Review article

Journal of the Intensive Care Society
0(0) 1–14

DOI: 10.1177/1751143720963757
journals.sagepub.com/home/jics
SAGE

Outcome measurement tools for communication, voice and speech intelligibility in the ICU and their clinimetric properties: A systematic review

Charissa J Zaga[1,2], Bridie Cigognini[1], Adam P Vogel[2,3,4] and Sue Berney[5,6]

Abstract

Purpose: To identify outcome measurement tools used to evaluate communication, voice and speech intelligibility in the mechanically ventilated ICU population. Secondly, to evaluate, synthesise and compare the clinimetric properties of the tools identified.

Materials and methods: A systematic review of articles was undertaken via electronic databases in two parts. Eligibility criteria for selection: part one – quantitative or mixed methods studies which assessed communication, voice or speech intelligibility; part two – studies which evaluated a clinimetric property for one of the tools identified in part one. Two independent reviewers assessed articles for inclusion and used the consensus-based standards for health status measurement instruments (COSMIN) risk of bias checklist.

Results: The part one search yielded five included studies comprised of eight outcome measurement tools. The part two search yielded 22 included studies comprised of nine tools. Few studies had adequate reliability and measurement error properties. No studies established responsiveness. A notable proportion of studies utilised tools that have no clinimetric properties.

Conclusions: There is a relatively small number of studies which have established clinimetric properties for outcome measurement tools that evaluate communication, voice and/or speech intelligibility, and a fewer number which have done so in the mechanically ventilated ICU population.

Keywords

ICU, mechanical ventilation, communication, speech, voice, outcome measurement tool

Introduction

Critically ill patients experience impaired communication, usually due to the presence of an endotracheal tube or tracheostomy tube. About one-third of intensive care unit (ICU) patients experience difficulty communicating[1] and half of the mechanically ventilated patients in ICU meet minimum criteria for communication during sustained periods of wakefulness.[2] Communication interventions have the potential to enable patients to not only express their basic wants and needs, but also to engage in conversations with their loved ones and with health care professionals about their care, such as, consent for procedures and withdrawal of care which is an unfortunate reality of a proportion of ICU admissions. A range of communication interventions have been studied in this population, including communication board, electrolarynx, high-technology augmentative and alternative communication (AAC) device, one-way valve in-line with the ventilator and ventilator-adjusted leak speech

[1]Department of Speech Pathology, Austin Health, Melbourne, Australia
[2]Centre for Neuroscience of Speech, University of Melbourne, Melbourne, Australia
[3]Department of Neurodegeneration, Hertie Institute for Clinical Brain Research, Tübingen, Germany
[4]Redenlab, Melbourne, Australia
[5]Department of Physiotherapy, Austin Health, Melbourne, Australia
[6]Department of Physiotherapy, School of Health Sciences, University of Melbourne, Melbourne, Australia

Corresponding author:
Charissa J Zaga, Austin Health, 145 Studley Road, Heidelberg, Melbourne, Australia.
Email: Charissa.zaga@austin.org.au

(VALS).[3–5] A recent systematic review of the feasibility, utility and safety of communication interventions with mechanically ventilated ICU patients found that while the level of evidence is generally low, there is a promising signal that communication interventions are feasible, have utility and are safe to perform in this population.[3.] However, of the 48 included studies, there was heterogeneity relating to the communication intervention and the outcome measurement tool used. The majority of studies measured outcomes using perceptual or subjective judgements.[3] Higher quality studies are required to establish how best to provide intervention with critically ill patients who are usually voiceless, and need speech pathology intervention to facilitate communication and participation in their healthcare.

Use of outcome measurement tools with established clinimetric properties, such as validity and reliability, is favourable over subjective clinical judgements. A reliable and valid tool facilitates accurate, reproducible and specific measurement of a desired outcome. The utilization of outcome measurement tools in healthcare is essential, for patients, clinicians and health services alike – to identify which interventions are beneficial, which are not, to improve clinical outcomes as well as service delivery and appropriate resource management.[6] For clinicians, an assessment of the quality of outcome measurement tools available provides guidance as to the most reliable and responsive to determine efficacy, progress and utility of given interventions. This is beneficial first and foremost for patients, and additionally for the demonstration of the value-add of speech pathology in the ICU setting.

The communication needs of ICU patients are subject to change in a high-acuity setting, associated with their medical instability and illness severity. Factors such as fluctuating alertness and requirement for sedation, ICU-acquired weakness, fatigue and cognitive profile influence the choice and timing of particular therapeutic interventions. In the absence of objective outcome measurement tools, currently, clinicians, researchers and patients have limited knowledge of which therapeutic interventions add value, minimise deterioration or restore function.

