Multi-method versus single method appraisal of clinical quality indicators for the emergency medical services

Howard, Ian

2020-12-01

Howard , I , Cameron , P , Castrén , M , Wallis , L & Lindström , V 2020 , ' Multi-method versus single method appraisal of clinical quality indicators for the emergency medical services ' , International Journal for Quality in Health Care : journal of the International Society for Quality in Health Care , vol. 33 , no. 1 , mzaa171 . https://doi.org/10.1093/intqhc/mzaa171

http://hdl.handle.net/10138/341384
https://doi.org/10.1093/intqhc/mzaa171

acceptedVersion

Title: Multi-method vs single method appraisal of clinical quality indicators for the Emergency Medical Services

Running head: Multi-method QI appraisal

Abstract word count: 248

Manuscript word count: 2992

Authors and affiliations:

- Ian Howard BEMC, MPhil (Corresponding author)
  - Department of Clinical Science and Education, Södersjukhuset
    Karolinska Institutet
    Stockholm
    Sweden
  - Division of Emergency Medicine
    Stellenbosch University
    Stellenbosch
    South Africa
    ianhoward@outlook.com
    +974 3359 9143

- Peter Cameron MD, FACEM
  - School of Public Health and Preventive Medicine
    Monash University
    Melbourne
    Australia
    peter.cameron@monash.edu

- Maaret Castrén MD, PhD
  - Department of Emergency Medicine and Services
    Helsinki University Hospital, Helsinki University
    Helsinki
    Finland
    maaret.castren@hus.fi

- Lee Wallis MBChB, FRCEM
  - Division of Emergency Medicine
    University of Cape Town
    Cape Town
  - Division of Emergency Medicine

Stellenbosch University
Stellenbosch
South Africa
lee.wallis@uct.ac.za

- Veronica Lindström RN, PhD
  - Department of Neurobiology, Care Sciences and Society, section of Nursing Karolinska Institutet
    Stockholm
    Sweden
  - Academic EMS
    Stockholm
    Sweden
    veronica.lindstrom@ki.se

Multi-method vs single method appraisal of clinical quality indicators for the Emergency Medical Services

ABSTRACT

Background

Quality Indicator (QI) appraisal protocols are a novel methodology that combines multiple appraisal methods to comprehensively assess the "appropriateness" of QIs for a particular healthcare setting. However, they remain inadequately explored compared to the single appraisal method approach. This paper aimed to describe and test a QI appraisal protocol versus the single method approach, against a series of QIs potentially relevant to the South African Prehospital Emergency Care setting.

Methods

An appraisal protocol was developed consisting of two categorical-based appraisal methods, combined with the qualitative analysis of the discussion generated during the consensus application of each method. The output of the protocol was assessed and compared with the application and output of each method. Inter-rater reliability of each particular method was evaluated prior to group consensus rating. Variation in the number of non-valid QIs and the proportion of non-valid QIs identified between each method and the protocol were compared and assessed.

Results

There was mixed IRR of the individual methods. There was similarly low to moderate correlation of the results obtained between the particular methods (Spearman's rank correlation=0.42,p<0.001). From a series of 104 QIs, 11 non-valid QIs were identified that were shared between the individual methods. A further 19 non-valid QIs were identified and not shared by each method, highlighting the benefits of a multi-method approach. The outcomes were additionally evident in the group discussion analysis, which in and of itself added further input that would not have otherwise been captured by the individual methods alone.

Conclusion

The utilization of a multi-method appraisal protocol offers multiple benefits, when compared to the single appraisal approach, and can provide the confidence that the outcomes of the appraisal will ensure a strong foundation on which the QI framework can be successfully implemented.

Multi-method vs single method appraisal of clinical quality indicators for the Emergency Medical Services

BACKGROUND

The Institute of Medicine defines healthcare quality as "*the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge*''[1]. Objectively assessing the extent to which this is achieved can be a challenging task, given that quality is a relatively abstract concept. It stands to reason, therefore that a central tenet to defining quality is the system used towards its measurement.

The measurement of healthcare quality provides an essential mechanism towards directing policy; benchmarking performance; guiding improvement initiatives and maintaining accountability of the system[2–4]. Consequently, multiple users will consume quality data in a variety of ways to achieve these aims. Therefore, for any measurement system to be successful, it is fundamental that it be comprehensive in its approach, yet simple in its design, and contextually relevant to provide an appropriate measure of quality.

Considerable progress has been made towards improving Prehospital Emergency Care (PEC) quality measurement, mainly through the development of PEC-specific quality indicators (QIs)[5–7]. In and of themselves, QIs cannot improve quality; they effectively provide clinicians and organizations with a quantitative basis to monitor, evaluate, and improve the quality of patient care, clinical support services, and organizational function[3,4]. Despite their advantages, the objective appraisal of quality measurement systems is often neglected, leading to the potential for implementation of inappropriate QIs. In the PEC environment, this is already evident in the literature where less than 15% of QIs have undergone some form of measure evaluation[7]. The consequences of inadequately assessed reliability, validity and bias in quality measurement can in the best-case scenario prove to be time-consuming and costly, and in the worst-case scenario potentially undermine the system in its entirety[8,9].

Several methodologies to appraise QIs have been described and utilized with considerable success[9–17]. While there is a level of overlap or commonality in the components they assess, the process towards their application can vary significantly[9–17]. Therefore, the potential exists for variation in the outcomes of these methodologies when applied to a common data set. QI appraisal protocols are a novel methodology that combines multiple appraisal methods to comprehensively assess the "appropriateness" of QIs for a particular setting[9]. There is limited evidence to suggest benefits with the use of such protocols; however, they remain inadequately explored compared to the single appraisal method approach. This paper aimed to therefore describe and test a multi-method QI appraisal protocol versus the

single method approach, against a series of QIs previously identified as potentially relevant to the PEC setting.

METHODS

The triangulation and integration of multiple data types have been increasingly recognized as a valuable approach to the study of healthcare delivery[18–20]. For this study, an appraisal protocol was developed consisting of two categorical-based appraisal methods, the Qualify appraisal tool and RAND Appropriateness method, combined with the qualitative analysis of the consensus application of each process, by a QI Appraisal Working Group. The protocol was tested against a series of QIs recently identified for potential relevance and used in the South African EMS setting and applied over three rounds (Figure 1). The final results of the protocol were compared and assessed against the outcomes of each method, with the rounds 1 and 3 serving as their own control test against the protocol (Table 3 provides the full list of QIs used for evaluation. See supplementary file for the data dictionary for each QIs).

For round 1, the Qualify QI appraisal tool was selected given its focus on feasibility[11,12], and consists of four-level Likert scale questions (1=Does not apply; 2=Rather does not apply; 3=Rather applies; 4=Applies) to assess 18 criteria amongst three categories: *Relevance*; *Scientific Soundness* and *Feasibility* (Table 1). For round 3, the Rand Appropriateness Method was included due to its practical focus (i.e., the data extraction)[14–17] to further rate the indicators by testing the definitions, data components and criteria for use developed for each QI against several clinical vignettes. Four categories (*Clarity*, *Necessity*, *Acceptability* and *Technical Feasibility*) were rated using a 9-point visual analogue scale, and data extraction assessed using a mock-up of a generic patient report form for the vignettes[9,13]. Two separate vignettes were developed for each of the QI categories included in the data extraction, and a "low-quality documentation" and "high-quality documentation" version developed for each vignette used during the assessment.

Both methods consisted of an evidence evaluation component as part of the appraisal process. To achieve this, the QIs were assessed for inclusion within local clinical practice guidelines (CPGs), and against the results of a literature review of the evidence base utilized for the development of PEC focused QIs. For the review, articles were identified by searching the following databases: PubMed; Embase; Cumulative Index to Nursing and Allied Health Literature (CINAHL); Web of Science; and the Cochrane Library. All searches were performed with no restrictions in terms of publication type or journal subset, date of publication, or patient age. Where applicable, searches were limited to English language articles and research involving human subjects only. Combinations and truncated variations of the following search terms were used for each database search: *Emergency Medical Service, prehospital emergency care, ambulance service, quality indicator, quality measure,*

performance measure, and performance indicator. Appropriate wildcards were used to account for singular and plural forms of each of the search terms. Variations in spelling were additionally used in varying combinations to broaden the search.

