Reviewer's report

Title: Do cognitive aids reduce error rates in resuscitation team performance? Trial of Emergency Medicine Protocols In Simulation Training (TEMPIST) in Australia.

Version: 0 Date: 22 Jul 2019

Reviewer: Mark Fan

Reviewer's report:

First, I would like to congratulate the authors for taking on this work and producing helpful evidence to support the use of cognitive aids in emergency procedures. I believe the article will be of interest to clinicians, policy-makers, hospital management, and human factors specialists alike.

While I think the article should be published, I would like further clarification in several areas. I think the author's responses (and if appropriate, additions to the paper), may help other readers understand the various issues at play more easily. The paper will also benefit from a copy-edit as there are some grammatical issues that compromise readability in a few places.

I do not have expertise in randomized controlled trials or statistical analyses, so will defer to the authors and the editorial team if any of my requests for clarification are the result of a misunderstanding on my part. My apologies if that is the case.

My comments are organized by numbered item below:

1. Consistent Terminology for "Primary Outcome Measures"

In the abstract, the following is stated: "The primary outcome measures were a list of 15 tasks per scenario. Each team was measured on 60 key processes." I thought it was odd that a "list" was the outcome measure.

I would recommend rewording these two statements. As an example:

"The primary outcome measure was the percentage completion of 15 key tasks per scenario; each team completed 4 scenarios." The authors cite Arriaga et al, who use "failure of adherence to critical steps"; this could be another possible phrase. It may also be worth highlighting that while all scenarios had 15 critical steps, the steps themselves varied by scenario.

I would also recommend that the authors standardize on a specific language for these 15 tasks. They are referred to as 'tasks' or 'key processes' in the abstract, but also as '15 primary outcome measures' or even 15 'key outcome measures' in the methods. In the statistical analysis the framing shifts to error rates. The authors may also consider whether they wish to distinguish
between errors (e.g., inaccurately performing the 'key task') versus omissions (i.e., missing the step completely), as they may have different significance for the reader. For example, in the discussion, paragraph 1 sentence 4, it is stated there was a 54% reduction in 'errors' made across all four scenarios, but I personally would think about the results differently if it said a 54% reduction in 'omissions' of key tasks. I think readers will catch on regardless, but consistency in the language throughout the paper will minimize their cognitive workload.

2. Add Discussion Regarding Ad Hoc Groups/Teams; May Not Account for Rural vs Urban Settings

In the introduction, the authors highlighted that smaller urban and rural departments have lower exposure to patients in extremis, and lower staffing levels. This seems to be a critical point, as busy urban centres might have different responses to the implementation of the cognitive aid than rural centres. The discussion then states that "the next step is to adopt this type of intervention in practice". While I agree that some pilot testing of the handbook in real settings is a valuable next step, the paper would be strengthened by discussing how differences in urban vs rural setting have not yet been thoroughly investigated; it was not controlled for in this trial.

For example, there is no indication that individuals from rural settings were purposively grouped together into teams. As a result, teams likely mixed clinicians more familiar with in extremis patients (due to being stationed in a tertiary setting), and less experienced clinicians. This grouping could theoretically boost the team's performance, and wash out urban vs rural differences.

If anything, the tool could possibly be more effective in rural settings where in extremis patients are less common. The strong results in the tricylic overdose scenario, hypothesized to be due to its rarity, support this idea.

While I believe the strength of the results suggest the handbook will likely be useful in both settings, I feel that this limitation should be acknowledged or addressed in some fashion in the discussion, at a minimum to encourage future work to account for this. This is somewhat related to my next point.

3. Acknowledgement that other factors affect checklist implementation

In the introduction, it is stated "Despite numerous recommendations for the use of checklists in critical care medicine, adoption of such practice has overall been slow." This point is important for readers considering implementation or adoption of the handbook, but it is not explored in the introduction or discussion.

The introduction continues by mentioning the Checklist Manifesto by Dr. Gawande. However, implementation of the surgical checklist is non-trivial. Urbach et al found that there was no significant impact in mortality or complications after it was implemented (https://www.nejm.org/doi/full/10.1056/nejmsa1308261). Dr. Leape in response to Urbach et al's
paper states to this paper states that "...changing practice is not a technical problem that can be solved by ticking off boxes on a checklist but a social problem of human behaviour and interaction." (https://www.ohin.ca/Leape.pdf).

Given that the authors suggest that the next step is implementation in practice, the details of how the implementation should occur deserves mention in the discussion. This is related to my next point.

