to specific patients and certainly limits their immediate generalizability to the trauma population. Similarly, with other body regions, one of the challenges is the sheer number of tools used. Looking at disease-specific foot and ankle scores, the authors re-emphasize that these instruments are designed to be sensitive to the unique characteristics of one distinct disease state. This is a fundamental point that pervades all disease-specific outcomes assessment—information gathered with these instruments cannot be generalized to other disease states or populations. For example, one of the most widely used scores pertaining to ankle fractures, while featuring in a huge range of publications, cannot be used to assess forearm trauma.43 Moving to region-specific scores, the authors highlight that region-specific instruments are not as inclusive as generic measures because items on these instruments apply to only one region of the body. The advantage of these regional tools, however, is that they are not as restrictive as disease-specific instruments and as such can be used in the assessment of more than one condition.

The two most widely used regional specific instruments used in the foot and ankle are the American Orthopedic Foot and Ankle Society scale (AOFAS) and the Foot Function Index (FFI). These tools, despite widespread use, have considerable limitations. The AOFAS tool, although popular and featuring widely in the literature in many different studies, has been demonstrated to have concerns with validity, consistency, and responsiveness.44,45 The FFI in contrast has been correlated for outcome assessment with the SF36, thus supporting its validity as a healthcare measure. The FFI, as stated by Farrugia et al,42 is commonly thought of as a rheumatological condition assessment tool, although it is not that specific and can be used in other conditions. The authors conclude their assessment of foot and ankle tools by stating that more work is required in producing meaningful assessment systems due to the lack of validity or restriction of use of many of the available scoring systems.

### Discussion

The use of scoring systems has the potential to be a valuable indicator of an individual’s quality of health, and furthermore to assess the major trauma service offered by a provider. However, in order to implement the appropriate use of regular QOL outcome assessments in major trauma, it is necessary to overcome a number of key issues.

First, the right tool must be selected for a given task. The first distinction in choosing the right tool is between a general and a specific instrument and in the specific tool, choosing one that is either disease specific or region specific. Generally speaking, generic assessments such as the SF36 or EQ5D capture broader aspects of an individual’s health, while specific tools tend to focus on a particular illness or injury to a given body region and its effect on the individual’s health. The use of a generic tool allows for the identification of unexpected positive or negative results of an intervention and also enables comparison across injuries.

However, a limitation is that these are less sensitive to changes in various aspects of an individual’s health in comparison to that of a specific tool such as the PRWE score.
for wrist trauma. In an attempt to assess the different aspects of a patient’s clinical condition, combining a generic tool with a specific one can be useful. For example, since 2009, all patients undergoing hip replacement, knee replacement, and varicose vein and groin hernia surgery have been asked to participate in a combination of specific and general PROMs.\(^4^6\)

Here, the same EQ-5D questionnaire is used to assess the overall health of individuals before and after each surgery, with additional specific PROMs (such as the Oxford Knee Score for knee arthroplasty) being used for individual surgeries. The data from the EQ-5D shows that there has been a limited scope for improvement in outcomes for varicose vein surgery and hernia repair when compared to hip and knee replacements, as patients undergoing surgery for hernias and varicose veins typically already have a high pre-operative health score.\(^4^5\) This reflects the lack of sensitivity of a generic tool like EQ-5D to detect clinically significant improvements made to a specific condition, thus exhibiting the need for a specific PROM to be used in conjunction. Despite this, the EQ-5D is an excellent tool for the comparison of the effect of health interventions on the overall QOL of a patient, and the effect of major trauma generally spans a wide array of factors contributing to a person’s overall health.

The challenge remains to identify a suitable specific assessment tool for major trauma patients, as the specific illness and intervention can vary widely between each patient, for example, head injury with conservative management versus mid-shaft femur fracture with intra-medullary nailing. Thus, the solution might be to select a generic assessment tool for all trauma patients with specific PROMs offered to common injuries and their management. However, there are a great number of PROMs available for each injury as demonstrated by the fact that there are 139 unique outcome scales used in foot and ankle surgery alone.\(^4^8\) The selection of an appropriate PROM for each injury would therefore be a very laborious task.

Once an appropriate tool has been identified, a further key issue arises in the use of the data gathered from the population. In the above example of the four surgeries, PROMs are used to compare an individual’s QOL before and after a particular intervention. Owing to the unpredictable nature of trauma, it is not possible to obtain a score before the major trauma incident and retrospective data collection would be too unreliable. Instead, assessment tools can be useful in long-term follow-up of trauma patients. Multiple studies using PROMs have highlighted the health deficits in patients many years after the major trauma incident\(^4^9,5^0\) and the regular use of these assessment tools would be useful to capture aspects of health that might still be affected long after the incident.