Inclusion Criteria

For this study, a QI was defined as *any measure that compared actual care against ideal criteria; or a tool used to help assess quality and/or performance*. The following minimum criteria were utilized when identifying studies for further analysis:

- Research that focused on the development and/or implementation of prehospital focused QIs
- The primary aim of the research was to describe, analyse, discuss or provide evidence for prehospital focused QIs

Exclusion Criteria

Non-English research, studies that examined disaster management/ major incident response QIs, or research aimed at inter-facility transport measures of care were excluded. Furthermore, secondary research that examined QIs developed as part of a primary study already included in the analysis was excluded.

Article Review

Eligible articles were identified and analysed independently in three parts by the primary author (IH) and two participants of the working group. Full-text articles remaining after a title and abstract review were independently reviewed for the satisfaction of the inclusion and exclusion criteria, and to determine whether they provided evidence for at least one indicator. The level of evidence for each article and QI was assessed and presented using the Oxford Centre for Evidence-based Medicine Levels of Evidence[21].

Data for parts 1 and 2 were collected over three rounds of group discussion of a QI Appraisal Working Group. An initial introductory round was conducted to familiarize the Working Group with the QUALIFY tool, Rand methodology, results of the literature review, and provide the data dictionary for the QI set. The data dictionary was the primary documented utilized by the Working Group for the application of each appraisal method and outlined 19 definable components for each QI. (Figure 2)

Prior to Round 1, the QUALIFY tool was independently applied by each member of the Working Group, who then met to discuss their individual scoring and apply a final consensus summary score during Round 1. Prior to Round 2, the Working Group similarly independently assessed the results of the literature review and then met to apply a final consensus rating of the evidence during Round 2. For round 3, the Working Group met to compare their individual data extraction results and rate the QIs using the Rand method. The Working Group meetings were recorded and later transcribed for the final part of data

collection – content analysis of the discussion generated surrounding the consensus appraisal process for Rounds 1 to 3.

Setting and Population

Traditionally, quality in the PEC setting has been exclusively reported based around response time targets[22–25]. This is no different from what is found in South Africa, where the utilization and reporting of clinically focused QIs by the Emergency Medical Services (EMS) are wholly lacking. Towards this, several clinically focused QIs have recently been identified for potential relevance to the SA PEC setting[26]. These QIs were used to test the appraisal protocol, with the secondary aim of identifying those QIs appropriate for use in the SA PEC setting.

The QI Appraisal Working Group consisted of nine experts chosen for their intricate knowledge of the South African PEC setting and to align with minimum panel size recommendations for each methodology[10,13]. All the participants were South African trained and post-graduate educated Emergency Care Practitioners (ECPs) with > 10 years of operational experience each. Six of the participants' primary experience and occupation were in quality governance and improvement within PEC, and the remaining three were primarily involved in clinical operations. The Working Group were given one month between each round with which to work through the information and data collection required for each subsequent round.

Data Analysis

Descriptive statistics were utilized to describe and summarize categorical based appraisal data. For the QUALIFY tool, mean scores per category, and the number of criteria scoring 3 (Rather applies) or 4 (Applies) were calculated for each QI. For the Rand method, consensus scores per category, and the proportion of categories scoring ≥7 were calculated. Inter-rater reliability (IRR) for each criterion of both the QUALIFY tool and Rand method were calculated using percentage agreement and Gwet's AC1.

A final composite score was calculated for each QI, for each method. For the QUALIFY tool, this was calculated using a weighted mean of the appraisal categories after consensus, due to the differences in the number of criteria per class. To be considered a valid indicator, the QI had to score ≥3 based on the final composite score. For the Rand method, the unweighted mean of the appraisal categories after consensus was used. To be considered a valid indicator, the QI had to score ≥7 based on the final composite score. A second group of QIs were identified consisting of those scoring on the validity threshold (3.0-3.1 for the Qualify tool; 7.0-7.1 for the Rand method) for which caution was recommended before full implementation.

Correlation between the final composite scores was calculated using the Spearman's rank correlation. The consensus derived proportion of non-valid QIs, and QIs for which caution was recommended, identified by each individual method and the protocol, were calculated and assessed against each other using the z-test. 95% confidence intervals were calculated where necessary and a p-value of 0.05 used as a cut-off for the strength of evidence. All data were entered and analysed using a combination of Microsoft Excel 2010 (Microsoft Corp., Richmond, WA, USA) and Stata version 16 (StataCorp. College Station, TX: StataCorp LLC).

Conventional content analysis, as described by Hsieh and Shannon, was utilized to sort and analyse the group discussions generated during the three rounds[27]. Recordings and transcripts were created for each round, and each transcript reread for content familiarisation. First-level coding was conducted through the extraction of meaning units from each transcript and summarised into codes using open-coding from each interview. Once completed, similar codes were combined and organized to develop clustered subcategories pertaining to each appraisal tool. Transcriptions were analysed using MAXQDA software for data storage; extraction of meaning units and subcategory development (MAXQDA, 2016; Sozialforschung GmbH, Berlin, Germany).

RESULTS

The Working Group appraised a total of 90 *clinical* and 14 *non-clinical* (n=104) QIs using each method, over the three rounds. There was a high level of validity of the QIs assessed across the majority of the appraisal criteria for both methods, the results of which were moderately correlated between each method.

Round 1 - QI Appraisal Tool

There was mixed IRR of the criteria found prior to the group consensus. *Validity* and *Understandability & interpretability for medical personnel* scored perfect agreement by the Working group. In contrast, *Data Collection Effort* (% agreement=22%, IRR=0.01) and *Understandability & interpretability for patients and interested public* (% agreement=28%, IRR=0.09) and scored the lowest (Table 2). Of the 104 QIs assessed, eight (7.7%) scored less than the validity threshold on the final composite score (≥3). All eight scored relatively high for *Relevance* and *Scientific Soundness* yet scored poorly for *Feasibility*. A further 15 QIs scored on the validity threshold (3.0-3.1).

To appraise the *Indicator Evidence* criterion within the *Scientific Soundness* category, the QIs were evaluated for inclusion within local CPGs. There was considerable representation of the QIs amongst the SA national EMS CPGs (Table 3). Seventy-nine QIs (76%) were accounted for in the CPGs, of which 76 (73%) had evidence directly supporting their use. Those QIs not represented were found to be either structure-based QIs; clinical bundle-based QIs; or those QIs focusing on sentinel events and patient safety.

Round 2 – Literature Review

The literature search identified a total of 1624 potential articles for review (Figure 3). Following a title and abstract review, 1528 articles did not meet inclusion criteria and were excluded, leaving 89 articles for full-text review. Following the removal of duplicate texts, and research not meeting the inclusion criteria (n=57) 31 articles remained for the full-text review. The literature review found an evidence base for 11 of the 15 Clinical subcategories and the 2 Non-clinical subcategories, plus an additional 4 subcategories not included in the QI appraisal, covering 311 indicators (Table 4). More than half (59%) were developed through a consensus/expert opinion-based approach, with fewer developed via more robust and higher quality levels of evidence such as systematic reviews or cohort and case control-based studies (10% each).

Round 3 – Rand Method

As with the appraisal tool, there was mixed IRR in the individual rating prior to the consensus rating, with *Acceptability* scoring the highest (% agreement=90%, IRR=0.9) and *Technical Feasibility* the lowest (% agreement=47%, IRR=0.32). Eleven QIs (10.6%) scored below the validity threshold, and a further eight QIs scored on the validity threshold (7.0-7.1). In total, from a series of 104 QIs, eight were identified as non-valid and three identified for which caution was recommended before full implementation, that were shared between the appraisal methods. A further 19 QIs were identified as non-valid and not shared by each method.