4. Discussion or justification needed for not training the teams prior to use of the handbook

While the handbook was designed such that an individual would read the steps aloud, it seems the groups were not told this was how it was supposed to be used, nor who would be responsible for reading aloud. Dedicating an individual to this role would conceivably change who is accountable for tasks, change patterns of task delegation and who feels responsible for what. For example, in the discussion it is suggested that a nursing role could possibly be the "resuscitation guide manager". In my experience, nurses are typically more familiar with the location of supplies, and executing tasks (e.g., IV access, medication administration), so if the number of staff available is low, tying up the nurse with the handbook may actually place a greater task load on a physician, who may be less experienced and efficient with some of the required tasks. The physician might also be less accustomed to having a competing role dictating the flow of events, as the resuscitation guide manager is in a way co-leading the scenario. As a result, the effect of training is an important area for future work, and it would be useful if the authors emphasized this more, perhaps prior to implementation in live clinical settings.

It was also not reported if any teams spontaneously used the handbook in the way it was intended (designated individual to speak the steps aloud). I recognize that within the word limit this may be a challenge, but these are important details related to how the handbook would be implemented in practice, and so any discussion that can be added in this regard would help make the findings more practical for institutions wishing to adopt the handbook. If teams naturally designated someone to reading steps aloud, that might minimize the depth or format of training clinicians on how to use the handbook and speak to its usability.

Also to be discussed prior to implementation of the tool is where the handbook would be placed, how clinicians would be cued to use the handbook, or if they would remember to take the time to scan the table of contents to determine if the patient condition is present in the handbook. Presumably the nature of the simulation itself minimized the need for participants to answer these questions.

If any of these issues can be commented on in the discussion, I think it would strengthen the paper. Appropriate training on the use of the handbook turned out to be a major discussion point with regards to the participants' weaker performance in the VT scenario, so the lack of training needs to be touched on in greater detail.

5. Consider Deleting Second 'Objective' Statement in Methods Section
The objective of the study is stated at the end of the introduction and then midway through the 'intervention' suggestion in the Methods. Is this by design? The second occurrence seems out of place, and could probably be deleted.

6. Description of Scenario Randomization May Need to Be Amended

I see that a previous reviewer commented on line 6 of the methods where it states that "…accounts for all possible four scenario and two handbook combinations". The reviewer found it confusing and I see that the authors responded and said that they rephrased this, but unfortunately I still find it a bit confusing in the body of the methods.

Please correct me if I am misunderstanding, but in order to account for ALL possible combinations, a minimum of 144 different teams would have had to participate.

This is because the 1) scenario sequence can be counterbalanced in up to 24 possible combinations (1234, 1243, 1324, 1342.…etc.) and 2) the handbook (H) and non-handbook (N) order can be counterbalanced in up to 6 possible combinations (HHNN, HNHN, HNNH, NNHH, NNNH, NHNN)

As a result, \(6 \times 24 = 144\).

Since the authors wrote "Four different randomization sequences in blocks of six were created…" (Methods - Random Allocation, 2nd sentence), I am assuming the authors selected just 4 of the possible scenario sequences, and therefore could not rule out the possibility that the sequence of the scenarios may have impacted the results (e.g., learning effects, fatigue). Perhaps the methods should be amended so that the language is more clear. For example, the sentence above could be amended to say: "Using SPSS v22, one investigator (MR) generated 4 different scenario sequences. Each scenario sequence was associated with a block of 6 combinations of handbook (H) and non-handbook (N) conditions across the four scenarios (i.e., HHNN, HNHN, HNNH, NNHH, NNNH, NHNN)." This is just one possibility; I leave it to the authors to develop their preferred wording, and ensure that the 'scenario sequence' language is clarified throughout the manuscript.

Given the infeasibility of trying to counterbalance all possible scenario sequences, what the authors have done is entirely appropriate (and I doubt further testing with alternate scenario sequences would have much impact on the results), but I think it is a bit hard for the reader to follow as currently written. Since at least one other reviewer has commented on this, I think a worked example of the scenario and handbook sequencing in the appendix would help interested readers wrap their head around how the authors approached this and avoid future questions. If feasible, perhaps a 1 sentence statement in the limitations section to mention that the effect of the scenario sequence was not fully counterbalanced would be a nice way to resolve this issue, but I don't consider this mandatory given that partial counterbalancing is normal in these types of situations.
I would also at minimum amend the sentence that says "... which accounts for all possible four scenario and two handbook combinations", because not ALL possible combinations were addressed, unless I have misunderstood how this was done.

7. Consider Removing Duplicate Explanation of Randomization

Randomization is discussed in the methods twice, once briefly in the "trial design", and again in more detail in the "random allocation" section. Suggest keeping this just to the latter section, and if it needs to be addressed elsewhere, perhaps that section can point to the "random allocation" section. I defer to the authors and editors if RCT reporting must follow a specific format; I am only pointing out that the duplication might be confusing to some readers.