The aims of this systematic review were two-fold: (1) to identify the outcome measurement tools utilised in studies of critically ill mechanically ventilated adults which evaluated communication function and/or speech intelligibility and (2) to evaluate the clinimetric properties of the outcome measurement tools identified in part one.

Materials and methods

Protocol

The review was registered on PROSPERO (CRD42019136852). The search for this systematic review was conducted in two parts. Part one involved the identification of outcome measurement tools which have been used to evaluate communication, voice or speech intelligibility in the mechanically ventilated ICU population. Part two involved a second search to identify studies that examined the clinimetric properties of the measurement tools identified in part one. The review design was consistent with earlier published work in the ICU setting.[7]

Part one – Identification of measures

Medline, Embase, PsycINFO, Scopus and Web of science electronic databases were searched by one reviewer using a systematic, comprehensive and reproducible search strategy, devised with the assistance of a professional university librarian (see Supplementary Figure 1). The search was last run on 9 June 2019. Two independent reviewers determined eligibility against pre-determined criteria (see Table 1). A list of outcome measurement tools was generated from the results of part one.

Part two – Clinimetric properties of outcome measurement tools

Medline, Embase, CINAHL and Web of Science electronic databases were searched by one reviewer using a systematic, comprehensive and reproducible search strategy, devised with the assistance of a professional university librarian (see Supplementary Figure 2) with the last search run on 25 July 2019. The study selection and data extraction followed the same methodology described for part one. Two independent reviewers used relevant items of the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) risk of bias checklist to evaluate the methodological quality and risk of bias of the included studies. The COSMIN checklist is a standardised tool for assessing the individual studies on measurement properties of Patient Reported Outcome Measures (PROMs).[8] Since its original development, the COSMIN checklist was revised for use in systematic reviews of PROMs seeking to assess risk of bias of studies on measurement properties.[9] The COSMIN comprised a box for each measurement property which contains standards to assess the quality of a study on that specific measurement property, of which a score is derived.[8]

Since this systematic review investigated clinician-reported outcome measures, not patient-reported outcome measures, reliability, measurement error, hypothesis testing, criterion validity and responsiveness were the relevant items of the checklist. An overall quality score for each item was obtained by using the lowest score recorded.[9] Agreement between the two reviewers was estimated using percentage agreement and the kappa statistic.[10]

Table 1. Part one: pre-determined inclusion and exclusion criteria.

Characteristics	Inclusion	Exclusion
Design	Quantitative and mixed-methods study designs, randomised control trials, cohort studies, case–control studies, case series	Studies not published in a peer-reviewed journal, descriptive commentary (reviews, editorials, narratives), conference abstracts
Participants	Adults >18 years of age, admitted in the intensive care unit, mechanically ventilated at the time of participation in the study	
Intervention	Communication intervention[a]	
Outcome measurement instruments	Primary outcome measure related to communication, voice and/or speech intelligibility	
Publication	English language No date restrictions	

[a]An intervention which facilitates communication.

Results

Part one – Identification of outcome measurement tools

This search identified five included studies comprised of eight outcome measurement tool (see Figure 1). Seventeen studies met the exclusion criteria (see Supplementary Table 1). Percentage agreement for title and abstracts was 88% ($\kappa = 0.75$) for full-text was 94% ($\kappa = 0.88$). The outcome measurement tools identified from the included studies were the Ease of Communication Scale (ECS)[11,12] ($n = 1$), Therapy Outcome Measure for Voice Impairment (TOMS)[13] ($n = 1$) Adapted TOMS[14] ($n = 1$), ICU Functional Communication Scale (ICU-FCS)[13] ($n = 1$), Electrolarynx Effectiveness Score (EES)[15,16] ($n = 1$), Assessment of Intelligibility of Dysarthric Speakers (AIDS)[15] ($n = 1$), Grade Roughness Breathiness Asthenia and Strain Voice Profile (GRBAS)[13] ($n = 1$) and a Questionnaire[17] ($n = 1$). The characteristics of the included studies are summarised in Table 2.

Part two – Clinimetric properties of outcome measurement tools

Study selection and study characteristics. The aim of this second search was to yield relevant articles which examined the clinimetric properties of the outcome measurement tools that were identified in the part one search. As such, the search terms for part two were only devised at the completion of the part one review. The first iteration of this search did not yield any relevant full-text articles. The search strategy was subsequently revised to specifically include the outcome measurement tools identified in part one, as keyword phrases. This second iteration of the search yielded 5217 records (see Figure 2). Twenty-two studies met the inclusion criteria (Table 3), including nine outcome measurement instruments which included the GRBAS[13,18–28] ($n = 12$), TOM[29] ($n = 1$), TOM-AAC[30] ($n = 1$), The Clinician Interview-Based Impression (CIBI)[31] ($n = 1$), Loewenstein Communication Scale (LCS)[32] ($n = 1$), Questionnaire[17] ($n = 1$), AIDS[15,33,34] ($n = 3$), ICU-FCS[13] ($n = 1$) and EES[15,16] ($n = 2$). Percentage agreement for title and abstracts 90% ($\kappa = 0.78$) and full-text 100% ($\kappa = 1$). The characteristics of the included studies are summarised in Table 4.

