Author’s response to reviews

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

Authors:

Michael Schnetz (mps93@pitt.edu; schnetzmp@upmc.edu)

Harry Hochheiser (harryh@pitt.edu)

David Danks (ddanks@cmu.edu)

Douglas Landsittel (dpl12@pitt.edu)

Keith Vogt (vogtkm@upmc.edu)

James Ibinson (ibinsonjw@upmc.edu)

Steven Whitehurst (WhitehurstSL@anes.upmc.edu)

Sean McDermott (mcdermotts3@upmc.edu)

Melissa Giraldo Duque (giraldoduquem2@upmc.edu)

Ata Kaynar (kaynarm@upmc.edu)

Version: 1 Date: 17 Oct 2018

Author’s response to reviews:

Dear Dr. Vedula,

We would like to thank you and the reviewers for considering our recently submitted manuscript for publication and providing comments to improve the work. In response to this feedback, we have significantly revised our manuscript. The specific revisions are described below for each provided comment/critique. The corresponding changes in our manuscript are identified with Figure/Table numbers and/or Page/Line numbers. Changes made in the text of the manuscript are denoted using red text.
1. Were the physiological variables captured synchronously, i.e., for a given instance in time, were data simultaneously captured for MAP, BIS, and inhaled anesthetic concentrations. If no then what was done to synchronize the data in time?

- Unfortunately, our data were not consistently captured synchronously. Despite being monitored together during surgery, MAP, BIS, and end tidal inhaled concentration measurements were commonly recorded within one to two minutes of each other, but not at the same exact time. To synchronize the information within these roughly overlapping variables, we used a sliding window approach following data normalization. In the revised manuscript, we provide a figure (Additional Figure 1) that outlines our data processing steps. Non-synchronous data capture, per se, does not prevent the generation of TVI values. Page 9 line 194 explicitly describes the sliding window approach we used.

2. The last paragraph of the introduction describes the Methods, which do not belong here.

- This paragraph has been removed from the revised manuscript as suggested. All Methods-related information is contained within the Methods section of the revised manuscript.

3. On average, what fraction of TVI values were missing per procedure during cardiopulmonary bypass?

- The median proportion of missing TVI values due to cardiopulmonary bypass is 0.33. We have included this in our revised Results section in Table 1.

4. How did you determine whether or not variables were normal (and report mean vs. median)?

- We used histogram analyses in evaluating data distributions. For variables whose plots suggested a roughly normal distribution, we calculated a mean, otherwise we calculated a median. We clarify this fact in our revisions (Page 11, line 238).

5. I don't understand the 10000 bootstrap samples. What are they and why were they needed?

- We used bootstrapping to generate 95% confidence intervals for our calculated median statistics (e.g. median procedure length in each TVI pattern). Bootstrapping was performed by first generating 10,000 samples each containing 1,000,000 randomly selected values from
the original dataset. For each sample, a median was calculated. The 2.5 and 97.5 quantiles of the population of 10,000 medians defined 95% CI. Bootstrapping was not necessary to generate 95% CI for mean statistics. We used bootstrapping as a flexible, nonparametric approach to compare medians associated with TVI patterns using 95% confidence intervals without making strong assumptions about the data.

6. I would like to see some more systematic descriptive analyses before proceeding to analyze associations with outcomes. For example, what was the variance in variables within and across clusters? Does this approach really separate the data?

   - In our revised manuscript, we compare the variable distributions for each cluster to that of the total study population using violin plots (Figure 3). TVI, compared to MAP, BIS, and MAC variables, best separates surgeries between clusters.

7. Mortality analysis - the outcomes are incongruent - the first three are about death, the last is about being alive. Why is it this way?

   - We thank the reviewer for this comment and agree the presented outcomes were incongruent. We have removed the survival data from Table 3. In our revised manuscript, we only report mortality data.

8. Why is the proportion of profiles associated to different clusters clinically meaningful?

   - The proportion of profiles associated with different clusters is not clinically meaningful and these data have been removed from Table 2.

9. To analyze association with outcomes, I would suggest a prediction modeling approach instead of doing a k-means and comparing proportion of profiles assigned to different clusters.