Comparison of Categorical Appraisal Methods

When final consensus validity scores were compared, there was poor to moderate correlation of the results between the QUALIFY tool and Rand method (Spearman's rank correlation=0.42, p<0.001). Ninety-two of the 104 QIs (88%) (78 *clinical* and 14 *non-clinical*) were appraised to be valid and feasible for the SA PEC setting, based on the results of this study. Of this group, an additional 21 QIs (13 *clinical* and eight *non-clinical*) were assessed to be on the threshold of validity, in which caution is recommended before full implementation. There was little evidence to support a statistical difference in the proportion of non-valid QIs identified between the Qualify tool and the Rand method [difference=-0.03; (95%CI -0.12:0.05,p=0.47)]; between the Qualify tool and the protocol [difference=-0.05; (95%CI -0.13:0.03,p=0.25)]; or between the Rand method and the protocol [difference=-0.02; (95%CI -0.11:0.07,p=0.66)]. There was likewise little evidence to support a statistical difference in the proportion of QIs in which caution is recommended, identified between the Qualify tool and the Rand method [difference=0.07; (95%CI -0.02:0.15,p=0.12)]; or between the Qualify tool and the protocol [difference=-0.06; (95%CI -0.16:0.04,p=0.27)]. There was, however, strong evidence to support a statistical difference between the proportion of QIs in which caution is recommended, identified between the Rand method and the protocol [difference= -0.13; (95% CI -0.22:-0.03,p=0.009)].

Discussion Group Content Analysis

Several observations highlighted during the group discussions were found to be important considerations regarding the appraisal protocol and its ability to assess the appropriateness of the QIs. For the QUALIFY tool, *Relevance* and *Scientific Soundness* were perceived to be characteristics inherent to the QIs (and supporting data components) themselves, and as a result, were generally appraised to be highly applicable across all QIs and criteria (Table 5). In contrast, *Feasibility* was judged to be more of a gauge of the system in which the QIs would be implemented and as such, scores were found to be on average lower amongst these criteria [1.1, 1.2]. Somewhat related to this was the broader issue of context and the importance of selecting those indicators that best suited the local setting, before full implementation (1.3, 1.4]. Despite the focus on the appraisal of the QIs, on several occasions, the discussion steered towards the need for EMS organizations in SA to improve their quality systems in general if such measures are to be implemented [1.5, 1.6].

For the Rand method, the importance of having completed the practical data extraction using the case vignettes made a difference in the QI rating [2.1,2.2]. This expanded further into a general conversation about applying the QI framework, the quality system in which they'd be used and documentation in general [2.3 – 2.6].

DISCUSSION

The simplicity and practicality of QIs as a system of quality measurement has led to their widespread adoption in healthcare[4,14,28–34]. They align with Donabedian's conceptual framework for healthcare evaluation, predicated on the belief that an effective structure gives rise to effective processes of care, which in turn result in improved outcomes[8]. Importantly, there is no evidence to suggest that QIs have the potential to reduce mortality and morbidity if implemented and utilized effectively[35,36]. Within the PEC setting, patient exposure times are generally limited, and the delivery of care mainly based around processes as opposed to outcomes. The utilization of QIs as a measure of quality are, therefore ideally suited to this environment.

Despite these advantages, the implementation of inappropriate or poorly tested QIs - even in well-established quality systems - has been reported to be both time-consuming and costly to correct[9,14]. Furthermore, the clinical implications are potentially varied and far-reaching. Inappropriate QIs implemented within a particular setting will potentially lead to improper changes to clinical care that could unnecessarily and negatively impact patient safety. Moreover, inappropriately implemented QIs could additionally avert focus from unmonitored issues more suitable for the setting in question, thus too impacting patient safety. Consequently, QI appraisal has been identified as an essential step toward understanding the appropriateness of these measures for a particular healthcare field or

setting, before full implementation. The results of this study support these notions through the application of QI appraisal protocol against a series of QIs. Further to this, the results support the value in adopting a multi-method approach towards QI appraisal, compared to the single method approach.

Our observations found the multi-method approach to be advantageous in that the methods complemented each other's strengths and compensated for each other's weaknesses. While the Qualify tool appraised the QIs from a greater number of viewpoints, the Rand approach offered insight into the practical application of the QIs not available with the Qualify tool. This was additionally evident in the group discussion analysis, which in and of itself added further input towards understanding and appraising the appropriateness of the QIs that would not have otherwise been captured or understood by the categorical methods alone[18,37].

Despite these advantages, the application of the protocol required a significant investment in time and staff resources. The overall benefits of such an approach are therefore heavily dependent on the availability of these resources. This availability will likely vary significantly, depending on the quality system setting within which the protocol will be applied. These "system-focused" factors, therefore, have the potential to exert as much influence on the validity of the QIs as the setting in which the QIs will be implemented[38,39].

The outcomes of the appraisal have identified a significant number of QIs assessed to be valid and feasible for the SA PEC setting. The majority are centred around clinically focused processes of care, measures that are lacking in current performance assessment in EMS in SA. The importance and potential influence of the quality system in which the QIs will be implemented was further highlighted across all the methodologies. Quality system-focused assessment criteria, on average, scored lower than those criteria assessed to be characteristics inherent to the QIs themselves. This was reaffirmed during the qualitative discussion analysis, where system focused factors were a common discussion point.

Limitations

While the specific results of this study may not be readily generalizable to other settings or services, the concept of employing appraisal protocols to identify maximally relevant and appropriate QIs for use in a particular setting is well demonstrated in this paper. However, their outcomes are dependent on several factors that would need to be considered. Firstly, the inclusion of other appraisal tools in an appraisal protocol will likely produce different results to those observed in this study. The tools included in our protocol were chosen given their primary focus; however, other services or settings may place greater emphasis on qualities found in other appraisal tools. Secondly, the composition of the working group applying the tools and protocols could potentially use the same tools in a different manner. The make-up of the group should be focused on the setting and service type and where

possible, include individuals with experience and training in quality assessment and improvement or patient safety. Lastly, results will likely vary depending on the QIs under study, their purpose and focus.

CONCLUSION

Measurement forms a central part of every healthcare quality system. Regardless of the approach used, the framework must be comprehensively assessed for appropriateness for the setting in which it will be employed. Understanding and accounting for this as a factor is vital towards ensuring both successful implementation and ongoing utilization of quality systems in any setting. The utilization of a multi-method appraisal protocol offers significant benefit towards achieving this, when compared to the single appraisal approach, and can provide the confidence that the outcomes of the appraisal will ensure a strong foundation on which the measurement framework can be successfully implemented and employed.

ABBREVIATIONS

PEC: Prehospital Emergency Care
QI: Quality Indicator
SA: South Africa
CPG: Clinical Practice Guideline
IHI: Institute for Healthcare Improvement
EMS: Emergency Medical Services
ECP: Emergency Care Practitioner
ACS: Acute Coronary Syndrome
CI: Confidence Interval
CINAHL: Cumulative Index to Nursing and Allied Health Literature