Risk of bias results. Percentage agreement for risk of bias assessment of included studies was 98% ($\kappa = 0.823$).

Study results are summarised in Table 5. According to the COSMIN risk of bias checklist, 4/9 (44%) outcome measurement tools utilised in the included studies had 'doubtful' or higher quality scores for at least one of the clinimetric properties evaluated. Five of nine (55%) outcome measurement tools did not have any established clinimetric properties. Six of nine (66%) outcome measurement tools were studied in the ICU setting. Specifically, GRBAS, AAC-TOM and TOM were studied in alternative settings such as outpatient clinics. The highest scored properties were reliability and measurement error. For reliability, 4 of 22 studies were rated 'adequate', 3 'doubtful' and 15 'inadequate'. For measurement error, 4 were 'adequate', 1 'very good', 2 'doubtful' and 15 'inadequate'. Criterion validity was not adequately established in any of the included studies, with only 1 study rated 'very good', 8 'doubtful' and the remaining 13 'inadequate'. Hypothesis testing was rated 'adequate' in only 1 study, 'doubtful' in one study and the remaining 20 were rated 'inadequate'. None of the 22 included studies established responsiveness, with 'inadequate' ratings for all. The GRBAS scale scored the highest for reliability and measurement error overall. No clinimetric properties were established for 5/9 (55%) of tools (AAC-TOM, AIDS, EES, ICU-FCS, or the Questionnaire[17]). The quality

Table 2. Studies which met inclusion criteria.

Author(s)	Design	Methodology	Intervention	Outcome measure	Instrument
Happ et al. (2014)	Quasi-experimental 3 phase sequential cohort study	Mixed-methods	1) Usual care 2) Basic communication skills training for nurses 3) Additional training in AAC devices and SP consultation	Frequency, quality, success and ease of communication	Adapted Ease of Communication Scale (ECS)
Maringelli et al. (2013)	Cohort	Quantitative	Electronic Basic Communication Table with TheGrid2 AAC package (Gaze-controlled communication system)	Ability to communicate/interact, the ability to understand others, the perceived anxiety and depression in relation to the communication deficits and genera evaluation of the importance of communication in ICU	Questionnaire
McGrath et al. (2016)	Case series	Quantitative	Above cuff vocalisation	Audible voice	Adapted TOMS
McGrath et al. (2019)	Feasibility study	Quantitative	Above cuff vocalisation	Audible voice	GRBAS Scale; Therapy Outcome Measure for Voice Impairment (TOMS); ICU Functional Communication Scale (FCS)
Rose et al. (2018)	Prospective single group feasibility study in 3 units	Quantitative	Electrolarynx	Feasibility and successful communication (defined as the ability to generate intelligible and comprehensible speech)	Electrolarynx Effectiveness Score (EES); Assessment of Intelligibility of Dysarthric Speech (AIDS)

AAC: augmentative and alternative communication; SP: speech pathologist.

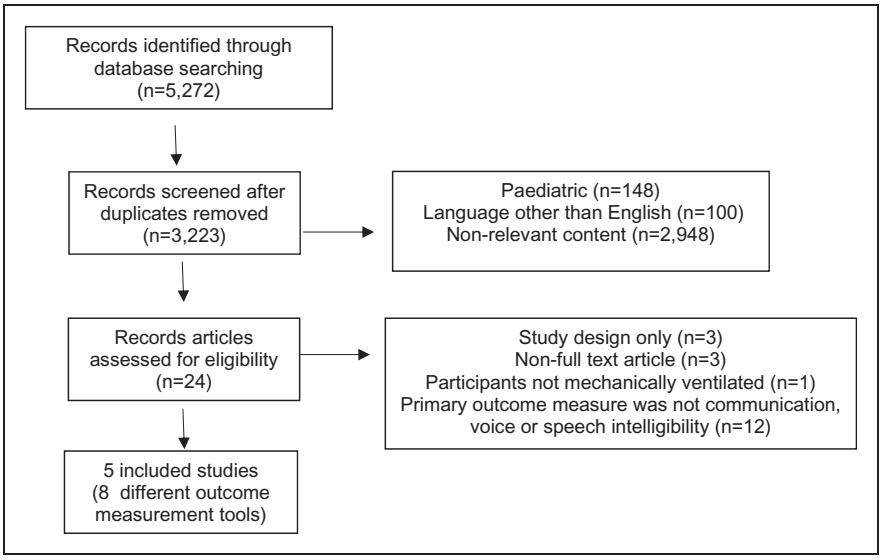

Figure 1. Part one: PRISMA diagram. PRISMA: preferred reporting items for systematic reviews and meta-analyses.