   - In our revised manuscript, we compare TVI to the triple low state as prediction models for 30-day mortality. Previous studies have shown cumulative exposure to the triple low state is an independent predictor of 30-day mortality. This comparison is shown in Figure 7. For the TVI model, we used the median TVI value generated during surgery. For the triple low state model, we calculate the number of profile windows that met triple low state criteria (MAP
<75 mmHg, BIS <45, and MAC <0.8) for each surgery. TVI better discriminates between patients that died and those that survived the 30 days following surgery than the triple low state in our study population.

10. There is a lot of analyses described in this paper. At this time, it is hard for me to suggest ways to streamline the presentation to make it easily accessible to readers. Kindly consider this observation while revising your manuscript.

- We thank the reviewer for this comment and have reduced the number of analyses presented in the revised manuscript. In total, we present 16 total figures/tables including those as additional material compared to 28 total figures/tables in the original submission.

11. The TVI values, presumably multiple numbers for each procedure, are correlated within each procedure. Isn't this correct? I'm trying to understand whether your input to the k-means accounts for this correlation in the data. If not then consider discussing how it affects your findings and conclusions.

- Yes, that is correct. TVI values within a given surgery are not independent measurements as they share several key features like being generated from one patient undergoing one specific procedure. Our analysis using k-means clustering is not adjusted to account for such correlation because our goal is to distinguish patients according to their pattern of repeating TVI values generated over time. We specifically address this issue in our revised discussion section as suggested by the reviewer (Page 22, Line 483).

12. Discussion - fourth paragraph - "Our findings are consistent with previously published data."

- I'd like to see more discussion on what was similar and different between your work and previous research in terms of the question and methodology.

- We have revised both the Introduction (page 5, Line 106-124) and Discussion (page 23, Line 514-534) sections in our revised manuscript to highlight the similarities and differences between our work and that of previous investigations.

Briefly, the established model for combined MAP, BIS, and MAC levels is the triple low state. Although it provides valuable risk information, the triple low state is ill-equipped to broadly characterize the collective behavior of these variables during surgery because it 1) has only been evaluated in select surgical populations (e.g. non-cardiac surgery), 2) is based on thresholds and thus uses only a small subset of the total data available, and 3) does not account for how
combinations occur across time. TVI overcomes these limitations and can be used to more broadly evaluate MAP, BIS, and MAC values that occur together during surgery and identify their association to key patient, procedure, and outcome characteristics.

13. I think the major limitation to consider in your approach is whether and how well you have addressed or utilized the fact that observations within a procedure are correlated, basically time series data. Your contributions, pending clarification on its comparison with Sessler, et.al., is an index derived from routinely captured physiological measurements in the operating room, and a descriptive analysis of the index. The utility of the index to predict clinical events, I think, should be evaluated with time series analysis methods to develop and validate prediction models.

- We agree with the reviewer’s general description of our contributions represented in the manuscript. In the revised version, we show TVI is a better predictor of 30-day postoperative mortality than cumulative exposure to the triple low state (Figure 7). Considering the many excellent prediction models for postoperative mortality that already exist such as RSI and RQI, TVI may be used in an alternative way that is clinically useful. TVI may be used to further explore the relationship between a patient’s response to inhaled anesthetics and intraoperative hypotension. Our work shows patients that were most sensitive to inhaled anesthetic administration also exhibited the lowest intraoperative MAP levels, and as a result, the most intraoperative hypotension (IOH). IOH is an important modifiable risk factor linked to end-organ damage and postoperative death. In the future, TVI could be used as a tool to better characterize IOH in terms of event frequency, depth, and location, which in turn may help better predict and prevent IOH exposure. The discussion of this possibility is included in our revised discussion section (page 24, line 536-553).

14. I don’t understand the sixth paragraph in the Discussion - the first paragraph listing the contributions of this work.

- The sixth paragraph has been removed from our revised discussion section.

15. In the descriptive analysis, I’m not clear whether TVI patterns within surgical procedures were analyzed. If not then why not?

- We did not analyze TVI patterns within specific surgical procedures because we were limited by the number of surgeries available for a given procedure. The most common procedures in
our dataset are associated with only a few hundred surgeries each. Evaluating TVI expression patterns on an individual procedure basis may be useful to evaluate the index’s ability to distinguish patients at high and low risk of 30-day mortality, for example. However, a larger dataset is required for this type of analysis.

