Reviewer’s report

Title: The Triple Variable Index combines information generated over time from common monitoring variables to identify patients expressing distinct patterns of intraoperative physiology.

Version: 0 Date: 27 Aug 2018

Reviewer: Matthias Görges

Reviewer's report:

Dr. Schnetz and colleagues describe the development of data fusion approach to identify adults undergoing general anesthesia, who are at risk for undesired outcomes, particularly 30-day mortality. They do so by integrating data from hemodynamic, hypnotic-drug, and depth of hypnosis indicators by normalizing those data, identifying patterns, and clustering patients with similar characteristics. The variables selected mirror those of the triple-low concept introduced by Sessler, in which they were associated with increased mortality. The paper is nice to read and likely of importance regarding the use of big-data analytics and prediction in perioperative care.

General comments:

a) While I quite like the paper, I am a bit unclear whether it fits the scope of this journal: "Articles on the methodology of epidemiological research, clinical trials and meta-analysis/systematic review are particularly encouraged, as are empirical studies of the associations between choice of methodology and study outcomes." It seems that your approach is very anesthetic (and outcome)-focused, rather than methodology focused?

b) Some details of the data wrangling need to be explained with more detail, and for some of the design decisions justifications would be useful. Particularly, I am wondering if a simple additive score, versus a weighted score, might be important. Similarly, what the effect of the window length, and therefore loss of variability might be is a question of interest.

c) I am a bit confused about the order in which you explore associations. Is the main goal to identify a risk score, with predictive value, for which there are associations with known risk factors, or is it mainly to characterize the profiles of the groups clustered by your approach? The discussion leans towards the former, while the methods and results to the latter?

d) The manuscript is very long, but I am not sure what to suggest to shorten it, short of suggesting to split it into multiple publications exploring different aspects.
Specific comments:

Title: The second part of it is a bit flipped, as you should be interested in patients with certain physiological characteristics, and those with undesired outcomes, not the other way around?

Abstract, P3, L50: interoperative monitoring variables (not systems, only BIS is a system)

Abstract, P3, L63: Please comment on sampling frequencies here. I guess that you weren't able to find more triplets as the NIBP cuff cycle time limits this?

Abstract, P3, L69: Unless report the ages in the other groups, 54.5 years is not meaningful to report here. Readers might argue that 75+ is old, 55 is not.

Abstract, P4, L75: I feel the reporting is flipped again, the patterns [thus eventually useful to predict] are associated with the undesired outcomes, not the other way around?

Abstract, P4, L77: I am not sure what you are saying here. Are you saying that for patients with normal MAP and low-ish BIS, your indicator was able to discriminate those at high risk for mortality from those that were not?

Background, P5, L93: MAP (multiparameter [hemodynamic] monitor) and MAC ([end-tidal] gas analyzer) are not monitoring systems, they are derived by them. Even BIS is not as you actually mean their DoH index here.

Background, P6, L119: I am wondering about how you weight these factors, which you might explain in the methodology but could hint at here already?

Background, P6, L119: Aims deserve their own paragraph.

Background, P6, L124: Does it help the reader to pre-empt the methods here?

Methods, P6, L147: Type of surgery was by service/speciality, or by CMS procedure code, or by something else like billing codes?

Methods, P6, L148: And their timing, which is relevant more than a binary response to associate with the variables under investigation. Also this makes me wonder if you considered (even crudely) the pharmacokinetics of the drugs in their effect, which obviously continues past their discontinuation of use/bolus.
Methods, P6, 156: You later mention you were mainly interested in 30 day mortality, so why collect it for up to two years past the procedure. The more (temporarily) remote the event from the outcome, the weaker the association should becomes.

Methods, P7, L160: This sounds reasonable, but how did you deal with the fact that an arterial line will give you q/1sec data, while the NIBP might only be cycling q/5min? Also please report sampling frequencies of your AIMS for these variables and if you carried forward the NIBP or only calculated your index when a new measurement was obtained.

Methods, P8, L166: But we know that mac-age-sum exists for a reason, as at both extremes of age anesthetic MAC might be quite different (kids need much more, elderly much less drug). Similarly I am wondering why you didn't use age-adjustments for MAP - I know pediatric nomograms exist, but I don't know if adult ones do? If not future work?

Methods, P8, L169: A MAC of 0.001 is very low, so you probably would get that for over 30min following an inhalational induction when converting to TIVA.

Methods, P8, L170: Can you elaborate more on the data cleaning, including how you dealt with measurement artifacts like ABP transducer repositioning, electrosurgical interference in BIS etc. Also capturing BIS=0, as in SR=100 might actually be true, and particularly undesirable.

Methods, P8, L178: Did you test for that, I would intuitively have assumed that MAP would be bimodally distributed (either high or low not much in-between).