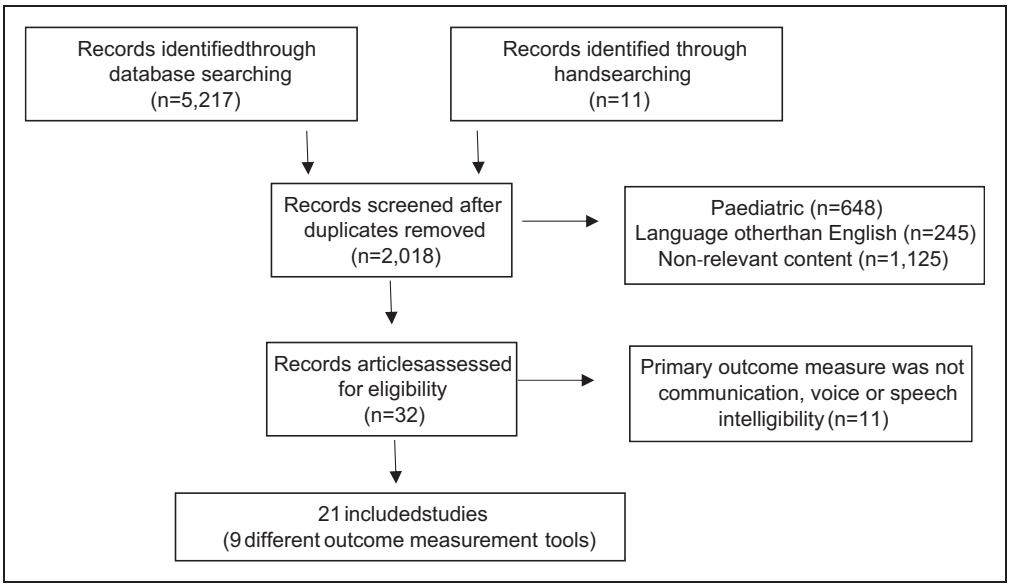

Figure 2. Part two: PRISMA diagram. PRISMA: preferred reporting items for systematic reviews and meta-analyses.

Table 3. Part two: pre-determined inclusion and exclusion criteria.

Characteristics	Inclusion	Exclusion
Design	Quantitative and mixed-methods study designs, randomised control trials, cohort studies, case–control studies, case series	Studies not published in a peer-reviewed journal, descriptive commentary (reviews, editorials, narratives), conference abstracts
Participants	Adults >18 years of age in the Intensive Care Unit, mechanically ventilated at the time of participation in the study	
Intervention	Did not form part of the eligibility criteria	
Outcome measurement instruments	Primary outcome measure related to communication, voice and/or speech intelligibility; assessment of clinimetric properties of an outcome measurement instrument identified in part one	
Publication	English language No date restrictions	

Table 4. Part two: characteristics of included studies.

Authors	Location	No. of participants	Study setting	Tool(s)	Gender	Mean age (range/SD)	Airway	ICU LOS	Illness severity	Intervention
Yorkston and Beukelman (1981)	USA	$n=14$	Clinic	AIDS	Unknown	Unknown	Normal	N/A	N/A	Speech samples
Sparker et al. (1987)	USA	$n=23$	Majority in ICU or intermediate care unit	AIDS	15 M 8 F	45 years (14–78)	Tracheostomy	Unknown	Unknown	Speaking tracheostomy tube
Rose et al. (2018)	Canada	$n=24$	2 × ICU 1 × specialised weaning unit	AIDS, EES	15 M 9 F	62 years (13.9)	Tracheostomy	93 days (38–140)	Unknown	Electrolarynx
Nilsen et al. (2014)	USA	$n=5$	ICU	CIBI	Unknown	$\geqslant 60$ years	Intubated	Unknown	Unknown	Video recorded observations of interactions
Tuinman et al. (2015)	USA	$n=15$	ICU	EES	7 M 8 F	57 years (Unknown)	Intubated $n=13$ Tracheostomy $n=2$	Unknown	Mean APACHE II = 23	Electrolarynx
Maringelli et al. (2013)	Italy	$n=15$	ICU	Questionnaire	8 M 7 F	(50–65)	Intubated $n=7$ Tracheostomy $n=8$	>2 weeks	Unknown	Eye gaze AAC device
McGrath et al. (2019)	UK	$n=10$	1 × Cardiothoracic ICU 1 × General ICU	GRBAS, ICU FCS, TOM	7 M 3 F	60 years (28–83)	Tracheostomy	28 (22–132)	Mean APACHE II = 18	Above cuff vocalisation
Bhuta et al. (2004)	USA	$n=37$	Clinic	GRBAS	Unknown	Unknown	Normal	N/A	N/A	Voice samples
Brinca et al. (2015)	Portugal	$n=3$ normal speakers $n=27$ dysphonic speakers voice recordings from $n=15$ used	Clinic	GRBAS	All female	44 years (20–72)	Normal	N/A	N/A	Voice samples
De Bodt et al. (1997)	Belgium	$n=9$	Clinic	GRBAS	3 M 6 F	(22–65)	Normal	N/A	N/A	Voice samples
Karnell et al. (2006)	USA	$n=103$ voice recordings from $n=34$ used	Clinic	GRBAS	42 M 61 F	(17–90)	Normal	N/A	N/A	Voice samples
Hakkesteegt et al. (2008)	The Netherlands	$n=294$ dysphonic speakers $n=118$ normal speakers	Clinic	GRBAS	Dysphonic: 98 M 196 F Controls: 49 M 69 F	Dysphonic: 44 years (14–87) Controls: 44 years (20–79)	Normal	N/A	N/A	Voice samples