16. I would like to see a reasoned suggestion on next steps to develop and validate the TVI index for clinical use.

- As mentioned above, we include a more complete, detailed description of the potential utility of TVI as a predictor of 30-day postoperative mortality as well as a tool to investigate the relationship between a patient’s response to inhaled anesthetics and IOH exposure in our revised discussion section (page 24, line 536-553)

BMC Medical Research Methodology operates a policy of open peer review, which means that you will be able to see the names of the reviewers who provided the reports via the online peer review system. We encourage you to also view the reports there, via the action links on the left-hand side of the page, to see the names of the reviewers.

Reviewer reports:

Narges Ahmidi (Reviewer 1): The authors executed an impressive task of collecting a useful database for their study. However the analysis of the data is insufficient and not justified. The clinical relevance of their analytical decisions also were not mentioned. The paper itself lacked proper description of analysis which made it hard to comment properly how to improve them, for example, the definition of basic variable TVI profile was not given properly, which made the rest of the paper confusing. The decision to use an unsupervised algorithm was ambiguous where they could choose a supervised algorithm, because they had already reported outcome for patients.

It's possible that the authors had very good reasons for the set of analysis they did, but this was not communicated with the reader and therefore the paper just read as a report of analysis done and not their justifications and reasoning.

- We thank the reviewer for her comments and we have revised our manuscript to address the identified deficiencies. The clinical relevance of our work is that TVI provides a better tool to assess MAP, BIS, and MAC data, which when considered together provide important information about clinical outcomes. The current model used to assess these variables, the triple low state, has several key limitations, namely it 1) has only been evaluated in select
surgical populations (e.g. non-cardiac surgery), 2) is based on thresholds and thus uses only a small subset of the total data available, and 3) does not account for how combinations occur across time. TVI overcomes these limitations and can be used to more broadly evaluate MAP, BIS, and MAC values that occur together during surgery and identify their association to key patient, procedure, and outcome characteristics. In addition, we show TVI is a better predictor of 30-day mortality than cumulative exposure to the triple low state (Figure 7). We have specifically revised our Introduction (page 5, line 106-124) and discussion (page 23, lines 514-553) sections to highlight the distinction between TVI and the triple low state and why this is clinically relevant.

A TVI profile is the collection of sequential TVI values generated from MAP, BIS, and MAC values measured and recorded over time during surgery. The creation of a TVI profile is described on Page 10, line 190-217. In our revised manuscript, we also include a figure that visually demonstrates how we create TVI profiles from MAP, BIS, and MAC data (Additional Figure 1). We used an unsupervised algorithm because our experimental goal was to identify patterns of TVI expression, not TVI values or profiles that shared the same outcome. This is an important difference because TVI expression is meaningful beyond its association to postoperative death. TVI patterns, for example, are associated with different levels of exposure to intraoperative hypotension (IOH) during surgery, an important adverse event related to end organ damage including myocardial infarction, acute kidney injury, and stroke. In our revised discussion (Page 24, line 536-553), we discuss how TVI may be used as a tool to better understand the relationship between a patient’s response to inhaled anesthetics and IOH exposure in ways that may improve our ability to predict and prevent IOH events.

Elias Rizk (Reviewer 2): This is an interesting analysis and development of a new intraoperative tool to reveal key information intraoperatively and affect patient outcome over time. This method was based on a analysis of a very large cohort that allowed the authors to reach significant conclusions. The next step is to use the TVI as a positive or negative predictive value in different hospital settings to predict patient outcomes. If this is found to be significant then this would be ideal since intraoperative monitoring tool to allow anesthesiologist interventions to directly affect outcomes.

- We thank the reviewer for their comments.
Matthias Görges, Ph.D. (Reviewer 3): 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?

- MAP, BIS, and MAC variables, when considered together, provide important clinical information related to postsurgical outcomes. The only established model for the combination of these variables is the triple low state. We present a new approach to assessing these variables in combination and in our revised manuscript directly compare approaches as prediction models for 30-day postoperative mortality (Figure 7). The Triple Variable Index outperforms the triple low state in this comparison. We agree that the Triple Variable Index represents an unadjusted measure of a patient’s response to inhaled anesthetics and a portion of our characterization is focused on outcomes. However, the significance of our work is that we present a new approach to assessing MAP, BIS, and MAC behavior that occurs during surgery and in ways not possible using the triple low state model. We clarify how the triple low state is limited as a methodology in our revised Introduction (Page 5, lines 106-124) and highlight the differences between the triple low state and our presented Triple Variable Index in our revised discussion (Page 23, lines 514-553).