Another way by which data from patient-related QOL scoring can be used would be to make comparisons of outcome measures between different interventions and care providers in order to identify statistically significant outliers resulting in the evaluation of the intervention. For example, Barnsley Hospital NHS Foundation Trust was identified as a negative outlier in primary hip replacements (based on the Oxford hip score) in during the period 2011–2012. After evaluation of their enhanced recovery program and delivery of physiotherapy and implementation of the subsequent suggested changes, they were no longer found to be below the negative outlier threshold in during 2013–2014.\(^5^1\) Similarly, PROM data were used in Northumbria NHS Healthcare Foundation Trust to identify a better brand of prosthesis and surgical technique in patients undergoing knee surgeries.\(^5^2,5^3\) Due to the lack of pre-intervention outcome measures, the comparison would have to be the post-intervention outcome measure by a certain health provider or a certain intervention within the context of an individual’s demographics and the severity of the major trauma.

Finally, the regular use of PROMs in major trauma will produce large amounts of data but for this data to be useful, it must be of sufficient high quality. A poor compliance in filling out the PROM questionnaire for example can result in a population bias, thus leading to an invalid conclusion; the more severely injured might be less likely to respond and may therefore be underrepresented. Additionally, multiple lengthy questionnaires may act as deterrents to filling in the PROM questionnaires for certain individuals. To avoid such population bias, a significant amount of resources would need to be assigned to appropriately follow-up the trauma patients; to maximize the response, a “less is more” method would need to be adopted with short single forms. Among the discussed PROMs, the EQ-5D is the shortest and takes the least time to complete. It is also most useful for comparison as the EQ-5D; in addition to being used widely in the NHS for QALY calculations, it has also been recommended as an outcome indicator in major trauma in England, with a pilot study to follow.\(^2^1\) In addition to long-term follow-up, if the data are to be used for comparison among services and the available interventions, the quality of the data recorded for the severity and type of major trauma injury would be a critical factor. Unfortunately, like PROMs, there are many trauma-scoring systems to grade major trauma, such as the Abbreviated Injury Scale, Injury Severity Score, and Organ Injury Scaling, each with their own comparative merits and pitfalls.\(^5^4\)

The discussion of a PROM is never complete without the mention of the validity and reliability of the measure. In broad terms, reliability means the ability of a test to reproduce similar results.\(^5^5\) It can be further subdivided into internal consistency, which represents the correlation between different domains of a survey, and test–retest reliability, which is the measure of responses obtained by the same individual at a later date.\(^5^5\) The reliability of a measure is usually expressed as numerical correlation indices ranging between 0 and 1.0 with 1.0 meaning results exactly the same when the test is repeated.\(^5^6\) The reliability of some PROMs has been mentioned in literature; examples of general outcome measures are shown in Table 1.

Validity, on the other hand, is the ability of a test to measure what is supposed to be measured, and in general, there is no single “gold standard” for testing the validity of a PROM and inference.
against a standard should be made.\textsuperscript{55,56} Some categories were identified as parameters for testing the validity of a construct, including “content validity” and “construct validity,” although the details of which are beyond the scope of this review.\textsuperscript{55,56}

### Conclusion

Patient-Reported Outcome Measures of QOL are increasingly being used in clinical settings to improve services in many specialties and major trauma should be no exception. With the aims of the NHS Outcomes Framework in mind, PROMs can be used for long-term follow-up of patients and can further serve as a useful indicator for improving recovery from injuries and trauma. When used in conjunction with an appropriate trauma outcome score, it can also be used to compare different interventions and different trauma service providers, in order to identify statistical outliers for further investigation. The EQ-5D is the shortest, yet a valid and reliable PROM in trauma. Given that it is widely used in the NHS and has also been recommended to be used as an outcome indicator in major trauma survivors in England, the EQ-5D is the perfect candidate for regular PROM data collection in major trauma patients in England. The challenge remains in choosing the injury- or region-specific tool to go with it and how that interface for each injury group is weighted, particularly in the polytrauma patient. This should be the focus for multicenter or registry studies in the future.

### Author Contributions

Conceived and designed the experiments: WE. Analyzed the data: WE, KH, and BA. Wrote the first draft of the manuscript: KH, PR, and BA. Contributed to the writing of the manuscript: WE. Agree with manuscript results and conclusions: WE, KH, PR, and BA. Jointly developed the structure and arguments for the paper: WE. Made critical revisions and approved final version of the manuscript: WE. All authors reviewed and approved of the final manuscript.

### REFERENCES