(continued)

Table 4. Continued

Authors	Location	No. of participants	Study setting	Tool(s)	Gender	Mean age (range/SD)	Airway	ICU LOS	Illness severity	Intervention
Lu and Mateson (2014)	USA	*n* = 39 dysphonic *n* = 21 normal speakers	Clinic	GRBAS	Dysphonic: 15 M 24 F Controls: 1 M 20 fF	Dysphonic: 32 years (2.7) (18–81) Controls: 25 years (0.9) (21–35)	Normal	N/A	N/A	Voice samples
Millet et al. (1998)	The Netherlands	*n* = 65	Clinic	GRBAS	Unknown	Unknown	Normal	N/A	N/A	Voice samples
Nemr et al. (2012)	Brazil	*n* = 50 dysphonic *n* = 10 normal speakers	Clinic	GRBAS	Unknown	Unknown	Normal	N/A	N/A	Voice samples
Saenz-Lechon et al. (2011)	USA	*n* = 53 normal speakers *n* = 173 dysphonic speakers	Clinic	GRBAS	Controls: 21 M 32 fF Dysphonic: 70 M 103 F	Male (26–59) Female (21–52)	Normal	N/A	N/A	Voice samples
Webb et al. (2003)	UK	*n* = 65 dysphonic speakers *n* = 5 normal speakers	Clinic	GRBAS	Dysphonic: 30 M 35 F Controls: 3 M 2 F	Dysphonic (18–87) Controls: (26–56)	Normal	N/A	N/A	Voice samples
Wuyts et al. (1998)	Belgium	*n* = 14	Clinic	GRBAS	7 M 7 F	(7–65)	Normal	N/A	N/A	Voice samples
Borer-Alafi et al. (2002)	Israel	*n* = 42	ICU and rehabilitation hospital	LCS	27 M 15 F	31 years (17–72) (13.86)	Unknown	236 days (26–632) (153.68)	Mean GCS 4.74 (1–8)	N/A
John et al. (2000)	UK	*n* = 73 speech pathologists, different patient groups	Clinic setting	TOM	N/A	N/A	Normal	N/A	N/A	N/A
Enderby (2014)	UK	N/A	N/A	TOM-AAC	N/A	N/A	Normal	N/A	N/A	N/A

AAC: augmentative and alternative communication; AIDS: Assessment of Intelligibility of Dysarthric Speakers; APACHE: Acute Physiology and Chronic Health Evaluation; CIBI: The Clinician Interview-Based Impression; EES: Electrolarynx Effectiveness Score; FCS: Functional Communication Scale; GCS: Glasgow Coma Scale; GRBAS: Grade Roughness Breathiness Asthenia and Strain Voice Profile; LCS: Loewenstein Communication Scale; TOM: Therapy Outcome Measure for Voice Impairment.

Table 5. Synthesised evaluation of clinimetric properties of outcome measurement tools.

