Validation of a classification system for causes of death in critical care: an assessment of inter-rater reliability

Elliott Ridgeon, Rinaldo Bellomo, John Myburgh, Manoj Saxena, Mark Weatherall, Rahi Jahan, Dilshan Arawwawala, Stephanie Bell, Warwick Butt, Julie Camsooksai, Coralie Carle, Andrew Cheng, Emanuel Cirstea, Jeremy Cohen, Julius Cranshaw, Anthony Delaney, Glenn Eastwood, Suzanne Eliott, Uwe Franke, Dashiell Gantner, Cameron Green, Richard Howard-Griffin, Deborah Inskip, Edward Litton, Christopher MacIsaac, Amanda McCairn, Tushar Mahambrey, Parvez Moondi, Lynette Newby, Stephanie O'Connor, Claire Pegg, Alan Pope, Henrik Reschreiter, Brent Richards, Megan Robertson, Helen Rodgers, Yahya Shehabi, Ian Smith, Julie Smith, Neil Smith, Anna Tilsley, Christina Whitehead, Emma Willett, Katherine Wong, Claudia Woodford, Stephen Wright and Paul Young (ICU-DECLARE Investigators)

Mortality is a standard outcome measure in clinical trials involving critically ill patients, and cause-specific mortality, as well as overall mortality, is commonly reported. Reporting a specific cause of death may provide mechanistic insights to interpret observed differences in mortality. Although cause-specific mortality is commonly reported, there are currently no validated systems to classify the cause of death in critically ill adults.[1,2] Establishing such a system may improve the comparability and standardisation of outcome measures in trials, and improve the understanding of treatment effectiveness, by promoting summaries of effectiveness in systematic reviews and meta-analyses.[3,4]

Our aim was to assess the inter-rater reliability of a novel system (the Intensive Care Unit Deaths Classification and Reason [ICU-DECLARE] system) designed to categorise the cause of death in patients who died in an ICU, based on previously used systems.[5-7] We hypothesised that this system would provide substantial agreement between raters, defined as an index of agreement (kappa) > 0.6,[8] and that such agreement would be consistent between health care professionals from different disciplines and countries.

Methods

Case scenario development

One hundred clinical scenarios were developed from the discharge summaries of medical records of patients who died consecutively in a single ICU over a period of 3 years (2012–2015), and whose cases were without ongoing interest from the coroner (see Supplementary Appendix 1; Supplementary Appendices are online at cicm.org.au/Resources/Publications/Journal). Ten scenarios were assessed for usability in a pilot study at a single institution, with six clinicians (see Supplementary Appendix 2).

Selection of raters

A convenience sample of 40 raters (20 ICU specialists and 20 research coordinators) was recruited after an email advertisement to clinical trial group and research centre mailing lists. Within these groups, we prospectively

ABSTRACT

Objective: Trials in critical care have previously used unvalidated systems to classify cause of death. We aimed to provide initial validation of a method to classify cause of death in intensive care unit patients.

Design, setting and participants: One hundred case scenarios of patients who died in an ICU were presented online to raters, who were asked to select a proximate and an underlying cause of death for each, using the ICU Deaths Classification and Reason (ICU-DECLARE) system. We evaluated two methods of categorising proximate cause of death (designated Lists A and B) and one method of categorising underlying cause of death. Raters were ICU specialists and research coordinators from Australia, New Zealand and the United Kingdom.

Main outcome measures: Inter-rater reliability, as measured by the Fleiss multirater kappa, and the median proportion of raters choosing the most likely diagnosis (defined as the most popular classification choice in each case).

Results: Across all raters and cases, for proximate cause of death List A, kappa was 0.54 (95% CI, 0.49–0.60), and for proximate cause of death List B, kappa was 0.58 (95% CI, 0.53–0.63). For the underlying cause of death, kappa was 0.48 (95% CI, 0.44–0.53). The median proportion of raters choosing the most likely diagnosis for proximate cause of death, List A, was 77.5% (interquartile range [IQR], 60.0%–93.8%), and the median proportion choosing the most likely diagnosis for proximate cause of death, List B, was 82.5% (IQR, 60.0%–92.5%). The median proportion choosing the most likely diagnosis for underlying cause was 65.0% (IQR, 50.0%–81.3%). Kappa and median agreement were similar between countries. ICU specialists showed higher kappa and median agreement than research coordinators.

