

Article details: 2020-0146	
Title	<b>Quality of end-of-life communication in two high-risk ICU cohorts: a retrospective cohort study</b>
Authors	Tammy L. Pham MPAS MSc, Allan Garland MD MA
Reviewer 1	Dr. John You
Institution	Trillium Health Partners
Reviewer comments and author response	<p>Major comments:</p> <p>1. The cohorts studied are novel (patients from nursing homes, patients receiving ECMO) but seem to create a generalizability problem. If they were selected because they are at high risk for mortality, would it not have been simpler and also a more generalizable sample to study any ICU patient age 80 or older from a more recent time frame since, as they authors state in the Introduction, 80+ year old ICU patients have poor post-ICU outcomes? The authors should consider discussion in the Interpretation section of the potential limitations of this cohort selection choice for this study.</p> <p>---&gt;REPLY: We agree that these two cohorts are not representative of those in ICUs, or even those in ICUs at high risk of death. We have substantially modified the Introduction to explain why these two cohorts were chosen, and in the Interpretation we have added some language as well. Specifically:</p> <ul style="list-style-type: none"> <li>• In Introduction: “Thus, these specific cohorts were chosen because they both have high mortality rates mandating a high level of attention to EOL communication, and because they represent relatively opposite extremes of other characteristics including age and burden of comorbidity.”</li> <li>• In 1st paragraph of Interpretation: “This is despite the fact that the two cohorts are quite disparate in characteristics such as age, and the burden of comorbid illness.”</li> <li>• In Limitations section of Interpretation: “Although it cannot be assumed that our findings generalize to other ICU cohorts, the disparate nature of these specific cohorts is not consistent with our findings being particular to a given type of patient or clinical situation.”</li> </ul> <p>2. It is surprising that there were only 5 nursing home patients admitted to 4 ICUs over a 5 year period (2000-2004), as reported in Table 1. Also, there were no ECMO patients during this time period. Even data from as far back as 2005 are quite old and of questionable relevance to current practice. I wonder if the time frame should be revised to either 2010 – 2017 (presumably the observed improvements in quality of communication still took place during these more recent years) or 2005 – 2017.</p> <p>---&gt;REPLY: Regarding the ECMO numbers, we did not capture ECMO use in our clinical ICU database until 2008. Regarding the NH numbers, from 2006 onwards we obtained this data in a detailed data dump from the provincial department of health, while before that we obtained it from a less detailed data dump from the Winnipeg Regional Health Authority. Obviously the latter reporting mechanism missed many such transfers. To ensure that the NH numbers were not ALSO influenced by a change in referral/transfer patterns over time, we have looked at those numbers for all 6 Winnipeg hospitals, not just the 2 tertiary hospitals included in our manuscript --- and indeed those lower numbers before 2006 are seen in all 6. We have added information about this in the Methods, and added a sensitivity analysis excluding those prior to 2006 -- which gave a similar answer to the primary analysis (Table E4).</p> <p>3. Also surprising is the finding that (unadjusted) quality of communication scores</p>

are almost identical between the nursing home cohort and the ECMO cohort (48.5 vs 49.1, respectively,  $P=0.86$ ), given that the authors found age was a significant predictor of quality of communication (5.0 points per decade) and that the mean age of the nursing home cohort was 2 decades older (72 versus 52 years old). Shouldn't the unadjusted quality of communication scores in the nursing home cohort be at least a little bit higher than the ECMO cohort? Or do other differences between the 2 cohorts (differences in Glasgow Coma Scale and year of admission) essentially "cancel out" the age difference? The authors might consider providing further explanation of this since I suspect other readers may wonder the same thing.

--->REPLY: Indeed it is simply a coincidence, accounted for by the baseline differences between the groups (as can be verified from the model coefficients and some algebra). Our primary analysis is for the median, not the mean, and we have augmented Table 2 to include unadjusted medians, which are not as close to each other as are the means. In light of this, we have chosen not to comment about this coincidence in the Interpretation.

4. I think the regression analysis performed (Table 3) is multivariable, i.e. many covariates were included in the model, but the label "unadjusted p-value" in the right-most column of Table 3 is confusing since it suggested this is a univariate analysis. I think it might mean that these p-values are not adjusted for multiple comparisons/false discovery rate but suggest somehow clarifying this for readers.