8. Possible Figure 1 typos

The excluded total is (N=0), but the sub bullet says 1 team declined to participate. Should the total for total exclusions read (N=1) instead?

Also, in the third box it should read "Randomized (n=21) across all four scenarios".

9. Request for more Detail on how the Groups were Assembled

It is stated that the groups ranged in size from 3 to 6 individuals. What dictated the size of the group? This is a rather large spread - the largest team was 100% larger than the smallest team. Was this purposeful? Do you think it had an impact on how the handbook was used? Perhaps the distribution of team sizes could be noted in the appendix?

10. Clarify distribution of doctors and nurses in each team

In the "Random Allocation" section of the methods, it is stated that "Participants available for each session were grouped by the facilitator to ensure an even distribution of doctors and nurses in each group prior to randomization allocation". However, the smallest team was 3 individuals, so it is not possible for there to be an even split of doctors and nurses in this group. Were there also teams of 5 where the split was not even? Authors may want to clarify the language in this section to address this incongruity.

11. Request to use consistent language regarding 'facilitator' and clarify what they were responsible for during the simulation.

In the "Random Allocation" section of the methods, there is a 'facilitator', but also a sentence that states "The envelope was opened by a member of the study group just prior to the commencement of the scenario". Is this member also the facilitator? In the 'Blinding' section of
the methods there is a reference to 'investigator in the room'. Later in the 'Intervention' suggestion it is described how the teams were oriented to the simulation - was this done by the facilitator? In the discussion, the individual remaining in the simulation is referred to as the 'observer'. I think being more consistent with the terminology here will make all this easier to follow.

12. Can more detail be given about the high fidelity simulation laboratory?

I see the simulation lab includes mannequins. Was it a small consultation room, or a more spacious laboratory? Were all typical clinical supplies available in emergency departments available, or only the ones needed for the 4 scenarios? This is only to provide the reader the context how representative the environment was in terms of space around the bed, room size, lighting etc. How many cameras were used and where were they placed? Was a separate microphone used to supplement the camera? A screenshot of the video camera's view might be helpful here (possibly in appendix).

13. What happened if teams were not finished the scenario in 15 minutes? I assume any of the 15 key tasks not yet completed were instantly considered 'missed'. Did this happen to any teams? Maybe a small addition in the methods could answer this question.

14. "No staff member had received any formal training regarding the correct use of this cognitive aid prior to the trial". How did you determine this? This does not appear to be part of the demographic survey.

15. In the "outcomes" section it is stated that "Three investigators reviewed the video recordings and scored each outcome measure", but the appendix states that there were 4 raters (and no initials are present for the fourth rater). Perhaps the text in the main article could be worded differently? In addition, it appears based on the explanation in the appendix that one rater (DR) viewed ALL the recordings, and whereas raters 2 and 3 viewed a subset. Currently the first part of the sentence suggests that 3 raters reviewed all the recordings.

16. Consider deleting the last sentence of the methods - outcomes section. What happened during the debriefing? Was it unstructured? Semi-structured? If you don't have room to get into this, maybe it makes sense to remove this sentence from the paper, as I was expecting some comment on it in the Results. The word 'debrief' does not appear anywhere else in the paper.

17. Add Male gender to Table 1. Since all subsections add up to 100%, I would recommend adding "Male" to the gender section; it is currently omitted.
18. In the results (paragraph about table 3) there is a reference to ("See Table 3 and 2c"); I believe this is a typo? There is no table 2c.

19. Consider reorganizing discussion so that similar ideas are discussed together.

The discussion is currently quite long and topics are touched upon in one paragraph and returned to later. If allowed by the editors, I would recommend the addition of subheadings. This would allow thematically related material to be better grouped together and give the reader a context for each idea.

For clarity, I perceive the discussion to have 10 paragraphs, roughly summarized as follows:

a) A brief summary of the main outcome from the study

b) Speculation on why VT scenario had the least improvement (brief mention of tricyclic scenario at the end)

c) Statement that almost all participants would want to use this cognitive aid in actual practice

d) Statement that regardless of team composition, the cognitive aid was helpful

e) A large paragraph addressing 4 different points, including: 1) errors of judgment or omission are common contributors to patient safety, 2) high quality cognitive aids depend on the right content and design, and are supported by appropriate training (which was not thoroughly discussed or planned for in the trial), 3) The handbook went through an iterative design process, and 4) Linear cognitive aids seem to be superior in prior research. [As an aside, it seems odd to define what a checklist is at this point of the manuscript - perhaps this should be done in the introduction?]

f) Describing a limitation of the study (e.g., cannot extrapolate to the value of the handbook in real clinical situations).

g) A return to the issue of linear flowcharts.

h) Discussion of additional limitations.

i) A return to discussing the scenario specific findings (tricyclic scenario vs VT scenario)

j) Discussion of next steps and implementation challenges.