FIGURES

Figure 1: Quality Indicator appraisal protocol

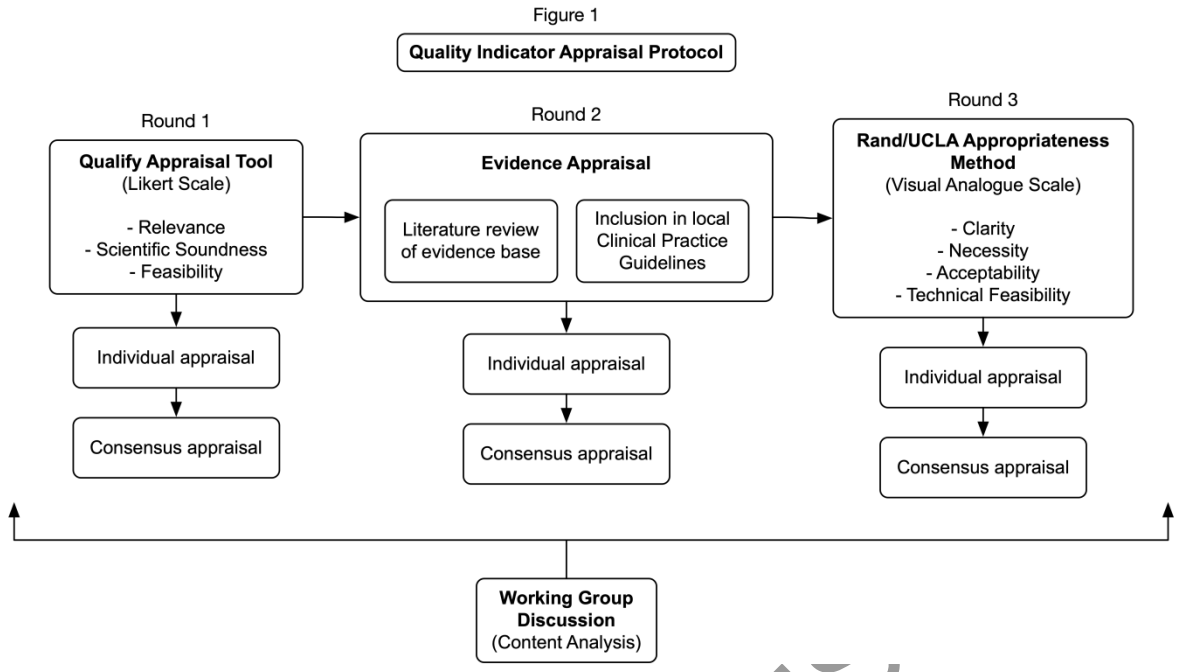
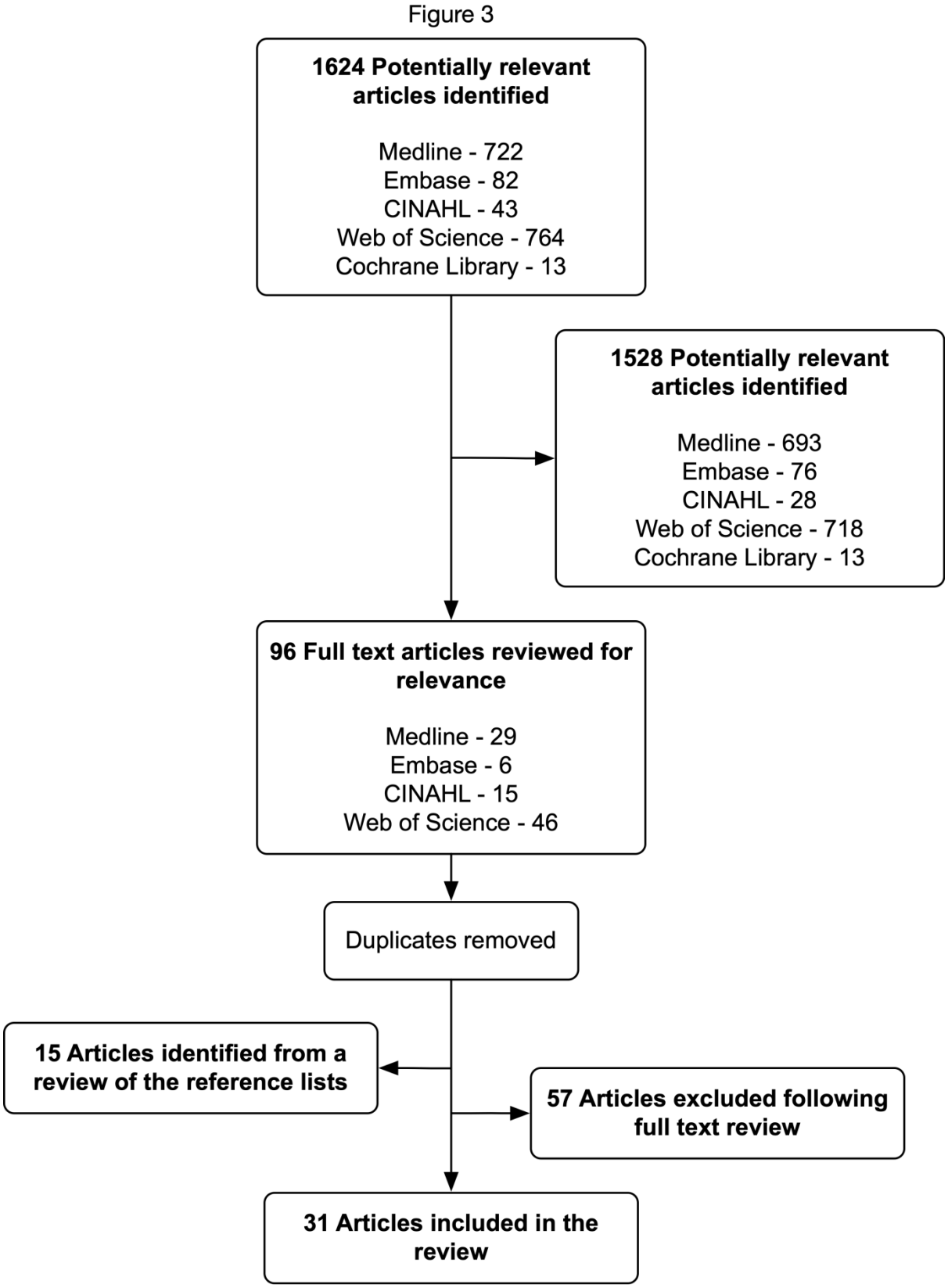


Figure 2: Data Dictionary components

Figure 2

Abbreviated Name	Abbreviated QI name
Definition	Basic description/purpose of the QI
Domain	Primary area of focus of the QI
Subdomain	Secondary area, within the Domain that is the subarea of focus
Clinical Pathway/Service Pathway	Identifies the Domain and Subdomain within which the QI is positioned
Measure Type	Structure, process or outcome focus
Target Population	Domain level population on whom the QI is measured/applied
Unit of Analysis	EMS component under study/assessment for quality and performance
Numerator Statement	Description of the subset of the subdomain population on whom the quality indicator is measured/applied
Denominator Statement	Description of the subdomain level of population on whom the quality indicator is measured/applied
Case Mix/Risk Adjustment	Suggested differentiation amongst the denominator population for greater accuracy (i.e.: stratification)
Exclusion Criteria	Denominator cases to be excluded when applying the QI
Measure Calculation	The equation for calculating the QI
Numerical Reporting Format	Suggested format in which the numerical results should be reported
Graphical Reporting Format	Suggested format in which the results should be displayed/visualized
Reported Indicator	Suggested output in which results should be described
Data Source	Suggested data source to obtain the data required for calculating the QI
Suggested Reporting Period	Timeframe, number of successive cases or other grouping strategies cases should be aggregated for reporting purposes
Recommended Review Period	Suggested time period at which the QI should be reviewed for validity and feasibility

Figure 3: Selection of articles for review



DECLARATIONS

Ethics approval and consent to participate

Ethical approval for the study was granted by the Stellenbosch University Health Research Ethics Committee (HREC) (Ref no. S15/09/193). Written consent for participation was provided by each of the participants before data collection. The datasets used and analysed during the current study available from the corresponding author on reasonable request.

Consent for publication

Not applicable/required

Competing interests

The authors declare that they have no competing interests.

Funding

Nil

Authors' contributions

IH, PC, MC, LW and VL conceived the study. IH, conducted the data collection and analysis. IH drafted the manuscript, and all authors contributed to its revision. All authors have read and approve of the final manuscript, and consent to its publication. IH takes responsibility for the paper.

Acknowledgements

Nil

Data Availability Statement

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

General

1. National Academies of Sciences; Engineering and Medicine. Crossing the global quality chasm: Improving health care worldwide. Washington, DC: The National Academies Press; 2018.
2. Mant J. Process versus outcome indicators in the assessment of quality of health care. Int J Qual Heal Care 2001;13(6):475–80.
3. Mainz J. Defining and classifying clinical indicators for quality improvement. Int J Qual Heal care 2003;15(6):523–30.
4. Mainz JAN. Developing evidence-based clinical indicators : a state of the art. Int Jounral Qual Healthc 2003;15(Suppliment 1):5–12.
5. Maio RF, Garrison HG, Spaire DW, et al. Emergency medical services outcomes