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.

- We have included a new Figure (Additional Figure 1) in our revised manuscript that details how raw data, after artifact removal, are converted to TVI values and plotted across time for
In creating TVI, we were required to make several basic assumptions related to combining MAP, BIS, and MAC data into a single index value. Specifically, we decided to use an equally weighted score for each variable, calculated a mean within our sliding window, and defined our window length as five consecutive data (MAP, BIS, and/or MAC) measurements. We agree each assumption affects our TVI signal and potentially our overall results. In our revised discussion, we address each issue and its potential impact (Page 22, lines 492-512). Manipulating variable weights may be particularly important towards better predicting clinical outcomes and feature selection/extraction techniques may be applied to define such weights. Increasing window length would reduce the amount of variation between TVI values within an individual surgery, however we would argue this would not significantly affect the overall TVI differences observed between patterns. We acknowledge it is important to test the effect of these assumptions in future studies.

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?

- We thank the reviewer for this comment and have clarified the order of our associations in our revised manuscript. Our goal is to generate a data-driven index from MAP, BIS, and MAC monitoring data and characterize groups of surgeries that share a similar pattern of index expression across time. Our index is predictive of 30-day mortality, but our goal was not to develop a novel predictive model for postoperative death. Instead, postoperative mortality is one of many key clinical characteristics that help define each identified expression pattern. The first paragraph of our revised Discussion section more clearly emphasizes this point (page 20, lines 435-442).

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.

- We have reduced the number of analyses presented in the revised manuscript. In total, we present 16 total figures/tables including those as additional material compared to 28 total figures/tables in the original submission.
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?

- Title has been changed to emphasize the identification of patients demonstrating distinct patterns of physiology: “The Triple Variable Index combines information generated over time from common monitoring variables to identify patients expressing distinct patterns of intraoperative physiology.”

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

- ‘Systems’ changed to ‘variables’.

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?

- Sample frequency explicitly noted. Continuously monitored variables, such as end tidal inhaled anesthetic concentration and blood pressure captured using an arterial line, are not recorded continuously, thus limited the amount of available data. NIBP does not appear to be the “rate-limiting” step for generating TVI data.

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.

- Age data, along with data related to the other characteristics, reported for each TVI pattern.

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?

- Wording revised to report patterns associated with outcomes.

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?
This experiment has been removed from the revised manuscript. As suggested by the editor above, the association between TVI expression patterns and postoperative death was further evaluated as a prediction model and compared to that of the triple low state (Figure 7).

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.

- “System” terminology removed (page 5, line 98).

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?

- MAP, BIS, and MAC values were weighed equally to generate TVI data. We revised the introduction to explicitly state this (Page 6, line 120).

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

- Individual paragraph was dedicated to aims (page 6, line 126).

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

- No, providing “Method” information here is not particularly helpful. We have removed this information from the Background section.

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

- We extracted two pieces of data related to the performed surgical procedure: the type of procedure (e.g. laparoscopic cholecystectomy) and the surgical specialty that performed the procedure (e.g. general surgery). We have revised the section to explicitly state this (page 7, line 152)

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.

- We agree with the reviewer that the time of administration and pharmacokinetic properties represent significant factors one should consider in evaluating a medication’s effect on TVI expression. Evaluating medication administration as a binary event (administered/not administered) has clear limitations, however, this type of analysis is required to identify how medications are being used between patterns on a very basic level. To our knowledge, our study is the first to evaluate the use of commonly administered intraoperative medications across multiple drug classes that potentially affect MAP, BIS, and MAC combinations. Sessler and colleagues attempted to adjust their triple low state model according to only two potential effectors (propofol and fentanyl), while other triple low state studies did not attempt to account for any intraoperative medications. We argue our analysis, although limited in clear and important ways, represents a key first step from which more sophisticated hypotheses can be generated and tested in future studies. We have included in our revised discussion the need for follow-up studies to address this important aspect of the work (page 23, line 504-512).