--->REPLY: You are correct. To avoid confusion we have altered this column heading in Table 2 and Table E1 to read "p-value, not adjusted for multiple comparisons".

5. The ACCEPT audit (reference #19 by Heyland DK et al, for which I was also a co-investigator) did not enroll ICU patients but actually enrolled patients (and their family members) from general medical wards with serious illness. As a result, it is not surprising that the quality of communication or frequency of individual quality indicators (Table E2) would be different (i.e. lower) when compared to the high-risk ICU cohorts included in this retrospective cohort study, given a lower baseline risk of death amongst the medical ward patients and – potentially – less available time/higher patient volumes on medical wards that might prevent clinicians from more frequently engaging patients/families in these conversations. The authors could either omit this analysis or revise the corresponding sections of the manuscript to account for this, especially the interpretation. [Editor's note: If you intend to compare your study results with those of the ACCEPT audit, you will need to include that information as a study objective in the Introduction and add in the appropriate information in the Methods section and Interpretation section, including Limitations subsection.]

--->REPLY: First, thank you for pointing out that we had misrepresented the ACCEPT cohort as being ICU-specific. We have now fixed that in the final sentence of the 1st paragraph of the Interpretation section.

Second, it is precisely because the ACCEPT population was less acutely ill that we raised it --- i.e. the lower end-of-life communication score being even lower in that cohort is consistent with our finding of lower scores in ICU patients with a better (short-term) prognosis. Your hypothesis that ward attendings have less time available per patient to discuss EOL issues is interesting, but we are not aware of any evidence that directly speaks to as accounting for the low communication scores in ACCEPT (e.g. that is not mentioned in the ACCEPT paper). We have chosen to deal with your concern by altering the relevant sentence in the 1st

paragraph of the Interpretation to read “Although other possible explanations exist, the same phenomenon could explain the even lower average composite score of  $29 \pm 5$  reported by Heyland et al.(19) in their study of selected, less acutely ill ward admissions in 12 Canadian hospitals.”

Regarding the editor’s note about where to include this information, please see our response to Editor comment #2.

6. The Interpretation section could be further improved by including more concrete discussion about what are the clinical implications of the findings, i.e., what should ICU clinicians do differently based on these findings?, Or if they are not sufficient enough to be practice changing given limitations (essentially single centre study, older data) then what are the implications of these findings for further research in this field.

--->REPLY: We have augmented the 1st paragraph of the Conclusion section to expand upon and further highlight the implications of our findings.

7. On the one hand, I do agree that it is tempting to conclude from these findings that lower risk patients are being denied the opportunity for higher quality communication and that more needs to be done to rectify this imbalance. However, firstly, what is the magnitude of the difference that was observed between, for example 70 year olds and 50 year olds? Was it a clinically meaningful difference in quality? Secondly, it not appropriate that limited resources (clinician time) be triaged and allocated to the highest risk patients for whom these discussions are the most urgent? Surely, the desirable state isn’t for quality of communication to be higher for the least sick patients. So, a more in-depth discussion of the study implications might be warranted.

--->REPLY:

- Addressing your first point, the model-based difference in median score (out of 100) for a 20 year difference in age is 10 points. As you indicate the MIC (minimally important change/difference) of the composite score would be extremely valuable to know for interpreting our findings. Terwee et al. have done extensive empiric work on the MIC (e.g. CB Terwee, LD Roorda, J Dekker, SM Bierma-Zeinstra, G Peat, KP Jordan, P Croft and HC de Vet. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 63(5):524-34, 2010). They identified 5 different methods, including the commonly used  $MIC=0.5SD$  method. They found that the various methods provide widely divergent values that are highly dependent on the population characteristics. They further make the point that all the distribution-based methods identify the minimal detectable change/diff, not the minimal important change/diff, and that no theoretical arguments have been advanced for why 0.5SD (or any other data distribution-based value) should be important for subjects. Their conclusion is that in the absence of a gold standard provided by an available external “anchor”, that they cannot recommend any of the statistical methods. While no formal work has been done to identify such an anchor for this instrument, we note that 10 points on this 0-100 scale is 10% of the full scale range, and that in your 2017 CMAJ paper that validated this instrument, the highest ranked and lowest ranked hospital differed by 22 -- which also suggests that a 10 point difference is relevant. We have now included, as the 6th sentence of the 1st paragraph of the Interpretation, that “The magnitude of variation of median score by these factors is approximately 10% of the full scale 0-100 range, corresponding to a 5 point difference in GCS or a 20 year difference in age.”