Tool	No. of studies	Author(s)	Reliability	Measurement error	Criterion validity	Hypotheses testing	Responsiveness	Findings
AAC-TOM	$n = 1$	Enderby (2014)	Inadequate	Inadequate	Doubtful	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✗ Clinimetric properties established
AIDS	$n = 3$	Yorkston and Beukelman (1981)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✓ Pearson product-moment correlations and Kendall rank-order correlations calculated ✗ Percentage agreement between judges calculated ✓ Comparison of subgroups ✗ Comparator instrument ✗ Intervention
		Sparker et al. (1987)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✓ Used with mechanically ventilated ICU patients ✗ Clinimetric properties established
		Rose et al. (2018)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✓ Used with mechanically ventilated ICU patients ✗ Clinimetric properties established
CIBI	$n = 1$	Nilsen et al. (2014)	Doubtful	Doubtful	Inadequate	Inadequate	Inadequate	✓ Used with mechanically ventilated ICU patients ✗ Comparator instrument ✗ Comparison of subgroups ✗ Intervention
EES	$n = 2$	Tuinman et al. (2015)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✓ Used with mechanically ventilated ICU patients ✗ Clinimetric properties established
		Rose et al. (2018)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✓ Used with mechanically ventilated ICU patients ✗ Clinimetric properties established

(continued)

Table 5. Continued

Tool	No. of studies	Author(s)	Reliability	Measurement error	Criterion validity	Hypotheses testing	Responsiveness	Findings
GRBAS	$n = 11$	Bhuta et al. (2004)	Doubtful	Inadequate	Doubtful	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✗ Inter or intra-rater reliability measures ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups
		Brinca et al. (2015)	Adequate	Adequate	Doubtful	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups
		De Bodt et al. (1997)	Doubtful	Adequate	Doubtful	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups
		Hakkesteegt et al. (2008)	Inadequate	Inadequate	Very good	Adequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups
		Karnell et al. (2007)	Adequate	Adequate	Inadequate	Doubtful	Inadequate	✗ Used with mechanically ventilated ICU patients ✓ Pearson correlation calculated for reliability ✓ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups

(continued)

Table 5. Continued

Tool	No. of studies	Author(s)	Reliability	Measurement error	Criterion validity	Hypotheses testing	Responsiveness	Findings
		Lu and Mateson (2014)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✓ Normal controls ✗ Kappa agreement calculated ✓ Used Agreement Coefficient for reliability measure ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups
		Millet et al. (1998)	Inadequate	Inadequate	Doubtful	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✓ Pearson correlation calculated ✗ Kappa agreement calculated ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups
		Nemr et al. (2012)	Inadequate	Inadequate	Doubtful	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✓ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups
		Saenz-Lechon et al. (2011)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups
		Webb et al. (2004)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups
		Wuyts et al. (1999)	Adequate	Very good	Doubtful	Inadequate	Inadequate	✗ Used with mechanically ventilated ICU patients ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups

(continued)

Table 5. Continued

Tool	No. of studies	Author(s)	Reliability	Measurement error	Criterion validity	Hypotheses testing	Responsiveness	Findings
ICU FCS	$n = 1$	McGrath et al. (2019)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✓ Used with mechanically ventilated ICU patients ✗ Clinimetric properties established
LCS	$n = 1$	Borer-Alafi et al. (2002)	Adequate	Adequate	Inadequate	Inadequate	Inadequate	✓ Used with mechanically ventilated ICU patients ✓ Adequate reliability and measurement error properties ✓ Comparison made between subgroups ✓ Predictive validity measured ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention
Questionnaire	$n = 1$	Maringelli et al. (2013)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate	✓ Used with mechanically ventilated ICU patients ✗ Clinimetric properties established
TOM	$n = 1$	John and Enderby (2000)	Inadequate	Doubtful	Doubtful	Inadequate	Inadequate	✓ Used with mechanically ventilated ICU patients ✗ Comparator instrument ✗ Comparison to gold standard ✗ Intervention ✗ Comparison of subgroups

AAC: augmentative and alternative communication; AIDS: Assessment of Intelligibility of Dysarthric Speakers; CIBI: The Clinician Interview-Based Impression; EES: Electrolarynx Effectiveness Score; FCS: Functional Communication Scale; GRBAS: Grade Roughness Breathiness Asthenia and Strain Voice Profile; LCS: Loewenstein Communication Scale; TOM: Therapy Outcome Measure for Voice Impairment.

ratings were based on clearly outlined questions in the COSMIN risk of bias checklist, per clinimetric property, including design requirements, statistical methods and other.[9]

Discussion

Part one of this systematic review identified just five studies in which communication was a primary outcome. The main finding of part two was that the majority of the identified outcome measurement tools either do not have established clinimetric properties or have not been examined in the critical ill population who have an artificial airway. To the authors knowledge, this is the first systematic review which has examined the clinimetric properties of outcome measurement tools used to measure communication, voice and speech intelligibility in the ICU. As such, comparison with earlier work is not possible.