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.

- In evaluating the association between TVI expression and postoperative death, we made no initial assumptions about the data. In looking at the available evidence from the triple low state studies, 30-day postoperative mortality was the most common period used to evaluate postoperative death. However, Willingham et al and Kertai et al demonstrate postoperative mortality differences beyond 30 days and up to 2 years following surgery between their triple low state exposure groups. Based on these data, we sought to evaluate postoperative mortality within the 2 years following surgery. In our revised manuscript, we compare the Triple Variable Index and the triple low state as prediction models for 30-day postoperative mortality. We use 30-Day mortality for this particular analysis because this was the outcome used to establish the triple low state as an independent risk factor for postoperative death. We have revised our Methods section to clarify this issue (Page 12, line 267-286).

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.
Our anesthetic record system does not record all measurements that occur during surgery and MAP, BIS, and end tidal inhaled concentrations measurements were not always measured/recorded simultaneously. To combine MAP, BIS, and MAC data that exists in this way, we used a non-overlapping, sliding window that calculates an average value for each variable within a given period of time (page 9, Line 194). A window is defined by five consecutive MAP, BIS, and/or MAC measurements. The sample frequencies for each variable and the median length, in minutes, of the sliding window are presented in our revised manuscript in Table 1. In addition, we provide a new figure (Additional Figure 1) that shows how raw MAP, BIS, and MAC data, after artifact removal, are processed to generate TVI profiles.

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?

- We acknowledge that many factors affect the relationships between MAP, BIS, and MAC variables. This is important to consider when associating these data to clinical outcomes, for example, as factors like age confound the association. However, our goal was not to identify MAP, BIS, and MAC combinations as independent predictors of death. Instead, we sought to develop a better model to assess the unadjusted variable combinations that occur during surgery compared to the current model, the triple low state. This is important because we know MAP, BIS, and MAC combinations provide important clinical information beyond blood pressure and anesthetic depth monitoring, yet this combined behavior has not been fully characterized. We would expect a decrease in the difference in TVI values between patterns if age-adjusted MAC values were used. This is because MAC values would increase relative to BIS and MAC values as age increased. The TVI values in the depressed pattern would likely increase relative to the TVI values in the other patterns because the depressed pattern was associated with the oldest patients. Future studies are needed to assess the exact effect of using age-adjusted MAC in the generation of TVI data. We have specifically addressed this issue in our revised discussion (Page 23, Line 520-529).

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.

- Agreed, however this scenario does not seem to be well represented in our dataset. There exist 314 MAC measurements less than 0.005 and greater than 0.001 out of a total of 199,311 MAC measurements (0.1%). These 314 measurements were recorded in 204 individual surgeries and only 6 surgeries contained more than 5 such measurements. Considering these
surgeries as TIVA conversions, they represent approximately 0.1% (6 out of 5296) of study surgeries.

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.

- We agree that transducer repositioning and electrosurgical interference may have an effect on recorded MAP and BIS data, respectively. However, these events are not explicitly identified in the anesthesia electronic record system and therefore we cannot account for in our dataset (unless they result in measurement values outside those used to define artifacts in the dataset). BIS=0 appear as artifacts in our system. These measurements occurred at the very beginning and end of the monitoring period despite likely representing two fundamentally different phases of the anesthetic, induction vs emergence. It is possible, although difficult verify, that a BIS of zero represents a common artifact encountered in initiating and discontinuing BIS monitoring within our system. These points are clarified in the Method section (Page 8, Line 180-187).

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).

- We did not perform formal normality tests on these variables. We normalized these variables using a z-score because they each demonstrated an approximately normal distribution according to histogram analysis. We have more accurately stated this in our revised Method section (Page 9, line 191).

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?

- The sliding window we apply is based on sequential measurements of MAP, BIS, and end tidal inhaled anesthetic concentrations and is defined by 5 consecutive measurements. Because frequencies at which these variables are measured and recorded vary, the time represented by a given sliding window varies accordingly. A window representing measurements recorded more frequently will represent a smaller period of time compared to a window representing measurements recorded less frequently. We report the distribution (ie median, IQR) of window length, in minutes, in Table 1. We agree with the reviewer that median is less sensitive to extreme values than mean. It is possible calculating a median
instead of a mean will reduce the variation between windows because a median will be less sensitive to the range of values present in the underlying data. This may lead to more consistent TVI values captured across time for a given surgery and may have the effect of “smoothing” the observed data. Because median, like mean, reflects the underlying data, it seems unlikely the use of median instead of mean would significantly affect our overall findings: patients express relatively distinct TVI values across time during surgery. This effect of median compared to mean should be directly tested in follow up studies. We have noted this in our revised Discussion Section (Page 22, Line 496-504).