project I (EMSOP I): Prioritizing conditions for outcomes research. Ann Emerg Med 1999;33(4):423–32.
6. Spaite DW, Maio RF, Garrison HG, et al. Emergency Medical Services Outcomes Project (EMSOP) II : Developing the Foundation and Conceptual Models for Out-of-Hospital Outcomes. Ann Emerg Med 2001;37(6):657–63.
7. Howard I, Cameron P, Wallis L, Castren M, Lindstrom V. Quality Indicators for Evaluating Prehospital Emergency Care: A Scoping Review. 2017;
8. Donabedian A. Evaluating the Quality of Medical Care. Milbank Q 1966;44(Suppl):166–206.
9. Campbell S, Kontopantelis E, Hannon K, Burke M, Barber A, Lester H. Framework and indicator testing protocol for developing and piloting quality indicators for the UK quality and outcomes framework. BMC Fam Pract. 2011;12:85.
10. BQS - Institute of Quality and Patient Safety. QUALIFY: Instrument for the Assessment of Quality Indicators. 2007;(August):60.
11. De Cruppé W, Kleudgen S, Diel F, Burgdorf F, Geraedts M. Feasibility of 48 quality indicators in ambulatory care in Germany: A cross-sectional observational study. Z Evid Fortbild Qual Gesundhwes. 2015;109(9–10):682–94.
12. Bieler D, Hörster A, Lefering R, et al. Evaluation of new quality indicators for the TraumaRegister DGU using the systematic QUALIFY methodology. Eur J Trauma Emerg Surg [Internet] 2018;0(0):0.
13. Fitch K, Bernstein SJ, Aguilar MD, et al. The RAND/UCLA Appropriateness Method User's Manual. 2001;
14. Marshall M, Campbell S, Kirk SM, Brook RD, McGlynn E, Shekelle PG. Measuring General Practice: A demonstration project to develop and test a set of primary care clinical quality indicators Professors of General Practice. 2003;
15. Santori G, Fontana I, Valente R, Ghirelli R, Valente U. Application of the RAND/UCLA Appropriateness Method to Evaluate an Information System for Kidney/Pancreas Transplantation in Adult Recipients. Transplant Proc 2008;40(6):2021–3.
16. Bobrovitz N, Parrilla JS, Santana M, Straus SE, Stelfox HT. A qualitative analysis of a consensus process to develop quality indicators of injury care. Implement Sci. 2013;8:45.
17. Grudzen CR, Liddicoat R, Hoffman JR, Koenig W, Lorenz KA, Asch SM. Developing quality indicators for the appropriateness of resuscitation in prehospital atraumatic cardiac arrest. Prehospital Emerg care. 2007;11(4):434–42.
18. Curry LA, Nembhard IM, Bradley EH. Qualitative and mixed methods provide unique contributions to outcomes research. Circulation 2009;119(10):1442–52.
19. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. BMJ 2010;341(7783):1147–50.
20. Johnson RB, Onwuegbuzie AJ. Mixed Methods Research: A Research Paradigm Whose Time Has Come. 1973;
21. Howick J, Phillips B, Ball C, Sackett D, Badenoch D. Oxford Centre for Evidence-based Medicine Levels of Evidence. 2009;Available from: https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/
22. Department of Health KZN. 2017 / 18 Annual Report. Annu. Rep. 2016/17. 2018;1–417.
23. North West Department of Health. Annual Report. 2018;

24. Western Cape Department of Health. Western Cape Government,Department of Health Annual Report. 2017;
25. Limpopo Department of Health. Annual Report. Annu. Rep. 2017/18 Financ. Year. 2018;
26. Howard I, Cameron P, Wallis L, Castrén M, Lindström V. Identifying quality indicators for prehospital emergency care services in the low to middle income setting: The South African perspective. African J Emerg Med. 2019; 9(4): 185 - 192
27. Hseih H-F, Shannon SE. Three Approaches to Qualitative Content Analysis. Qual Health Res 2005;15(9):1277–88.
28. Kollberg B, Elg M, Lindmark J. Design and implementation of a performance measurement system in Swedish health care services: a multiple case study of 6 development teams. Qual. 2005;14(2):95–111.
29. Coleman P, Nicholl J. Consensus methods to identify a set of potential performance indicators for systems of emergency and urgent care. J Health Serv Res Policy. 2010 [cited 2014 Sep 30];15 Suppl 2:12–8.
30. Hoogervorst EM, van Beeck EF, Goslings JC, Bezemer PD, Bierens JJLM. Developing process guidelines for trauma care in the Netherlands for severely injured patients: results from a Delphi study. BMC Health Serv Res. 2013;13(1):79.
31. Abrams H, Belanger F, Carew D, et al. Prioritizing performance measurement for emergency department care : consensus on evidence- based quality of care indicators Steering committee. 2011;13(5):28–43.
32. Nakayama DK, Saitz EW, Gardner MJ, Kompare E, Guzik E, Rowe MI. Quality assessment in the pediatric trauma care system. J Pediatr Surg. 1989;24(2):159–62.
33. Mackenbach JP, McKee M. A comparative analysis of health policy performance in 43 European countries. Eur J Public Health. 2013. ;23(2):195–201.
34. Groene O, Klazinga N, Kazandjian V, Lombrail P, Bartels P. The World Health Organization Performance Assessment Tool for quality improvement in hospitals (PATH): An analysis of the pilot implementation in 37 hospitals. Int J Qual Heal Care 2008;20(3):155–61.
35. Jarvis JL, Gonzales J, Johns D, Sager L. Implementation of a Clinical Bundle to Reduce Out-of-Hospital Peri-intubation Hypoxia. Ann Emerg Med. 2018;72(3):272-279
36. Dusheiko M, Gravelle H, Martin S, Smith PC. Quality of disease management and risk of mortality in english primary care practices. Health Serv Res 2015;50(5):1452–71.
37. Van Harten WH, Casparie TF, Fisscher OAM. Methodological considerations on the assessment of the implementation of quality management systems. Health Policy (New York) 2000;54(3):187–200.
38. Kaplan HC, Provost LP, Froehle CM, Margolis PA. The model for understanding success in quality (MUSIQ): Building a theory of context in healthcare quality improvement. BMJ Qual Saf 2012;21(1):13–20.
39. Vaughn VM, Saint S, Krein SL, et al. Characteristics of healthcare organizations struggling to improve quality: Results from a systematic review of qualitative studies. BMJ Qual Saf 2019;28(1):74–84.

Quality Indicator Evidence Base Review

1. Norris RM. A new performance indicator for acute myocardial infarction. Heart.

2001;85(4):395–401.
2. Sobo E, Andriese S, Stroup C, Morgan D, Kurtin P. Developing Indicators for EMS Evaluation and Quality Improvement. Jt Comm J Qual Improv. 1992;27(3):138–54.
3. Siriwardena a N, Shaw D, Donohoe R, Black S, Stephenson J. Development and pilot of clinical performance indicators for English ambulance services. Emerg Med J. 2010;27(4):327–31.
4. Siriwardena AN, Shaw D, Essam N, et al. The effect of a national quality improvement collaborative on prehospital care for acute myocardial infarction and stroke in England. Implement Sci. 2014;9:17.
5. Myers JB, Slovis CM, Eckstein M, et al. Evidence-based performance measures for Emergency Medical Service systems – A Model for expanded EMS benchmarking. Prehospital Emerg Care 2008;12:141–51.
6. Colwell C, Mehler P, Harper J, Cassell L, Vazquez J, Sabel A. Measuring quality in the prehospital care of chest pain patients. Prehospital Emerg Care. 2009;13:237–40.
7. Rosengart MR, Nathens AB, Schiff M a. The identification of criteria to evaluate prehospital trauma care using the Delphi technique. J Trauma 2007;62(March):708–13.
8. Mohan D, Rosengart MR, Farris C, Fischhoff B, Angus DC, Barnato AE. Sources of non-compliance with clinical practice guidelines in trauma triage: a decision science study. Implement Sci. 2012;7:103.
9. HT S, Bobranska-Artiuch B, Nathens A, SE S. A systematic review of quality indicators for evaluating pediatric trauma care. Crit Care Med. 2010;38(4):1187-1196 10p.
10. Stelfox HT, Bobranska-Artiuch B, Nathens A, Straus SE. A systematic review of quality indicators for evaluating pediatric trauma care. Crit Care Med. 2010;38(4):1187–96.
11. Myers JB, Slovis CM, Eckstein M, et al. Evidence-Based Performance Measures for Emergency Medical Services Systems: A Model for Expanded EMS Benchmarking. Prehospital Emerg Care.2008;12(2):141–51.
12. Santana MJ, Stelfox HT. Development and evaluation of evidence-informed quality indicators for adult injury care. Ann Surg. 2014;259(1):186–92.
13. Dunford J, Domeier RM, Blackwell T, et al. Performance Measurements in Emergency Medical Services. Prehospital Emerg Care. 2002;6(1):107–13.
14. Nakayama DK, Saitz EW, Gardner MJ, Kompare E, Guzik E, Rowe MI. Quality assessment in the pediatric trauma care system. J Pediatr Surg. 1989;24(2):159–62.
15. Patterson PD, Lave JR, Martin-Gill C, et al. Measuring Adverse Events in Helicopter Emergency Medical Services: Establishing Content Validity. Prehospital Emerg Care. 2014;18(1):35–45. A
16. Hoogervorst EM, van Beeck EF, Goslings JC, Bezemer PD, Bierens JJLM. Developing process guidelines for trauma care in the Netherlands for severely injured patients: results from a Delphi study. BMC Health Serv Res. 2013;13(1):79.
17. Greenberg MD, Garrison HG, Delbridge TR, et al. Indicators for out-of-hospital Emergency Medical Services: The Paramedics Perspective. Prehospital Emerg Care.