Conclusions: The ICU-DECLARE system allowed ICU doctors to classify the proximate cause of death of patients who died in the ICU with substantial reliability.

Crit Care Resusc 2016; 18: 50-54

Table 1. Options for selection from the two lists of proximate causes of death in the ICU-DECLARE system (raters chose one from each list)

List A. Eight options	List B. Five options
Neurological	Cerebral
Arrhythmia	Cardiac
Cardiogenic shock	Bleeding
Distributive (septic) shock	Sepsis
Hypovolaemic shock	Other
Hypoxic respiratory failure	
Metabolic	
Other	

ICU-DECLARE = Intensive Care Unit Deaths Classification and Reason.

determined that we would recruit half from Australia and New Zealand, and half from the United Kingdom. No more than one ICU specialist and one research coordinator was selected from a single institution, to maintain independence of raters, as far as possible, when assessing cases.

Classification system

The ICU-DECLARE system required raters to select two proximate causes of death (Table 1) and one underlying (Table 2) cause of death. Lists of causes were adapted from those used in previous critical care trials,[5-7] with adjustments to make the lists as broadly applicable as possible to cases, and simple to use. These lists were not specifically related to formal death certification processes or coding practices (eg, International Classification of Diseases, 10th revision).

The proximate cause of death was defined as the organ system failure that was considered to have primarily led to death (eg, cardiac, neurological or metabolic causes, or bleeding or sepsis). For the proximate cause of death, raters chose one option from each of List A, which had eight options, and List B, which had five options.

The underlying cause of death was defined as the specific medical diagnosis that precipitated the proximate cause of death. For the underlying cause of death, raters chose one option from a list of 44 options.

Sample size

To determine sample size when using kappa, a null hypothesis value of kappa is required, along with a "goal" kappa. We set these at 0.4 and 0.7, respectively, and used 80% power. With these parameters, a sample size of 74 can be used.[9] Treating this as a minimum, we used a sample of 100 cases. Kappa typically increases with an increasing number of raters, and to ensure broad applicability and validity we used a convenience sample of 40 raters.

Our study received ethics approval from the New Zealand Central Health and Disability Ethics Committee (15/CEN/42).

Statistical analysis

There was no gold standard for the proximate or underlying cause of death in the case scenarios. To describe the distribution of the selections for each of the three lists, summarised across the case scenarios, the cause that was chosen by the largest proportion of raters was designated as the most likely actual diagnosis.

Table 2. Options for underlying cause of death in the ICU-DECLARE system (raters chose one from the list of 44 options)

TBI (unsurvivable primary injury)	Aortic valve disease	Pulmonary fibrosis
TBI (refractory intracranial pressure)	Mitral valve disease	Pneumonia
Haemorrhagic stroke	Pericardial tamponade	Aspiration pneumonitis
Ischaemic stroke	Sepsis with multiorgan failure	ARDS (pulmonary trigger)
Meningoencephalitis	Haemorrhage due to trauma	ARDS (non-pulmonary trigger)
Cerebral abscess	Haemorrhage not due to trauma	Pulmonary haemorrhage
Status epilepticus	Hepatic failure	Other respiratory
Hypoxic brain injury	Anaphylaxis	Vasculitis
Aneurysmal subarachnoid haemorrhage	Pancreatitis	Drug overdose
Metabolic encephalopathy	Massive pulmonary embolism	Anorexia/cachexia
Other neurological	Other cardiovascular disease	Diabetes
Acute myocardial infarction	Chronic obstructive pulmonary disease	Hypoadrenalism
Myocarditis	Cancer	Drug induced
Ruptured or leaking abdominal aortic aneurysm	Asthma	Renal failure
Ruptured or leaking thoracic aneurysm	Other cause not listed	

ICU-DECLARE = Intensive Care Unit Deaths Classification and Reason. TBI = traumatic brain injury. ARDS = acute respiratory distress syndrome.