I have tried to illustrate above that the discussion has separated some topics that might be better together. One option for the authors to consider could consist of the following (with sample subheadings):
* Overview of findings: Paragraphs A, C and D

* Value of Checklists/Cognitive Aids: Paragraph E, although ideally it would be broken into several paragraphs so that each idea remains distinct. Paragraph G would also be in this section, so that the topic of linear guides is discussed together.

* Scenario-specific Challenges: Paragraphs B and I

* Study Limitations: Paragraphs F and H (several of my earlier comments about how to actually implement the handbook could be addressed here)

* Future Work: Paragraph J

20. Consider elaborating on the importance (or non-importance) of completing key tasks in a specific order.

In the discussion, paragraph 2, it is stated "Participants likely did not follow the step-by-step approach in the handbook. They were unable to recall steps in the correct sequence so they made errors." Up to this point, in my mind, the primary outcome measure was the completion of the 15 key tasks regardless of sequence. If order of completion is important, this may need to be explained earlier in the article, perhaps in the introduction. For example, I see that some key tasks are actually dependent on earlier key tasks (e.g., if IV access is not obtained, then sequential medication administration steps may not be completed).

21. Consider discussing the paradox of expertise in future work.

In paragraph 2 of the discussion, the authors state that "Familiarity and dexterity with this cognitive aid is expected to increase with further practice and training". This is an interesting statement given that the scenario where participants had the most experience (VT) was the scenario that teams performed the worst in, with or without the handbook. This suggests that longitudinal research with the handbook may be required to assess whether extended familiarity and use of the handbook in real clinical settings is actually sustained. It is possible that extended familiarity with the handbook could cause a similar effect to that of participants receiving additional VT training (e.g., a degradation in completion of key tasks). This is a key question that needs to be evaluated, and I think it is worth mentioning, or citing other research that has been done on this topic.

22. Consider discussing user preference for flowcharts as a recommendation for future work.

The authors note that roughly a third of participants did not favour linear steps compared to flowcharts, and also cite Marshall et al's study on branched cognitive aids. The authors conclude that clinicians may need some time to get used to linear flowcharts. If the authors agree, I think this is
a good opportunity to suggest future studies on linear vs branched cognitive aids, particularly as there is no data from the study itself to explain why users have this reservation (unless there is data to share from the). At present, it appears that the participants' feedback on this issue is not being given serious consideration.

For example, the preference for linear steps could be an artefact of the scenarios used, where simulated patients presented with a single condition. Patients with comorbidities could potentially be better treated with a handbook that had branching steps, and this flexibility to address multiple competing demands may be a reason some clinicians had reservations about a simple linear process. I don't want to speculate, but the reservations of a third of participants should be explored further. It may be possible to perform effective branched cognitive aids with digital technologies, which Marshall et al alludes to as well. These electronic cognitive aids might also address some of the implementation challenges touched upon earlier (e.g., how participants recall which patient conditions are covered by the handbook), because charting patient vitals and history into a rapid triage screen could automatically pull-up the relevant prompts for users without their need to consider competing care pathways.

23. Better justify link in final paragraph of discussion with study results

In the final paragraph of the discussion, the authors suggest that the resuscitation guide could be someone with expertise in human factor sciences, and that the appointment of human factors directors on local health government levels are a practical next step. While I applaud the authors for suggesting these proactive steps, and agree in principle, I do not see a clear connection to the study results. It seems this last paragraph takes a detour into broader system level suggestions for non-technical skills training. I think a more reasonable conclusion would be that as future work is completed to answer questions regarding implementation success, pilot studies in live clinical environments, organizational leadership will be hardpressed to ignore the patient safety benefits, regardless whether they have human factors members on staff or not. It is also not clear that an individual with human factors experience is necessarily better positioned to act as a person reading the handbook steps aloud.

24. Reword 'human performance' in final paragraphs of manuscript.

The phrase 'human performance' makes its first appearance in the second last paragraph of the manuscript and seems to signal a change in tone. The authors begin to discuss non-technical skills training, and individual behaviours. What the study measured was completion of key tasks, rather than individual human characteristics. It may be preferable to keep the final paragraphs consistent with earlier language.

Again, I congratulate the authors for their work and am looking forward to seeing the final published version. I hope my comments help pre-emptively address questions that arise in readers' minds and round out the comprehensiveness of the paper, which is an important contribution to the literature.
Level of interest
Please indicate how interesting you found the manuscript:

An article of importance in its field

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal