

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Variation in preoperative stress testing by patient, physician, and surgical type – a cohort study
AUTHORS	Pappas, Matthew; Sessler, Daniel; Auerbach, A; Kattan, Michael; Milinovich, Alex; Blackstone, Eugene; Rothberg, Michael B.

VERSION 1 – REVIEW

REVIEWER	McCluskey, Stuart University Health Network, Anesthesia and Pain Management
REVIEW RETURNED	09-Feb-2021

GENERAL COMMENTS	<p>The authors are to be congratulated for a thorough review of their own practice. There is no doubt the need for stress testing is variable and the drivers are the patient and the practitioner. This is not novel.</p> <p>However, the most compelling information in the manuscript is in the final line of the abstract, i.e. Use of preoperative stress testing appears to have decreased over time If the author were able to identify driver of this observation, I am sure the manuscript would be of wide interest. However, the most compelling information in the manuscript is in the final line of the abstract, i.e. Use of preoperative stress testing appears to have decreased over time If the author were able to identify driver of this observation, I am sure the manuscript would be of wide interest.</p>
-------------------------	---

REVIEWER	Valle, Javier VA Eastern Colorado Health Care System
REVIEW RETURNED	12-Feb-2021

GENERAL COMMENTS	<p>In this manuscript, the authors perform an analysis of preoperative stress testing at a single center, over a ten-year period. They sought to define the variability in preoperative stress testing at their institution, and to identify specific factors associated with the ordering of these tests. The authors found that predicted surgical risk (through prespecified calculable scores), patient functional status, previous diagnosis of IHD and other clinical factors were associated with the use of preoperative stress testing, as well as extremes of SES, and specific surgical subtypes (urologic, vascular, aortic). After adjustment, individual provider influence remained an important predictor and contributor to variance in the use of preoperative stress testing. The authors are to be commended for an expansive analysis with a robust data set.</p>
-------------------------	--

Comments:

1. While the authors successfully demonstrate the point here, it is not surprising that physician variation is the key driver of preoperative stress testing. As the authors are aware, observational datasets are limited insofar as their inability to address and identify unmeasured variables, which are particularly challenging in assessing angina. In the absence of objective measures (i.e. SAQ), the details of an anginal assessment by history are limited in multiple ways, to the eyes of the abstractor: the history obtained (detailed, nuanced, or brusque and dichotomous), the documentation, and the patient's forthrightness/willingness to answer. Each of these are influenced by the physician, and so it is understandable why variation appears to occur at the physician level (in either direction)- some are better than others at eliciting history, some are better at documenting it, and some are better at interpreting it. As the authors state in their discussion (although somewhat buried in the text)- "It seems probable that clinical decision-making is more nuanced than we can discern from our data source."

2. The authors utilize a prediction score to assess the likelihood of having obstructive CAD (ref 16). This score was utilized among patients without prior CAD or revascularization, with stable angina undergoing either CT or conventional angiography. As many of the patients in this evaluation did have prior CAD, I am unsure of the appropriateness of this assessment to predict the presence of obstructive CAD. I am also unclear how the authors determined chest pain type (typical vs atypical) from their dataset, without chart review- and how to account for missingness in this scenario. As authors use this score to comment on the calibration of clinicians' prediction of the presence of obstructive CAD, this should be addressed.

3. The majority of patients were evaluated prior to the publication of the current guidelines for preoperative stress testing, and so while current guidance might argue against the use of these tests, decision-making at the time they were ordered/performed may not have been influenced by the weight of present recommendations. This should be mentioned in the limitations.

4. The authors mention that the surgical subtypes included in their model predicting use of stress testing include PV and aortic surgeries, but then also mention that the majority of these (vascular) patients undergo evaluation elsewhere in their institution. This raises the high likelihood that there is something about this cohort of patients (i.e. they are higher risk in general to be referred to the specialty preop clinic, or are inappropriate in some way for usual evaluation) that would set them apart. I would recommend a sensitivity analysis excluding these patients that are typically evaluated elsewhere, in order to identify patterns and trends in usual care.