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).
- Yes, happy to report an example as recommended by the reviewer. We have included this in Additional Figure 1.

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?
- We did not plot TVI signal within established induction/maintenance/emergence phases. Unfortunately, these data are difficult to retrieve in our anesthesia electronic medical record system and are victim to incompleteness. We agree with the reviewer that evaluating TVI expression relative to standard times/events would permit a better approach to directly comparing individual surgeries.

Methods, P10, L217: Quickly say why this approach is particularly suitable for your application/setting/etc.
- We have described why this approach is suitable in our revised Methods section (Page 10, Line 224).

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?
- We calculated (and reported) group characteristic statistics. We did not perform Wilcoxon tests.
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.

The triple low state criteria, as originally defined by Sessler and colleagues, are: MAP <75 mmHg, BIS <45, and MAC <0.8. We list these criteria in our Method section and provide reference to Sessler et al (Page 12, Line 267).

Methods, P13, L279: Did you attempt to adjust for the (temporal) remoteness of the intervention and outcome here?

- We did not attempt to adjust for the temporal remoteness between TVI expression and postoperative mortality. The triple low state literature connects triple low state exposure to death at multiple timepoints in the postoperative period. Based on these data, it’s unclear how to adjust the association between TVI expression and postoperative death as time away from surgery increases. As a simple alternative, we examined the association within a period where an association exists for the triple low state, 2 years following surgery. To the reviewer’s point, we show the association between expression and postoperative death does not exist after 30 postoperative days, suggesting expression provides transient mortality information. However, this was difficult to anticipate prior to performing the analysis.

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?

- In our revised manuscript, these analyses were replaced with another analysis that compared TVI expression to the triple low state in predicting 30-day mortality. We create an ROC for each model and compare their AUC. In Figure 7, we show TVI is a better predictor of 30-day postoperative mortality than cumulative triple low state exposure. Associated description of this analysis in provided in our revised methods (Page 13, Line 286).

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

- In lieu of our presentation and comparison of TVI expression and the triple low state as prediction models for 30-Day postoperative mortality, the utility of the analysis referenced here became significantly reduced. As a result, we removed it from the revised manuscript.
Methods, P13, L302: Wouldn't you expect better performance (at least sensitivity) if you had used a higher thresholds than the mode?

- In lieu of our presentation and comparison of TVI expression and the triple low state as prediction models for 30-Day postoperative mortality, the utility of the analysis referenced here became significantly reduced. As a result, we removed it from the revised manuscript.

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?

- We acknowledge the fact that BIS monitoring and inhaled anesthetics are not used for all patients in the study period and the decision to use either is not random. This limits the overall applicability of our findings to other patient populations. Our presented data represents a key first analysis to which the TVI expression from other patient populations, including those from different time periods and institutions, may be compared. It’s unclear how the risk of postoperative death associated with patients that underwent TIVA during our study period compares to that of patients that were administered inhaled anesthetics allowing TVI expression to be measured. Besides the difference in anesthetic technique, these patients likely differ in many ways that potentially influence their risk of postoperative death. To use TVI to assess risk of postoperative death in patients that may not be well represented in our study population, TVI expression should to be assessed in additional patient populations of interest.

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.

- IQR and SD noted in the revised Results section. Proportion of males in the study population explicitly noted.

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

- Sample frequency of each variable is a function of variable measurement and the data recording parameters for our anesthesia electronic record system. Continuously monitored variables such as end tidal anesthesia concentrations and BIS are only recorded at a set frequency. For these study surgeries, that frequency was approximately one measurement every 5 minutes. Non-continuously monitored variables, such as NIBP, and measurements
made manually were recorded as they were measured. For these reasons, MAP data were recorded at a different frequency than MAC and BIS data.