1997;1:23–7.
18. Meara PO. A generic performance framework for ambulance services : an Australian health services perspective. 2012;3(3).
19. Stelfox HT, Straus SE, Nathens A, Bobranska-Artiuch B. Evidence for quality indicators to evaluate adult trauma care: a systematic review. Crit Care Med. 2011;39(4):846–59.
20. Valenzuela TD, Spaite DW, Meislin HW, Clark LL, Wright AL, Ewy GA. Emergency vehicle intervals versus collapse-to-CPR and collapse-to-defibrillation intervals: Monitoring emergency medical services system performance in sudden cardiac arrest. Ann Emerg Med. 1993;22(11):1678–83.
21. Grudzen CR, Liddicoat R, Hoffman JR, Koenig W, Lorenz KA, Asch SM. Developing quality indicators for the appropriateness of resuscitation in prehospital atraumatic cardiac arrest. Prehospital Emerg care. 2007;11(4):434–42.
22. Patterson PD, Lave JR, Martin-Gill C, et al. Measuring Adverse Events in Helicopter Emergency Medical Services: Establishing Content Validity. Prehosp Emerg Care. 2013
23. Daudelin DH, Kulick ER, Amore KD, Lutz JS, Barrientos MT, Foell K. The Massachusetts Emergency Medical Service Stroke Quality Improvement Collaborative 2009 - 2012. Prev Chronic Dis. 2013;10.
24. Siriwardena AN, Shaw D, Essam N, et al. The effect of a national quality improvement collaborative on prehospital care for acute myocardial infarction and stroke in England. Implement Sci. 2014;9(17):1–9.
25. Oostema JA, Nasiri M, Chassee T, Reeves MJ. The quality of prehospital ischemic stroke care: Compliance with guidelines and impact on in-hospital stroke response. J Stroke Cerebrovasc Dis. 2014;23(10):2773–9. 0
26. Stelfox HT, Straus SE. Measuring quality of care: Considering measurement frameworks and needs assessment to guide quality indicator development. J Clin Epidemiol. 2013;66(12):1320–7.
27. Nakayama DK, Gardner MJ, Waggoner T. Audit Filters in Quality Assurance in Pediatric Trauma Care. J Pediatr Surg. 1993;28(1):19–25.
28. Patterson PD, Probst JC, Moore CG. Expected annual emergency miles per ambulance: an indicator for measuring availability of Emergency Medical Services resources. J Rural Health. 2006;22(2):102–11.
29. Greenberg MD, Garrison HG, Delbridge TR, et al. Quality indicators for out-of-hospital emergency medical services: the paramedics' perspective. Prehosp Emerg Care. 1997;1(1):23–7.
30. Gitelman V, Auerbach K, Doveh E. Development of road safety performance indicators for trauma management in Europe. Accid Anal Prev. 2013;60:412–23.
31. Willis CD, Gabbe BJ, Cameron PA. Measuring quality in trauma care. Injury 2007;38(5):527–37.

Table 1: Quality Indicator appraisal tool categories and criteria*

Category	No.	Subcategory Criterion
Relevance	R1	Significance: "The indicator covers aspects of quality of life, morbidity, or mortality."
	R2	Benefit: "Use of the indicator can have a positive effect on the quality of care."
	R3	Potential risks/side effects: "No risks are known/assumed which may result from the use of the indicator."
Scientific soundness	S1	Unambiguity of definitions: "The indicator is defined clearly and unambiguously."
	S2	Reliability: "It is a reliable measurement."
	S3	Risk adjustment: "The indicator is sufficiently adjusted to risk" (Are all factors that are not caused by the user taken into due account?)
	S4	Sensitivity: "The indicator provides sufficient sensitivity."
	S5	Specificity: "The indicator provides sufficient specificity."
	S6	Validity: "The indicator provides sufficient validity."
Feasibility	F1	Understandability and interpretability for patients and interested public
	F2	Understandability and interpretability for medical and nursing personnel
	F3	Possibility to influence the indicator manifestation: "The quality indicator refers to an aspect of care which can be influenced by the actors to be assessed."
	F4	Availability of data: "The data are documented by the service provider as a routine or can be collected with acceptable effort."
	F5	Data collection effort: "There is no data collection method available that provides at least equivalent results with less effort."
	F6	Implementation barriers: "Implementation barriers are unknown or covered by adequate measures."
	F7	Accuracy: "The correctness of the data can be verified."
	F8	Data integrity: "Is the individual data set intact?"
	F9	Completeness of the data: "Is it possible to verify that all occurring cases were recorded?"

*BQS – Institute of Quality and Patient Safety. QUALIFY: Instrument for the Assessment of Quality Indicators. 2007;(August)

Table 2: Inter-rater reliability analysis of individual appraisal by the Quality Indicator Appraisal Working Group

Methodology		% agreement [p value (95% Confidence interval)]			Kappa [p value (95% Confidence interval)]		
Quality Indicator Appraisal Tool							
Relevance							
R1	Significance	90%	[<0.001	(0.8675 - 0.9350)]	0.90	[<0.001	(0.8587 - 0.9334)]
R2	Benefit	83%	[<0.001	(0.7934 - 0.8746)]	0.82	[<0.001	(0.7704 - 0.8669)]
R3	Potential risks/side effects	41%	[<0.001	(0.3887 - 0.4395)]	0.25	[<0.001	(0.2065 - 0.2840)]
Scientific Soundness							
S1	Unambiguity of definitions	81%	[<0.001	(0.7818 - 0.8465)]	0.80	[<0.001	(0.7664 - 0.8390)]
S2	Reliability	49%	[<0.001	(0.4614 - 0.5181)]	0.30	[<0.001	(0.2647 - 0.3434)]
S3	Risk adjustment	71%	[<0.001	(0.6789 - 0.7340)]	0.66	[<0.001	(0.6248 - 0.6975)]
S4	Sensitivity	80%	[<0.001	(0.7695 - 0.8395)]	0.78	[<0.001	(0.7426 - 0.8269)]
S5	Specificity	88%	[<0.001	(0.8502 - 0.9126)]	0.87	[<0.001	(0.8395 - 0.9093)]
S6	Validity	100%		(1)	1.00		(1)
Feasibility							
F1	Understandability and interpretability for patients and interested public	28%	[<0.001	(0.2670 - 0.2959)]	0.09	[<0.001	(0.0646 - 0.1076)]
F2	Understandability and interpretability for medical and nursing personnel	100%		(1)	1.00		(1)
F3	Possibility to influence the indicator manifestation	45%	[<0.001	(0.4286 - 0.4714)]	0.35	[<0.001	(0.3233 - 0.3835)]
F4	Availability of data	65%	[<0.001	(0.6434 - 0.6630)]	0.48	[<0.001	(0.4487 - 0.5134)]
F5	Data collection effort	22%	[<0.001	(0.2104 - 0.2345)]	0.01	[<0.001	(-0.0133 - 0.0235)]
F6	Implementation barriers	49%	[<0.001	(0.4803 - 0.5069)]	0.11	[<0.001	(0.0775 - 0.1503)]
F7	Accuracy	49%	[<0.001	(0.4803 - 0.5069)]	0.11	[<0.001	(0.0775 - 0.1503)]
F8	Data integrity	49%	[<0.001	(0.4765 - 0.5030)]	0.35	[<0.001	(0.3283 - 0.3695)]
F9	Completeness of the data	49%	[<0.001	(0.4765 - 0.5030)]	0.35	[<0.001	(0.3283 - 0.3695)]

RAND method						
Clarity	85%	[<0.001	(0.8079 - 0.8854)]	0.83	[<0.001	(0.7865 - 0.8786)]
Necessity	48%	[<0.001	(0.4663 - 0.5033)]	0.39	[<0.001	(0.3663 - 0.4196)]
Acceptability	90%	[<0.001	(0.8682 - 0.9363)]	0.90	[<0.001	(0.8585 - 0.9347)]
Technical Feasibility	47%	[<0.001	(0.4401 - 0.4958)]	0.32	[<0.001	(0.2735 - 0.3568)]

Table 3: Quality Indicator appraisal results

r for Review	QI Class	Relevance	Scientific Soundness	Feasibility	Appraisal Tool Score	Total criteria Applies	Applicable CPG	Supported in CPG	Clarity	Necessity	Acceptability	Techn Feasib
ACS/STEMI												
gnosis of ACS/STEMI who endance	Process	3.7	3.8	2.5	3.1	12	Yes	No	9.0	7.0	9.0	9.0
gnosis of ACS/STEMI who sk factors assessed and	Process	3.3	3.8	3.1	3.4	16	Yes	Yes	5.0	5.0	7.0	4.0
gnosis of ACS/STEMI who	Process	3.7	4.0	2.4	3.1	10	Yes	Yes	9.0	6.0	6.0	6.0
gnosis of ACS/STEMI who	Process	3.7	4.0	3.1	3.5	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of ACS/STEMI who	Process	3.7	4.0	3.1	3.5	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of ACS/STEMI who or thrombolysis by	Process	3.7	3.8	1.8	2.8	10	Yes	Yes	5.0	5.0	7.0	4.0
gnosis of ACS/STEMI who l thrombolysis	Process	3.7	3.8	1.8	2.8	10	Yes	Yes	5.0	5.0	7.0	4.0
gnosis of ACS/STEMI who Facility with PCI	Process	3.3	4.0	1.8	2.8	9	Yes	Yes	5.0	5.0	7.0	4.0
gnosis of ACS/STEMI who eiving Cath Lab	Process	3.7	3.8	1.8	2.8	10	Yes	Yes	5.0	1.0	1.0	4.0
components of a defined score	Process	3.7	3.8	3.1	3.5	15	No	No	7.0	7.0	8.0	6.0
Acute Pulmonary Oedema												
gnosis of APO who were	Process	3.7	4.0	3.1	3.5	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of APO who	Process	3.7	4.0	2.6	3.3	11	Yes	Yes	9.0	9.0	9.0	2.0
gnosis of APO who had a	Process	3.7	4.0	2.5	3.2	11	Yes	Yes	9.0	5.0	7.0	4.0
Airway Management												
TI paralytic, following SpO2 > 10% from 0% overall	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	7.0	9.0	9.0	9.0
d by EMS personnel used post ETI	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
d via RSI by EMS gent was administered	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
d by EMS personnel dministered post-ETI	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
d by EMS personnel r was used post-ETI for	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
empted by EMS personnel y inserted as a final airway	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	8.0	9.0	9.0	9.0
empted by EMS personnel erted	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
d by EMS personnel with mmHg post-ETI > 10 mins	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	7.0	9.0	9.0	9.0
was unsuccessful when	Process	3.7	3.8	3.1	3.5	15	Yes	Yes	9.0	9.0	9.0	9.0
was unsuccessful when	Process	3.7	3.8	3.1	3.5	15	Yes	Yes	9.0	9.0	9.0	9.0

was successful when	Process	3.7	3.8	3.1	3.5	15	Yes	Yes	9.0	9.0	9.0	9.0
essfully intubated via RSI	Process	3.7	3.8	3.1	3.5	15	Yes	Yes	9.0	9.0	9.0	9.0
components of the composite Bundle score	Process	3.7	3.8	3.1	3.5	15	No	No	7.0	8.0	9.0	7.0
Anaphylaxis												
gnosis of Anaphylaxis and on documented who st	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of Anaphylaxis and on documented who olinergic bronchodilator	Process	3.7	3.7	3.0	3.3	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of Anaphylaxis tihistamine	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	7.0	9.0	9.0
gnosis of Anaphylaxis ticosteroid	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	6.0	9.0	9.0
gnosis of Anaphylaxis and ponse recorded who were	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	6.0	9.0	9.0	9.0
Asthma/Bronchoconstriction												
gnosis of with lung sounds assessed st treatment)	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of with a SpO2 documented	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of who were administered a	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of who were administered tor	Process	3.7	3.7	3.0	3.3	15	Yes	Yes	8.0	9.0	9.0	9.0
gnosis of who were administered a	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of recorded with silent chest/BP < 90 dministered IM Adrenalin	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	7.0	9.0	9.0	9.0
components of the triction composite bundle	Process	3.7	3.8	3.1	3.5	15	No	No	9.0	9.0	9.0	9.0
Burns												
gnosis of Burns with	Process	3.3	3.8	3.0	3.3	14	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of Burns with body ssessed and recorded	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
General												
es available per defined	Structure	3.7	3.8	3.1	3.5	16	No	No	7.0	8.0	8.0	5.0
oring devices available per od	Structure	3.7	3.8	3.1	3.5	16	No	No	7.0	8.0	8.0	5.0
toring devices available period	Structure	3.7	3.8	2.5	3.1	12	No	No	7.0	8.0	8.0	5.0
ylinders available per od	Structure	3.7	3.8	3.1	3.5	16	No	No	7.0	8.0	8.0	5.0
devices available per od	Structure	3.7	3.8	3.1	3.5	16	No	No	7.0	8.0	8.0	5.0
ators available per defined	Structure	3.7	3.8	2.5	3.1	12	No	No	7.0	8.0	8.0	5.0
consciousness with a	Process	3.7	3.8	3.1	3.5	16	Yes	Yes	9.0	9.0	8.0	9.0
2 < 95% who were xygen	Process	3.7	4.0	3.0	3.4	15	Yes	Yes	9.0	9.0	8.0	9.0
gnosis recorded	Process	3.7	3.8	3.0	3.4	15	No	No	9.0	9.0	9.0	9.0
Hypoglycaemia												

level < 5 mmol who were	Process	3.7	4.0	3.1	3.5	16	Yes	Yes	9.0	9.0	9.0	9.0
level measured and dministration	Process	3.7	4.0	3.1	3.5	16	Yes	Yes	9.0	9.0	9.0	9.0
Neonate/Paediatric												
d and recorded for	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
d and recorded for	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
visional diagnosis of Croup nhaled steroids	Process	3.7	4.0	3.0	3.4	15	Yes	No	9.0	9.0	9.0	9.0
visional diagnosis of Croup ised Adrenalin	Process	3.7	3.8	3.0	3.4	15	Yes	No	9.0	9.0	9.0	9.0
cility with specialist ces	Process	3.3	3.7	2.3	2.9	8	Yes	Yes	7.0	9.0	9.0	7.0
Obstetrics												
r prior to EMS arrival	Process	3.7	4.0	3.0	3.4	15	Yes	Yes	9.0	6.0	8.0	9.0
artum haemorrhage who	Process	3.7	4.0	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
isional diagnosis of ho were administered	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
r during EMS care	Outcome	3.7	4.0	3.0	3.4	15	Yes	Yes	9.0	8.0	9.0	9.0
OHCA												
gnosis of OHCA with a ed	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of OHCA who der CPR	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of OHCA who onic CPR advice	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	7.0	9.0	9.0	3.0
gnosis of OHCA with hm on arrival of EMS	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of OHCA with ng rhythm on arrival of	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of OHCA intubated	Process	3.7	3.8	2.4	3.1	10	Yes	Yes	8.0	9.0	9.0	9.0
gnosis of OHCA for whom ior to arrival at hospital	Process	3.7	3.8	3.1	3.5	15	Yes	Yes	9.0	8.0	9.0	9.0
gnosis of OHCA who were ROSC and Non-ROSC	Process	3.7	4.0	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of OHCA with	Process	3.3	4.0	2.9	3.3	13	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of OHCA with	Process	3.3	4.0	2.9	3.3	13	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of OHCA with over	Process	3.3	4.0	3.0	3.4	14	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of OHCA with discharge	Process	3.3	3.8	1.6	2.6	8	Yes	Yes	7.0	9.0	9.0	2.0
gnosis of OHCA with	Outcome	3.3	3.8	1.6	2.6	8	Yes	Yes	7.0	9.0	9.0	2.0
Pain Management												
asured via defined pain	Process	3.7	3.8	3.1	3.5	16	Yes	Yes	9.0	9.0	9.0	9.0
core threshold who were	Process	4.0	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
asured via defined pain inistration	Process	4.0	3.8	3.1	3.5	16	Yes	Yes	9.0	9.0	9.0	9.0
Seizures												
gnosis of Seizures with a recorded	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of Seizures who eptic for ongoing Seizures	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
Stroke/TIA												