5. Functional status is raised as a concern- many patients achieving > 4 METs underwent stress testing. It is unclear what proportion of these patients expressed symptoms concerning for angina at their usual workload, in which case stress testing would be reasonable to assess of underlying IHD and potentially warrant revascularization for symptomatic benefit. This should be addressed.

6. One of the more interesting pieces of the analysis (physician experience) was removed from the model, but is presented in the supplemental data. I would suggest incorporating this into the main manuscript as an avenue for further investigation- experience appears to lead to less preoperative stress testing, but does it lead to more appropriate preoperative stress testing?

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:

We added as much about this observation to our abstract as we are able while remaining within word limits. Since our initial submission, a cross-sectional analysis of claims data among patients who completed elective hip and knee arthroplasty also suggested declining rates of preoperative testing over a similar time period. We have added a paragraph of discussion (pages 17-18) to contextualize the two studies. The recently published analysis of claims data has the strength of capturing patients from across the country rather than our single-center data, but that analysis includes only two specific surgeries, fails to capture patients who considered but did not complete surgery, and lacks the detailed clinical data required to control for patient risk and functional status. The two analyses thus have complementary strengths and different limitations. Reduced stress testing appears to be a broad change in clinical practice which, our analysis suggests, is not explained by the release of 2014 ACC/AHA guidance.

Reviewer #2:

“1. While the authors successfully demonstrate the point here, it is not surprising that physician variation is the key driver of preoperative stress testing. As the authors are aware, observational datasets are limited insofar as their inability to address and identify unmeasured variables, which are particularly challenging in assessing angina. In the absence of objective measures (i.e. SAQ), the details of an anginal assessment by history are limited in multiple ways, to the eyes of the abstractor: the history obtained (detailed, nuanced, or brusque and dichotomous), the documentation, and the patient's forthrightness/willingness to answer. Each of these are influenced by the physician, and so it is understandable why variation appears to occur at the physician level (in either direction)- some are better than others at eliciting history, some are better at documenting it, and some are better at interpreting it. As the authors state in their discussion (although somewhat buried in the text)- "It seems probable that clinical decision-making is more nuanced than we can discern from our data source.”

Dr. Valle correctly points out ways in which physicians might differ in their patient assessment, and is correct that we do not have structured data on chest pain characteristics (the Seattle Angina Questionnaire or others). We added further text to the paragraph in question (pages 15) to make this more apparent to the reader.

2. The authors utilize a prediction score to assess the likelihood of having obstructive CAD (ref 16). This score was utilized among patients without prior CAD or revascularization, with stable angina undergoing either CT or conventional angiography. As many of the patients in this evaluation did have prior CAD, I am unsure of the appropriateness of this assessment to predict the presence of obstructive CAD. I am also unclear how the authors determined chest pain type (typical vs atypical) from their dataset, without chart review- and how to account for missingness in this scenario. As authors use this score to comment on the calibration of clinicians' prediction of the presence of obstructive CAD, this should be addressed.

We added detail of our approach to the methods (pages 8-9). We believe the patient populations are sufficiently similar to allow use when we calculate the pretest probability (patients without an existing diagnosis of coronary artery disease who are considering elective surgery) and in the cohort used to derive the prediction tool (patients without a history of previous myocardial infarction, without a history of revascularization, and without an acute coronary syndrome referred for elective angiography or CT angiography).

3. The majority of patients were evaluated prior to the publication of the current guidelines for preoperative stress testing, and so while current guidance might argue against the use of these tests, decision-making at the time they were ordered/performed may not have been influenced by the weight of present recommendations. This should be mentioned in the limitations.

In our revised paragraph discussing changes in testing rates over time and contextualizing our study with another recent analysis (described more fully above), we now point out that neither recent study suggests a clear change at the time of guideline release. It seems likely that updated guidance reflected, rather than caused, trends toward more limited use of preoperative stress testing. As Dr. Valle points out, there can certainly be time lags between publication and adoption, a limitation which we now point out in that paragraph.