Table 3. Kappa for agreement and percentage of raters choosing the majority diagnosis for proximate and underlying cause-of-death lists

	List A proximate cause		List B proximate cause		Underlying cause	
Raters	Kappa (95% CI)	Majority median, % (IQR)	Kappa (95% CI)	Majority median, % (IQR)	Kappa (95% CI)	Majority median, % (IQR)
All	0.54 (0.49–0.60)	77.5 (60.0–93.8)	0.58 (0.53–0.63)	82.5 (60.0–92.5)	0.48 (0.44–0.53)	65.0 (50.0–81.3)
Countries						
Australia and New Zealand	0.57 (0.52–0.62)	80 (65–95)	0.57 (0.53–0.63)	85 (65–95)	0.49 (0.45–0.54)	67.5 (52.5–85.0)
United Kingdom	0.52 (0.47–0.58)	75 (55–95)	0.58 (0.53–0.63)	85 (60–95)	0.47 (0.42–0.52)	65 (50–85)
Groups						
ICU specialists	0.64 (0.58–0.69)	90 (65–100)	0.66 (0.61–0.72)	90 (75–100)	0.55 (0.50–0.60)	75 (55–85)
Research coordinators	0.49 (0.44–0.54)	70 (55–90)	0.53 (0.49–0.58)	80 (63.5–90)	0.43 (0.39–0.47)	60 (45–80)

IQR = interquartile range. ICU = intensive care unit.

The proportion of raters who agreed with the most likely actual diagnosis was calculated for each list and for each of the case scenarios. The percentage of raters making these majority choices are summarised for each list (A, B and underlying). We report median values with interquartile ranges (IQRs).

The Fleiss multirater kappa statistic, together with bootstrap confidence limits, was used to estimate inter-rater agreement for the proximate causes of death and the underlying cause of death for all 40 raters across all 100 cases. We report kappa values with 95% confidence intervals.

We classified a kappa of 0.21–0.4 as fair agreement between raters; 0.41–0.6 as moderate agreement between raters; 0.61–0.8 as substantial agreement between raters; and > 0.8 as almost perfect agreement between raters.[8]

We treated missing data (failure to make a choice of proximate or underlying cause of death) as a separate category but also conducted a sensitivity analysis by excluding the missing data.

We conducted pre-planned subgroup analysis by health care profession (ICU specialists v research coordinators) and by country (Australia and New Zealand v the United Kingdom).

We used SAS, version 9.3 (SAS Institute), and the R statistical software (www.R-project.org), packages irr and boot.

Results

Across all raters and cases, agreement measured by kappa was moderate for proximate and underlying cause-of-death classifications (Table 3).

For the subgroups, there was no difference between countries, but ICU specialists showed greater agreement than research coordinators for proximate cause of death (Lists A and B) and underlying cause of death when measured by kappa. There was substantial agreement between ICU specialists across all three lists when measured by the percentage choosing the most likely diagnosis, and for proximate cause of death (Lists A and B) when measured by kappa.

The sensitivity analysis excluding missing data did not alter the results (Supplementary Appendix 3).

Discussion

Key findings

Using clinical case scenarios, and testing previously used but unvalidated tools with a group of ICU specialists and research coordinators who were involved in clinical research, there was substantial agreement between ICU specialists and moderate agreement between research coordinators in categorising the organ system leading to death and the underlying medical diagnosis causing death.

Relationship to previous studies

Cause-of-death classification systems in neonatology have been assessed for validity, using kappa, percentage agreement and other measures.[10] These systems report agreement ranging from 60% to 96%, which is comparable to the degree of agreement observed between ICU specialists in our study for proximate cause of death. However, the proportion of ICU specialists in our study who agreed with the majority diagnosis for underlying cause was 75%, which is lower than a previous neonatology study.[10] One explanation for this is that there may be a small number of causes of death in neonatology and that inter-rater reliability predictably increases with a decreasing number of available categories.[11] Other discrete diseases, such as stroke,[12] chronic obstructive pulmonary disease[13] and cancer surgery,[14] also have fewer categories for cause

of death. A lower rate of agreement may reflect the broad diagnostic casemix in the adult ICU patient population compared with other subspecialist groups.

Strengths and limitations

The main strength of our study was that we had a relatively large number of raters from different countries, institutions and training backgrounds, and they assessed causes of death using 100 real-life case histories.

We aimed to validate rating systems for use in a research setting and have not made any attempt to validate the classification system for use by non-research clinicians. We only categorised deaths that occurred in the ICU, for pragmatic reasons, and to allow scope for development of this system in other groups. There was no gold standard for diagnosis in each case, so the diagnostic accuracy of the system could not be tested in this way (against, for example, findings at a post mortem). No allowance was made for discussion of cases, and we recognise that such discussion would occur in a clinical trial setting if a rater was uncertain which category to choose; this may have affected the reliability of ratings observed. Independent ratings are also a pre-requisite for calculation of kappa. Future work could investigate the impact of allowing discussion.