gnosis of Stroke/CVA/TIA d and recorded	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9.0
gnosis of Stroke/CVA/TIA sment performed (e.g.:	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	7.0
gnosis of Stroke/CVA/TIA asurements recorded (X3)	Process	3.7	4.0	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9
gnosis of Stroke/CVA/TIA e Centre	Process	4.0	3.8	1.6	2.8	9	Yes	Yes	7.0	9.0	9.0	1
gnosis of Stroke/CVA/TIA	Process	3.3	3.8	2.3	3.0	8	Yes	Yes	7.0	9.0	9.0	2
components of the osite bundle score	Process	3.7	3.8	3.0	3.4	15	No	No	7.0	7.0	9.0	7
Trauma												
ma case with entrapment	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9
ma case with a BP < 90	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9
ma case with partial/full quet applied	Process	3.7	4.0	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9
ma case with a femur e	Process	3.7	3.8	3.0	3.4	15	Yes	Yes	9.0	9.0	9.0	9
ma case with a BP < 90 d TXA	Process	3.7	3.8	3.1	3.5	15	Yes	Yes	9.0	9.0	9.0	9
ma case with direct Trauma Centre	Process	3.7	4.0	2.3	3.1	9	Yes	Yes	7.0	9.0	9.0	9
Adverse Events												
le in EMS care	Sentinel Event	3.7	4.0	2.4	3.1	10	No	No	7.0	9.0	9.0	8
vents reported during	Sentinel Event	3.7	3.8	2.5	3.1	12	No	No	7.0	9.0	9.0	7
t/technical failures	Sentinel Event	3.7	3.8	2.5	3.1	12	No	No	7.0	9.0	9.0	6
pected extubations	Sentinel Event	3.7	3.8	2.5	3.1	12	No	No	8.0	9.0	9.0	4
crease in GCS of 3 or more	Sentinel Event	3.7	3.8	2.5	3.1	12	No	No	9.0	9.0	9.0	8
bation attempts	Sentinel Event	3.7	4.0	2.5	3.2	11	No	No	9.0	9.0	9.0	6
reports during EMS care	Sentinel Event	3.7	3.8	2.5	3.1	12	No	No	7.0	9.0	9.0	7
injury reports	Sentinel Event	3.7	3.8	2.5	3.1	12	No	No	7.0	9.0	9.0	7
n errors during EMS care	Sentinel Event	3.7	3.8	2.5	3.1	12	No	No	7.0	9.0	9.0	6
Communications/Dispatch												
ith defined ALS Dispatch	Structure	3.7	4.0	3.1	3.5	15	No	No	8.0	8.0	9.0	6
cessing time within	Structure	3.7	3.8	3.1	3.5	15	No	No	9.0	7.0	9.0	7
calls received per 10000	Structure	3.7	4.0	3.0	3.4	15	No	No	7.0	9.0	9.0	5
ed calls to the Service Call	Structure	3.7	4.0	3.1	3.5	15	No	No	7.0	9.0	9.0	6
in dispatch and/or olice/security escort	Process	3.7	4.0	2.6	3.3	11	No	No	8.0	9.0	9.0	7

hreshold of validity score

hreshold of validity score

Table 4: Literature review of evidence base

Indicator Category	Indicator subcategory	Total QIs	Indicator Type				Level of Evidence										Ref
			Structure	Process	Outcome	Sentinel Event	1a	1b	1c	2a	2b	2c	3a	3b	4	5	
Clinical	Acute Coronary Syndromes	25		23	2					4	5				2	14	[1–6]
	Airway management	8		8							2		1		1	2	[2,7–11]
	Acute Pulmonary Oedema	2		2						2							[5]
	Asthma	10		10						1						9	[2,3,11]
	General	18		15	3					2					4	12	[2,6–9,12–19]
	Hypoglycaemia	3		3												3	[3]
	Out of hospital cardiac arrest	44	4	38	2					2					3	39	[2,3,5,7–9,13,18,20–22]
	Pain management	1		1												1	[12]
	Seizures	2		2						2							[11]
	Stroke	11		11							3					8	[3,23–25]
	Trauma	16	3	11	2					4			5			6	[2,5,9,12,19,26]
Non-clinical	Adverse Event	25				25					9			11		5	[7,8,10,14,15,19,27]
	Deployable resources	15	13	2	2										5	13	[18,28,29]
	Dispatch/Call times	90	7	73	6					3	1		26	17	4	39	[2,7–9,12,15,16,19,25,30]
	Documentation	16	3	13								2		2	3	11	[7–9,12,15,18,19,31]
	Employee focused	16	16											2	2	12	[7–9,13,18,29]
	Service user rating/satisfaction	9		6	3										1	8	[13,18,29]
Total		311	46	218	20	25	0	0	0	20	20	2	32	32	25	182	
%			15%	70%	6%	8%	0%	0%	0%	6%	6%	1%	10%	10%	8%	59%	

2a. Systematic review of 2b and better studies
2b. Retrospective cohort study or prospective cohort with poor follow-up/low quality RCT
2c. "Outcomes" Research; Ecological studies
3a. Systematic review of 3b and better studies
3b. Non-consecutive cohort study/Individual case control study
4. Case series
5. Expert opinion

Table 5: Qualitative analysis of the Working Group discussion

Text erence	Sub-category	Supporting Quote
1.1	Relevance	"For me, because practically zero clinical indicators are used or reported publicly by EMS [Emergency Medical Services] in South Africa, their relevan and benefit was naturally going to be scored high"
1.2	Usability	"Whenever I was rating a category that I used or drew information from the data dictionary, there was always sufficient information that left no d planned for or accounted for. The difficult part was knowing how much variation there would be in different EMS organizations in South Africa in hov to extract this information and put it to use"
1.3	Context	"Whatever indicators are used by a service, it's important that they do a feasibility assessment of what's possible for them to achieve. We may be ab these will work for South Africa in general, but when it comes to actual implementation, a service is going to have to understand its surroundings and it sees"
1.4		"Like, the indicators involving direct transport to a CT [Computed Tomography] scanner for Stroke patients, or to PCI [Percutaneous Coronary Inter STEMI [ST Elevation Myocardial Infarction], those will only be applicable to certain metropolitan areas, and probably only for certain private services a general indicator for everyone to use"
1.5	Quality system	"This is a complete mind shift from what we currently know and how we measure quality in South Africa. If a service is serious about implementing few, they're going to have to admit that it's going to take an overhaul in their quality system, and that it's likely going to need more resources than w measuring response times at the moment"
1.6		"Outside of a few of the large private services, the provincial services are going to have to ramp up the effort around measuring quality. As simple a that these indicators are, there's probably not many of the provincial services that are ready to implement them"
2.1	Methodology	"You really get to see how these will be used from a practical point of view. I can see the benefit of how a simple system that's objective can make the It's not like how I used to remember it when we checked the case sheets, and it depended on how you felt at the time"
2.2		"Doing the data extraction made a big difference, because I remember, especially for the sentinel event indicators, I scored them quite low with th when we went through them and applied them to actual cases, it was much simpler than I thought it would be and so I scored them higher after being extraction"
2.3	Technology	"I think applying these indicators would be way easier with an electronic patient report form. It's going to take way more effort in doing it manually, benefits even if it's done this way"
2.4	Quality system	"I think when you're sitting down and applying the indicators to case sheets, the system does seem simple and straightforward enough to use. But w there? It's going to be a logistical challenge to get the paperwork together to do the assessment, but I feel like the bigger challenge is using the infor just as important as getting the information"
2.5	Transparency	"It seems like it's going to be easy to game the system. Like how I know the guys have done the things that they've written down. What sort of mech check that they've been truthful in their notes, especially if they now know they're being watched"
2.6	Technology	"I think [participant] was right about the electronic record, because we can build checks and balances into that sort of thing to monitor truthfulnes [respondent] mentioned. That also solves the legibility issue and whether or not enough information has been written. Look at when we used the examples, it was difficult to apply the indicators to those just because you didn't always have the right information to go on"