- Regarding your second point: Please note that, as stated in the manuscript, the two cohorts included in our analysis experienced 30% and 47% hospital mortality,

	<p>and that abundant evidence has shown that individual-level prognostication is, at best, mediocre. Together, these facts appear to mandate that all of these patients need attention to Advance Care Planning and the related EOL-related communications. Beyond that, however, as stated in the Conclusions, we believe that all patients (and/or their surrogates) who are at more than trivial risk of dying (which includes all patients in an ICU) be involved in these discussions. Supporting this is the ACCCM/ATS Policy Statement on shared decision-making in the ICU (ref #45) saying that “Clinicians and patients/surrogates should use a shared decision-making process to define overall goals of care”, which clearly mandates discussion with all ICU patients and/or surrogates. Furthermore, other expert opinions on these matters concurrently indicate that discussions should occur at or soon after ICU admission in all cases (e.g. ref#2). We have added to the Conclusion section the following sentence relevant to this: “Second, as existing guidelines and expert opinion<sup>2,45</sup> highlight the necessity for a shared decision-making model, direct and early communication should occur with all ICU patients and/or their surrogates, not just those who are believed most likely to die.”</p>
<b>Reviewer 2</b>	Dr. Simon Oczkowski
Institution	McMaster University, Hamilton, Ont.
Reviewer comments and author response	<p>1) How were the cutoffs for the scaled quality scores created (e.g., &lt;50% 'extremely low quality')? The fact that both the original validation cohort and this study identify all patient populations as extremely low quality suggests that either our expectations are too high (it is not reasonable to expect all patient care teams to meet all indicators for all patients) or we are poor at measuring these (as I discuss below) or clinicians are bad at this (I think this is true) or a combination (almost certainly). Providing transparency around how these cutoffs were decided will at least let the reader decide whether or not to accept the conclusions of whether or not the quality is high/low/very low etc.</p> <p>---&gt;REPLY: We began with the idea that all 18 elements were important -- derived from the original publication of these elements (ref#3) in which the expert group rated all 18 as being “moderately important or greater”. If indeed, none of the included items were irrelevant, then it would seem reasonable to believe that a score of 100 (out of 100) is desirable and attainable. Thereafter, the ranges chosen (extremely low quality &lt;50%; low quality 50-74%; medium quality 75-84%; high quality 85-100%) were based on this, but certainly have an element of arbitrariness to them.</p> <p>The fact that the subsequent study using this instrument in general wards (ref#21) observed that hospital-level scores ranged from 18 to 40 across the 12 Canadian hospitals highlighted the idea that clinicians are bad at this. While your other possible contributors (expectations are too high; these items/measurements are not a good representation of EOL communication) are also possible, the fact is that this scale is the only one of which we are aware that has done ANY validation at all.</p> <p>In response to your concern, we have added to these thoughts to the Limitations section: “Regarding the cutpoints, we began with the idea that all 18 scale elements are important, as judged by the expert group in the original description<sup>3</sup>. Given that, we believe it reasonable that a score perfect score of 100 is desirable and attainable, producing a reasonable decision that a score of &lt;50 is “extremely low.”</p> <p>2) One of the major differences between the ECMO and NH cohorts is that NH cohorts may be more likely to have participated in ACP. Many NH require a documented code/resuscitation status on admission, and given their advanced age and high baseline mortality rate outside of acute illness. By contrast, the ECMO</p>

population is younger and likely has a lower baseline mortality rate outside of acute illness (selection of who gets ECMO in the first place). ACP probably helps to prepare patients for in-hospital discussion and decision-making and may be part of the reason for differences in the documentation score between these groups in table 2. It is very uncommon for young, healthy patients to have advance directives on admission to ICU.

--->REPLY: We agree with the likelihood of your observation. Unless you feel very strongly about it, we prefer not to make this additional point in the Interpretation, as the sub-scale scores were subsidiary to our main messages.