Methods, P8, L183: Please also express this in time, as the timing is otherwise not known. I am not suggesting you should change it, but wouldn't a median not have been more robust to artifacts?

Methods, P9, L195: Can you show an example (and in it also the raw data as well as the z-scores that went into the fused TVI index).

Methods, P9, L197: Did you split these by procedure phase, particularly induction time being so variable that lack of stimulation might have a significant effect here?

Methods, P10, L217: Quickly say why this approach is particularly suitable for your application/setting/etc.

Methods, P11, 238: When you mean compared here, is this statistically using a pseudomedian, 95%CI and p-value, such as with the Wilcoxon test, or did you only calculate and report group characteristics statistics here?

Methods, P12, L272: Your definition is quite different from Sessler et al. - why? Also provide a reference to it, if you are using a published definition, which is different here.
Methods, P13, L279: Did you attempt to adjust for the (temporal) remoteness of the intervention and outcome here?

Methods, P13, L291: Were these results then compared to the TVI method to establish superiority/inferiority in predicting outcome, e.g. using receiver operating characteristics analyses and comparing the AUC with DeLong’s method for paired ROC? If not, what is the reason to calculate these?

Methods, P13, L296: Aren't you now mixing independent and correlated factors?

Methods, P13, L302: Wouldn't you expect better performance (at least sensitivity) if you had used a higher thresholds than the mode?

Results, P14, L311: Does the lack of DoH monitoring introduce bias? Similarly, if TIVA were better for patients than IH anesthesia, shouldn't you also look at a TVI where the risk contribution from MAC is zero?

Results, P14, L318: It might be nice to report the IQR (or SD) here and for procedure length too? Also 53.8% is technically more males, but not by much.

Results, P15, L325: Why was MAC samples less frequently than MAP? (Same question for BIS)

Results, P15, L332: So the minimum procedure length was 27 or 45 min?

Results, P16, L354: Sum, not product, or did you indeed multiply them and I missed that in the methods? (P9, L188 says sum).

Results, P16, L360: I'd sure hope so as otherwise there is no point in DoH monitoring. That said, I actually find the correlation rather weak?

Results, P17, L375: Isn't the association backwards again here? Or am I missing the point that you are not trying to find patterns with risky characteristics but rather simply wish to "label" the groups found by the k-clustering?

Results, P18, L405: We know Ketamine has paradoxical EEG effects, whereby it increases the DoH index while deepening clinical DoH. I don't know about dexmedetomidine, but it should in fact have the opposite effect on reducing BIS and MAC as it is synergistic with other sedatives?

Results, P18, L408: I am surprised to see Remi here, while all others make intuitive sense to me. - Might make an interesting discussion/future exploration.

Results, P18, L411: Well you excluded TIVA cases, which if you didn't would have shown interesting effects here.
Results, P19, L432: I am not surprised to see this, were you?

Results, P19, L446: Isn't this again an expected outcome, showing regression to mean when increasing the number of bins?

Results, P21, L465: Can part of this be explained by the supporting drugs needed to achieve this MAP, i.e. vasopressors and etAA?

Discussion, P21, L497: Pharmacokinetics? (and -dynamics for MAP, DoH etc might have different time constants)

Discussion, P23, L522: Is individual composite the right word? Also, what about different weights in the composite score or data transformations other than the z-score?

Discussion, P23, L533: So how could future researchers derive these weights?

Discussion, P25, L561: What is unrelated to TVI, MAC of 0.8?

Discussion, P25, L526: Yet you didn't perform any ROC of you method, which might allow you to comment on its positive (and negative) predictive value. Why not?

Table 1: Some of the presented data can be shortened by simply reporting median [q1,q3] (min, max) for all values.

Figure 1: It might be interesting to separate the anesthetic phases in induction, maintenance, emergence, here. Also what e they sorted by - the legend says randomly sampled, but you could sort by case length here?

Figure 3: The BIS and MAPs are surprisingly similar here; as ar the TVIs, where I would have expected the IQR to not overlap. Might simply be a large-n problem. Maybe consider showing the data as violin plots to allow the reader to appreciate the values in the whiskers?

Figure 4: The CBP periods are hard to see - maybe fill them in gray?

Figure 5: Again, I would have expected to see more discrimination in your score.

Figure 6: Would you want to speculate about the humps in MAC at -2, BIS at +3, which are interesting features?

Figure 7: It is surprising that MAC has almost zero effect on predicted morality, while MAP and BIS have significant ones.
Additional Materials - I have only glanced at most of these:

Additional Table 1, Additional Figure 1: Not sure these add much.

Additional Figures on different number of clusters - unless you do something with these, their utility is limited. Same for Table 7.

Additional Figure 9: This one looks interesting.

Additional Table 4: This one might be useful as well.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Unable to assess

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

Not relevant to this manuscript

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

Acceptable

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.