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

- The minimum procedure length cannot necessarily be calculated from the number of TVI values generated for the procedure. A given TVI value represents five consecutive monitoring measurements from MAP, BIS, and/or MAC variables. The time period represented by this value depends on the frequency of monitoring, but in our dataset would not be less than 5 minutes. The median length of time (as well as the first and third quartile) for the sliding windows used to calculate TVI values is shown in Table 1 in our revised Results section. The minimum procedure length in our dataset was 12 minutes.

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

- Z-scores were summed to generate TVI values. The sentence referenced in this comment has been clarified in our revised Results section.

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?

- Of the possible variable pairs (MAP, BIS, MAC), we agree the BIS-MAC relationship would be expected to show the highest correlation because the variables provide related information. The strength of correlation beyond this relative comparison is difficult to assess. There exist factors, such as electrosurgical interference and ketamine administration that would influence BIS while not necessarily affecting end tidal inhaled anesthetic concentrations. These factors, and possibly others, would be expected to reduce the BIS-MAC correlation.

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?

- We are attempting to label our identified clusters, elevated, mixed, and depressed, and identify their associated characteristics. In the analyses referenced in this comment, we evaluated the distribution of TVI values between groups of surgeries differentiated by ASA
physical status classification, repeat and emergent surgery, and patient age. However, considering the reviewer’s feedback and our experimental goal of identifying clusters and their associated characteristics, these analyses are not particularly useful. We have removed these analyses from our revised Results section.

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?

- Yes, we would expect dexmedetomidine (and other medications like opioids) to potentiate the effect of volatile anesthetics and may cause TVI expression to decrease by lowering the MAC level relative to BIS and MAP levels. Interestingly, the elevated pattern was associated with medications that may affect TVI expression different ways. Ketamine and dexemdetomidine are a good example of this. Although our study is the first to examine many commonly administered intraoperative medications in relation to MAP, BIS, and MAC combinations, we caution our analysis is limited and our results need to be interpreted with caution. We assess only whether a medication was administered at least once (and the median total dose administered) during surgery. As the reviewer points out, the pharmacodynamics and pharmacokinetics of medications would need to be considered to more clearly understand the relationship between medication administration and TVI expression. Prospective, controlled studies are required to identify cause-effect. We have revised our discussion section to emphasize this analysis, its limitations, and related future studies (Page 20 Lines 455-471, Page 23 Lines 504-512).

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.

- It is unclear why remifentanil is associated with depressed pattern surgeries, but it is possible remifentanil use may reflect some common practice pattern at our institution that pre-selects patients more likely to demonstrate a depressed TVI pattern. In general, opioids would be expected to potentiate the effect of volatile anesthetics thereby reducing TVI expression. As noted in our comment above, we are the first to report medication associations and these data can be used to conduct important follow up studies related to individual drug effects on TVI expression. We have revised our discussion section to emphasize this analysis, its limitations, and related future studies (Page 20 Lines 455-471, Page 23 Lines 504-512).
Results, P18, L411: Well you excluded TIVA cases, which if you didn't would have shown interesting effects here.

- TIVA cases were effectively excluded because they lack MAC data, a requirement for TVI data generation. A distinct analysis of non-volatile general anesthetics in a similar framework would be interesting, but was beyond the scope of this initial descriptive work.

Results, P19, L432: I am not surprised to see this, were you?

- We do not find our postoperative mortality associations unexpected considering the triple low state studies have demonstrated a clear association with 30-day postoperative mortality. However, it is unclear from previous studies how long from the time of surgery the association between TVI expression and postoperative death would remain significantly different between patterns. Kertai and colleagues performed a Kaplan-Meier analysis between groups exposed to different amounts of the triple low state during surgery. The unadjusted Kaplan-Meier curves were distinct between the lowest and highest exposure groups up to 2 years following surgery. In addition, these curves show a different rate of death between the groups over this time suggesting mortality was different during postoperative periods beyond 30 days. Our data suggest TVI patterns are characterized by a host of factors that may influence both short- and long-term death. Age and ASA physical status classification are examples. Although TVI patterns do not demonstrate a difference in mortality beyond 30 days following surgery, it was unclear if this was to be the case before completing the analysis.

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

- In light of the reviewer’s feedback and further consideration, we feel the analysis presented in Additional Table 6 is of limited value and have removed it from our revised Results section.