While currently without the examination of clinimetric properties, the ICU-FCS appears promising as an outcome measurement instrument for the verbal ICU patient, whether phonation be achieved by above cuff vocalisation,[13,14] VALS[35,36] or one-way valve in-line with the ventilator.[37,38] The LCS was studied in patients who suffered traumatic brain injury in order to differentiate minimally responsive patients from those in a vegetative state. The LCS was found to have 'adequate' reliability and measurement error according to the risk of bias assessment, and the LCS scores generated were found to predict individual communication rehabilitation potential ($p = 0.002$) which indicates a strong signal. The generalisability of this tool in conscious patients without traumatic brain injury or neurological impairment is currently unknown. The included studies identified outcome measurement tools with sound clinical bases; however, very little is known about their properties and this suggests questionable accuracy of their use in this population overall. Establishing clinimetric properties of an outcome measurement tool are essential, to ensure that the results of measurement are accurate, reproducible and consistent.

The strength of this review was the systematic and reproducible search strategy which was devised in conjunction with a senior academic librarian from the University of Melbourne. No limits were applied to the publication date within the search which enabled relevant articles to be included as early as 1981. The data extraction process and the utilisation of a high-quality risk of bias assessment tool which have been applied to other reviews of tools in the ICU setting.[7] While every effort was taken to maintain the scientific rigour of this systematic review, search strategy was limited to English only and as a result some articles may have been missed. The COSMIN risk of bias checklist dictates users to rate the overall quality based on the lowest score and, there may be instances where tools were underrated as a result. Lastly, the authors did not attempt to contact authors of the included studies for missing data or clarification thereof, which may have contributed to the risk of bias checklist scoring.

Recommended future directions are three-fold: (1) to establish clinimetric properties for outcome measurement tools that currently lack these, e.g. ICU-FCS (2) to adapt and examine outcome measurement tools in the ICU population, e.g. GRBAS or (3) develop an outcome measurement tool that has robust clinimetric properties and considers the ICU environment and critically ill population. A reliable and sensitive tool would facilitate accurate measurement of the efficacy of speech pathology interventions in the critically ill and drive improvement in clinical outcomes in this highly vulnerable population.

Conclusion

This systematic review found few outcome measurement tools that have been used to assess communication outcomes with mechanically ventilated patients in ICU. Furthermore, the quality of these tools in the ICU setting has not been widely established. The adaptation and examination of existing tools in the ICU setting or the development and testing of a comprehensive outcome measurement tool evaluating communication, speech intelligibility or voice of critically ill patients would be beneficial for patients, clinicians and researchers.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Charissa J Zaga https://orcid.org/0000-0002-6717-908X
Adam P Vogel https://orcid.org/0000-0002-3505-2631

Supplemental material

Supplemental material for the article is available online.

References

1. Freeman-Sanderson AL, Togher L, Elkins M, et al. Quality of life improves for tracheostomy patients with return of voice: a mixed methods evaluation of the patient experience across the care continuum. *Intensive Crit Care Nurs* 2018; 46: 10–16.