Implications

Our study supports the validity of the ICU-DECLARE system (adapted from previously used systems) to classify deaths in ICU patients, provided classification is performed by ICU specialists. The relatively lower degree of agreement observed among research coordinators may be due, in part, to lack of experience in classifying or describing cause of death. Development and validation of systems of this kind would facilitate standardisation of data and subsequent meta-analysis. Moreover, large-scale studies of the demographics of ICU populations[15] would be enhanced if all centres kept records of cause of death in a standardised way.

The similarity of findings from the five-category and eight-category proximate cause of death classification systems suggests that either classification system may be used. Users can take into account their needs for data resolution (five versus eight categories) and clinical usefulness of the categories in each list as applied to their particular project.

Our finding that the 44-category system for underlying cause of death only provided moderate agreement between raters suggests that further work is needed to optimise this component of the classification system.

Future directions

Independent rating of underlying causes of death by two or more ICU specialists, with resolution of disagreements by discussion, may improve the inter-rater reliability of the categorisation system, but this requires further study. Additional studies are also needed to establish the validity of these systems for deaths that occur after ICU discharge. The effect of training raters on the use of the system needs to be assessed.

Conclusions

Using the eight-category and five-category lists of proximate causes of death, the ICU-DECLARE system allows ICU specialists, but not research coordinators, to classify the proximate cause of death of patients who die in the ICU, with substantial reliability.

Competing interests

None declared.

Author details

Elliott Ridgeon[1]
Rinaldo Bellomo[2-4]
John Myburgh[5-7]
Manoj Saxena[5]
Mark Weatherall[1]
Rahi Jahan[8]
Dilshan Arawwawala[9]
Stephanie Bell[10]
Warwick Butt[11]
Julie Camsooksai[12]
Coralie Carle[13]
Andrew Cheng[6]
Emanuel Cirstea[14]
Jeremy Cohen[15,16]
Julius Cranshaw[17]
Anthony Delaney[3,18,19]
Glenn Eastwood[4]
Suzanne Eliott[20]
Uwe Franke[14]
Dashiell Gantner[3,21,22]
Cameron Green[23]
Richard Howard-Griffin[10]
Deborah Inskip[6]
Edward Litton[24]
Christopher MacIsaac[2,25]
Amanda McCairn[26]
Tushar Mahambrey[26]
Parvez Moondi[27]
Lynette Newby[28]
Stephanie O'Connor[29]
Claire Pegg[30]
Alan Pope[13]
Henrik Reschreiter[12]
Brent Richards[31,32]
Megan Robertson[33]
Helen Rodgers[34]
Yahya Shehabi[20]

Ian Smith[35]
Julie Smith[36]
Neil Smith[35]
Anna Tilsley[37]
Christina Whitehead[38]
Emma Willett[17]
Katherine Wong[27]
Claudia Woodford[39]
Stephen Wright[40]
Paul Young[1,41]