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

- In lieu of our presentation and comparison of TVI expression and the triple low state as prediction models for 30-Day postoperative mortality, the utility of the analysis referenced here became significantly reduced. As a result, we removed it from the revised manuscript.
Discussion, P21, L497: Pharmacokinetics? (and -dynamics for MAP, DoH etc might have different time constants)

- We agree with the reviewer. The pharmacodynamics and pharmacokinetics of medications would need to be considered to more clearly understand the relationship between medication administration and TVI expression. Prospective, controlled studies are required to identify cause-effect. We have revised our discussion section to emphasize our medication analysis, its limitations, and related future studies (Page 20 Lines 455-471, Page 23 Lines 504-512).

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?

- In our revised discussion, we note that the “ideal” variable weights used to generate TVI values may differ from our approach that used equal weights. We suggest the application of developed feature selection/extraction techniques that would allow us to identify the most relevant information within these variables to better predict clinical outcomes (i.e. 30-day mortality). Page 22, Lines 492-496.

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

- As noted in the above comment, we suggest the use of feature selection/extraction algorithms.

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

- The referenced analysis has been removed from the revised manuscript.

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?

- In our revised manuscript, we compared TVI expression to the triple low state in predicting 30-day mortality. We create an ROC for each model and compared their AUCs. In Figure 7, we show TVI is a better predictor of 30-day postoperative mortality than cumulative triple low state exposure. In our revised discussion section, we explicitly compare our work to the triple low state body of literature (Page 23 Lines 514-553)
Table 1: Some of the presented data can be shortened by simply reporting median [q1,q3] (min, max) for all values.

- Table 1 has been revised as suggested.

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?

- We acknowledge comparing TVI expression within standard phases of the anesthetic would be useful, unfortunately data demarcating such phases is difficult to extract and are incomplete in our anesthesia electronic record system. Figure 1 is presented, in part, to demonstrate that a sample of TVI profiles is difficult to interpret and applying machine learning techniques to these data is useful to identify meaningful patterns that can be studied further. Surgeries of each pattern exhibit different procedure lengths, therefore we did not order profiles according to procedure length as this may work to artificially group profiles according to similar patterns of TVI expression.

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?

- Boxplots have been converted into violin plots in our revised manuscript as suggested by the reviewer.

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

- CPB periods have been filled in grey in our revised manuscript as suggested by the reviewer.

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

- It is difficult to anticipate the differences that may be observed in TVI expression between groups of surgeries. In these data, we observed differences in TVI expression between noncardiac, cardiac without CPB, and cardiac with CPB surgeries (Figure 6 and Additional Table 3), however these differences were smaller than those that distinguish TVI patterns overall. This suggests the type of surgery (noncardiac versus cardiac) and CPB use represent
only a few of the potential factors that influence TVI expression and contribute to the observed differences between patterns.

Figure 6: Would you want to speculate about the humps in MAC at -2, BIS at +3, which are interesting features?
- The referenced analysis has been removed from the revised manuscript.

Figure 7: It is surprising that MAC has almost zero effect on predicted morality, while MAP and BIS have significant ones.
- The referenced analysis has been removed from the revised manuscript.

Additional Materials - I have only glanced at most of these:

Additional Table 1, Additional Figure 1: Not sure these add much.
- We have removed Additional Figure 1 and revised Table 1 (instead of showing confidence intervals, we show standard deviations). The purpose of Additional Table 1 is to demonstrate the similarity in MAP, BIS, and MAC values between the subset of surgeries included in TVI analysis and all of the surgeries available in the study period. These two groups of surgeries did not exhibit large differences in these variables suggesting we have not selected a subset of surgeries with completely unique MAP, BIS, and MAC characteristics compared to the total population of surgeries that took place during the study period.

Additional Figures on different number of clusters - unless you do something with these, their utility is limited. Same for Table 7.
- We agree with the reviewer and have removed these analyses from our revised manuscript.

Additional Figure 9: This one looks interesting.
- This figure has been revised to show MAP, BIS, MAC, and TVI differences between patterns across the intraoperative period. The revised figure is now Figure 4.
Additional Table 4: This one might be useful as well.

- We have retained this table in our revised manuscript.