3) It is unclear how the chart reviews were used to identify which quality indicators were met. The quality indicators only include 5 documentation items meeting that 13 of the goals of care communication items are not even intended to be documented. Thus there is a very high probability that many of these items/discussions may have occurred but not have been documented. Indeed, examining Table E2, the documentation scores appear to be much higher than the other GOCC items. This, to me is the major weakness of the study-- the GOCC could be perfect, but we will not know if they are not documented in the chart! This is a fundamental challenge of conducting research in this area. Both documentation and patient/family recall are often poor. Somewhat reassuring that this study's results reflect reality at the included study centres is that in the Heyland paper they used interviews with patients and families and the rates of GOCC were also low. I would expand upon this in the discussion.

--->REPLY: We agree, and briefly pointed that out in the Limitation of "we could not include end-of-life conversations that occurred, but were not documented." As suggested, we have added content (using much of your own wording, thanks!) to: "Second is that by using manual review of hospital charts, we could not include end-of-life conversations that occurred, but were not documented. This is a fundamental challenge of doing research in this area. However, reassuringly, a prior study that confirmed the presence of the same communication elements by interviewing hospitalized patients and their families, found even lower scores.<sup>21</sup>"

4) Lastly, I question the weighting of scores according to the Heyland paper. The weightings were developed by a panel of experts, without input from patients and families (at least, as far as I can tell from the manuscript!) Thus I think there is reasonable room for disagreement over whether or not these weightings are truly of value to patients and families or simply represent the views of a small group of experts. As well, the ratings themselves were subject to a significant ceiling effect (ie almost all items were rated very highly), again making weighting of questionable value. How substantial an effect did the weighting have on the final scores? The impact of weighing upon the scores is probably worth commenting on in the discussion, as other centres who try to use the quality indicators will probably want to know if weighting is useful above and beyond raw scoring.

--->REPLY: Thank you for this. A note --- it was the weightings applied to these 18 yes/no questions that were all high in the originating study by Sinuff et al., not the actual scores.

To address your concern about the weightings, as a sensitivity analysis we re-ran median regression analysis with all 18 items equally weighted, and scaled 0-100. The findings were similar, and this is now described in the Methods and the Results and shown in Table E3.

Regarding ceiling and floor effects, the standard identification of them (J Clin Epidemiol 60(1):34-42, 2007) is that >15% of the subjects have scores that are at (respectively) the maximum or minimum values. None of our subjects had those

	extreme scores.
<b>Reviewer 3</b>	Dr. Kelly Chessie
Institution	
Reviewer comments and author response	<p>good / important research topic / question. Nicely written article (clear, concise)</p> <p>1. page 2: any concern that the extreme low scores introduce a bias or limit ability to detect? or speak to a limit in the database? ---&gt;REPLY: As the scores are, in fact, in the middle of the full 0-100 range of the composite measure. Thus, there is no particular reason to be concerned about either floor or ceiling effects. Furthermore, the standard identification of ceiling and floor effects (J Clin Epidemiol 60(1):34-42, 2007) is that &gt;15% of the subjects have scores that are at (respectively) the maximum or minimum values. None of our subjects had those extreme scores.</p> <p>2. page 7: any thoughts as to why the difference between your cohort and Heyland's? ---&gt;REPLY: As now described in the 1st paragraph of the Interpretation, this is consistent with our finding of lower EOL communication in patients at lower risk of dying.</p> <p>3. page 4 (and 8): any concern with the ICUs not included? Any reason to think this introduced a bias to the data? e.g., they are in higher SES neighborhood or lower? ---&gt;REPLY: As stated in the "Setting" section of the Methods, we included all the ICUs in the only two tertiary care hospitals in Manitoba. Only these hospitals perform ECMO and it was important for eliminating unexplained variability EOL communication that the hospital/ICU infrastructure for the two clinical subgroups studied be the same. However, with such a small sample of hospitals in a single city, there must be important concerns about generalizability, which are discussed in the Limitations.</p> <p>4. page 5: description of scoring. is this weighting the standard scoring for this test? ---&gt;REPLY: Yes, as described in refs #3 and #21.</p>