1 Medical Research Institute of New Zealand, Wellington, New Zealand.
2 University of Melbourne, Melbourne, VIC, Australia.
3 Australia and New Zealand Intensive Care Research Centre, Monash University, Melbourne, VIC, Australia.
4 Department of Intensive Care, Austin Hospital, Melbourne, VIC, Australia.
5 George Institute for Global Health, Sydney, NSW, Australia.
6 St George Hospital, Sydney, NSW, Australia.
7 University of New South Wales, Sydney, NSW, Australia.
8 Intensive Care National Audit and Research Centre, London, United Kingdom.
9 Mid Essex Hospitals NHS Trust, Chelmsford, United Kingdom.
10 Critical Care Unit, Ipswich Hospital NHS Trust, Ipswich, United Kingdom.
11 Intensive Care Unit, Royal Children's Hospital, Melbourne, VIC, Australia.
12 Poole Hospital NHS Foundation Trust, Poole, United Kingdom.
13 Peterborough and Stamford Hospitals NHS Foundation Trust, Peterborough, United Kingdom.
14 South Tees NHS Trust, Middlesbrough, United Kingdom.
15 Royal Brisbane and Women's Hospital, Brisbane, QLD, Australia.
16 University of Queensland, Brisbane, QLD, Australia.
17 Royal Bournemouth Hospital, Bournemouth, United Kingdom.
18 Malcolm Fisher Department of Intensive Care Medicine, Royal North Shore Hospital, Sydney, NSW, Australia.
19 Northern Clinical School, Sydney Medical School, University of Sydney, Sydney, NSW, Australia.
20 Intensive Care Unit, Monash Medical Centre, Melbourne, VIC, Australia.
21 Department of Intensive Care, Western Health, Melbourne, VIC, Australia.
22 Department of Anaesthesia and Perioperative Medicine, Monash University, Melbourne, VIC, Australia.
23 Department of Intensive Care Medicine, Frankston Hospital, Melbourne, VIC, Australia.
24 Fiona Stanley Hospital, Perth, WA, Australia.
25 Intensive Care Unit, The Royal Melbourne Hospital, Melbourne, VIC, Australia.
26 Intensive Care Unit, St Helens and Knowsley Teaching Hospitals NHS Trust, Liverpool, United Kingdom.
27 Queen Elizabeth Hospital, Kings Lynn, United Kingdom.
28 Department of Critical Care Medicine, Auckland City Hospital, Auckland, New Zealand.
29 Intensive Care Unit, Royal Adelaide Hospital, Adelaide, SA,
30 Medway NHS Foundation Trust, Gillingham, United Kingdom.
31 Gold Coast Hospital and Health Service, Gold Coast, QLD, Australia.
32 Griffith University, Gold Coast, QLD, Australia.
33 St Vincent's Hospital, Melbourne, VIC, Australia.
34 Canberra Hospital, Canberra, ACT, Australia.
35 Hull and East Yorkshire Hospitals NHS Trust, Hull, United Kingdom.
36 Intensive Care Unit, Bendigo Hospital, Bendigo, VIC, Australia.
37 Department of Intensive Care Medicine, Middlemore Hospital, Auckland, New Zealand.
38 Westmead Hospital, Sydney, NSW, Australia.
39 Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom.
40 Freeman Hospital, Newcastle upon Tyne, United Kingdom.
41 Intensive Care Unit, Wellington Hospital, Capital and Coast District Health Board, Wellington, New Zealand.

Correspondence: paul.young@ccdhb.org.nz

References

1 Start RD, Bury JP, Strachan AG, et al. Evaluating the reliability of causes of death in published clinical research. *BMJ* 1997; 314: 271.
2 Earley A, Lau J, Uhlig K. Haphazard reporting of deaths in clinical trials: a review of cases of ClinicalTrials.gov records and matched publications — a cross-sectional study. *BMJ Open* 2013; 3: pii: e001963.
3 Ioannidis JP, Horbar JD, Ovelman CM, et al. Completeness of main outcomes across randomized trials in entire discipline: survey of chronic lung disease outcomes in preterm infants. *BMJ* 2015; 350: h72.
4 Koroshetz W. A core set of trial outcomes for every medical discipline? *BMJ* 2015; 350: h85.
5 Finfer S, Chittock DR, Su SY, et al. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med* 2009; 360: 1283-97.
6 Myburgh JA, Finfer S, Bellomo R, et al. Hydroxyethyl starch or saline for fluid resuscitation in intensive care. *N Engl J Med* 2012; 367: 1901-11.
7 Peake SL, Delaney A, Bailey M, et al. Goal-directed resuscitation for patients with early septic shock. *N Engl J Med* 2014; 371: 1496-506.
8 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.
9 Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005; 85: 257-68.
10 de Galan-Roosen AE, Kuijpers JC, van der Straaten PJ, Merkus JM. Fundamental classification of perinatal death. Validation of a new classification system of perinatal death. *Eur J Obstet Gynecol Reprod Biol* 2002; 103: 30-6.
11 de Keizer NF, Bonsel GJ, Goldfad C, Rowan KM. The added value that increasing levels of diagnostic information provide in prognostic models to estimate hospital mortality for adult intensive care patients. *Intensive Care Med* 2000; 26: 577-84.
12 Halkes PH, van Gijn J, Kappelle LJ, et al. Classification of cause of death after stroke in clinical research. *Stroke* 2006; 37: 1521-4.
13 McGarvey LP, John M, Anderson JA, et al. Ascertainment of cause-specific mortality in COPD: operations of the TORCH Clinical Endpoint Committee. *Thorax* 2007; 62: 411-5.
14 Waljee JF, Windisch S, Finks JF, et al. Classifying cause of death after cancer surgery. *Surg Innov* 2006; 13: 274-9.
15 Vincent JL, Marshall JC, Namendys-Silva SA, et al. Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit. *Lancet Respir Med* 2014; 2: 380-6. ❑