4. The authors mention that the surgical subtypes included in their model predicting use of stress testing include PV and aortic surgeries, but then also mention that the majority of these (vascular) patients undergo evaluation elsewhere in their institution. This raises the high likelihood that there is something about this cohort of patients (i.e. they are higher risk in general to be referred to the specialty preop clinic, or are inappropriate in some way for usual evaluation) that would set them apart. I would recommend a sensitivity analysis excluding these patients that are typically evaluated elsewhere, in order to identify patterns and trends in usual care.

We added the suggested sensitivity analysis. For space considerations, the results of said analysis are in the Supplemental Appendix, in a table comparing the mean marginal rates across all predictors other than surgical category. Broadly speaking, the results are similar when these patients are excluded. Effect directions are the same for all predictor variables and marginal rates changed only slightly. All changes in effect size are smaller than the smallest effect size in our base-case model (congestive heart failure).

5. Functional status is raised as a concern- many patients achieving > 4 METs underwent stress testing. It is unclear what proportion of these patients expressed symptoms concerning for angina at their usual workload, in which case stress testing would be reasonable to assess of underlying IHD and potentially warrant revascularization for symptomatic benefit. This should be addressed.

Dr. Valle points out that unstable angina is a specific problem that could prompt diagnostic testing in many settings, including our institution. We therefore made efforts to identify rates of unstable angina using clinical documentation. Many notes during this period used a structured template that included a field for "angina within 30 days", and we have used natural language processing to identify unstable angina in other records. These sections are now described on pages 9 (methods), 13 (results), and 17 (discussion). Clinical documentation of symptoms is far from perfect, but unstable angina appears to be present in a very small minority of visits to this clinic.

6. One of the more interesting pieces of the analysis (physician experience) was removed from the model, but is presented in the supplemental data. I would suggest incorporating this into the main manuscript as an avenue for further investigation- experience appears to lead to less preoperative stress

testing, but does it lead to more appropriate preoperative stress testing?

We added a suggestion that physician experience could be a fruitful avenue for further exploration (page 20). Between our prespecified analytic criteria (rejecting predictors based on information criteria) and other recent work that demonstrates decreasing rates of preoperative stress testing in other datasets (described above and on pages 17-18 of our revised manuscript), we believe that time is a better explanation than experience. We agree with Dr. Valle’s suggestion that the effects of experience on testing practices would be a fruitful avenue for further investigation, and we look forward to exploring it. However, that will require more analysis and space than will fit into the current manuscript.

VERSION 2 – REVIEW

REVIEWER	McCluskey, Stuart
REVIEW RETURNED	University Health Network, Anesthesia and Pain Management 03-Aug-2021
GENERAL COMMENTS	Congratulations on a well written and complete quality assurance document. It would be most interesting to know if the observations made have any implications to the future management of the clinic. For example, would the authors suggest closer adherence to current guidelines? It would also be valuable to know why this review was conducted? Was the clinic thought to be 'over-testing'?

VERSION 2 – AUTHOR RESPONSE

Reviewer #1:

This work is part of a larger research project investigating the causal effects of perioperative interventions on postoperative mortality. Advanced causal inference techniques that aim to separate the effect of an intervention (e.g., stress testing) from other confounders (e.g., risk of perioperative MI) require an estimate of the probability that a subject was assigned to the test/treatment. This paper identifies what prompts a patient to be selected for preoperative stress testing; our upcoming work uses the selection process identified in this paper to interrogate the causal effects of stress testing.

Many observational studies are plagued by unmeasured confounders, missing data, and uninterrogated selection processes. We hope that this paper, in conjunction with analyses still forthcoming, can demonstrate how our field can generate evidence of high quality using EHR data, in addition to generating direct evidence on the specific questions of preoperative stress testing. We will have more to say on the clinical implications of preoperative stress testing in upcoming papers; this paper is foundational.

Although I (Dr. Pappas) remain unconvinced that preoperative stress testing can play a useful role in preoperative decision-making, a number of my clinical colleagues are more sanguine than I am about such testing. My observation of our differences of opinion drew my attention to this question in 2016-2017, but I am not aware of a consensus view regarding the “correct” rate of preoperative stress testing or whether our group exceeded that.