2. Happ MB, Seaman JB, Nilsen ML, et al. The number of mechanically ventilated ICU patients meeting communication criteria. *Heart Lung* 2015; 44: 45–49.
3. Zaga C, Berney S, and Vogel AP. The feasibility, utility, and safety of communication interventions with mechanically ventilated intensive care unit patients: a systematic review. *Am J Speech-Lang Pathol* 2019; 28: 1335–1365.
4. Ten Hoorn S, Elbers PW, Girbes AR, et al. Communicating with conscious and mechanically ventilated critically ill patients: a systematic review. *Crit Care* 2016; 1–14. DOI 10.1186/s13054-016-1483-2.
5. Carruthers H, Astin F, and Munro W. Which alternative communication methods are effective for voiceless patients in intensive care units? A systematic review. *Intensive Crit Care Nurs* 2017; 42: 88–96.
6. Sansoni J. *Health outcomes: an overview from an Australian perspective*. Australian Health Outcomes Collaboration, Australian Health Services Research Institute, University of Wollongong, August 2016.
7. Parry S, Granger CL, Berney S, et al. Assessment of impairment and activity limitations in the critically ill: a systematic review of measurement instruments and their clinimetric properties. *Intensive Care Med* 2015; 41: 744–762.
8. Mokkink L, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19: 539–549.
9. Mokkink L, de Vet HCW, Prinsen CA, et al. COSMIN Risk of Bias checklist for systematic review of patient-reported outcome measures. *Qual Life Res* 2018; 27: 1171–1179.
10. McHugh M. Interrater reliability: the kappa statistic. *Biochem Med* 2012; 22: 276–282.
11. Menzel LK. A comparison of patients' communication-related responses during intubation and after extubation. *Heart Lung* 1997; 26: 363–371.
12. Menzel LK. Factors related to the emotional responses of intubated patients to being unable to speak. *Heart Lung* 1998; 27: 245–252.
13. McGrath BA, Wallace S, Wilson M, et al. Safety and feasibility of above cuff vocalisation for ventilator-dependant patients with tracheostomies. *J Intensive Care Soc*. Epub ahead of print 2018. DOI: 10.1016/j.bja.2017.11.059.
14. McGrath B, Lynch J, Wilson M, et al. Above cuff vocalisation: a novel technique for communication in the ventilator-dependent tracheostomy patient. *J Intensive Care Soc* 2016; 17: 19–26.
15. Rose L, Istanboulian L, Smith OM, et al. Feasibility of the electrolarynx for enabling communication in the chronically critically ill: the EECCHO study. *J Crit Care* 2018; 47: 109–113.
16. Tuinman P, ten Hoorn S, Aalders YJ, et al. The electrolarynx improves communication in a selected group of mechanically ventilated critically ill patients: a feasibility study. *Intensive Care Med*. Epub ahead of print 2015. DOI 10.1007/s00134-014-3591-2.
17. Maringelli F, Brienza N, Scorrano F, et al. Gaze-controlled, computer-assisted communication in Intensive Care Unit: "speaking through the eyes". *Minerva Anestesiol* 2013; 79: 165–175.
18. Lu FM, and Matteson S. Speech tasks and interrater reliability in perceptual voice evaluation. *Voice* 2014; 28: 725–732.
19. Nemr K, Simoes-Zenari M, Cordeiro GF, et al. GRBAS and Cape-V scales: high reliability and consensus when applied at different times. *J Voice* 2012; 26: 812.e17–22.
20. Bhuta T, Patrick L, and Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice* 2004; 18: 299–304.
21. Brinca L, Batista AP, Tavares AI, et al. The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli. *J Voice* 2015; 29: 776.e7–14.
22. De Bodt MS, Wuyts FL, Van de Heyning PH, et al. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice* 1997; 11: 74–80.
23. Hakkesteegt MM, Brocaar MP, Wieringa MH, et al. The relationship between perceptual evaluation and objective multiparametric evaluation of dysphonia severity. *J Voice* 2008; 22: 138–145.
24. Karnell M, Melton SD, Childes JM, et al. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *Voice* 2006; 21: 576–590.
25. Millet B, and Dejonckere PH. What determines the differences in perceptual rating of dysphonia between experienced raters? *Folia Phoniatr Logop* 1998; 50: 305–310.
26. Saenz-Lechon N, Fraile R, Godino-Llorente JI, et al. Towards objective evaluation of perceived roughness and breathiness: an approach based on mel-frequency cepstral analysis. *Logoped Phoniatr Vocol* 2011;36(2):52-9.
27. Webb AL, Carding PN, Deary IJ, et al. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Oto-Rhino-Laryngol* 2004; 261: 429–434.
28. Wuyts FL, De Bodt MS, and Van de Heyning PH. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice* 1999; 13: 508–517.
29. John AE, and Enderby P. Realiability of speech and language therapists using therapy outcome measures. *Int J Lang Commun Disord* 2000; 35: 287–302.
30. Enderby P. Introducing the therapy outcome measure for AAC services in the context of a review of other measures. *Disabil Rehabil* 2014; 9: 33–40.
31. Nilsen ML, Happ MB, Donovan H, et al. Adaptation of a communication interaction behavior instrument for use in mechanically ventilated, nonvocal older adults. *Nurs Res* 2014; 63: 3–13.
32. Borer-Alafi N, Gil M, Sazbon L, et al. Loewenstein Communication Scale for the minimally responsive patient. *Brain Inj* 2002; 16: 593–609.

33. Yorkston KB, and Beukelman DR. Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. *J Speech Hear Disord* 1981; 46: 296–301.
34. Sparker A, Robbins KT, Nevlud GN, et al. A prospective evaluation of speaking tracheostomy tubes for ventilator dependent patients. *Laryngoscope* 1987; 97: 89–92.
35. Garguilo M, Leroux K, Lejaille M, et al. Patient-controlled positive end-expiratory pressure with neuromuscular disease: effect on speech in patients with tracheostomy and mechanical ventilation support. *Chest* 2013; 143: 1243–1251.
36. Hoit J, Banzett RB, Lohmeier HL, et al. Clinical ventilator adjustments that improve speech. *Chest* 2003; 124: 1512–1521.
37. Freeman-Sanderson AL, Togher L, Elkins MR, et al. Return of voice for ventilated tracheostomy patients in ICU: a randomized controlled trial of early-targeted intervention. *Crit Care Med* 2016; 44: 1075–1081.
38. Sutt A, Cornwell P, Mullany D, et al. The use of tracheostomy speaking valves in mechanically ventilated patients results in improved communication and does not prolong ventilation time in cardiothoracic intensive care unit patients. *J Crit Care* 2015; 30: 491–494.