Decision: rejection

The committee was interested in the topic of your research. The following concerns were mentioned: ••

o The committee shared most of the clinical reviewers criticism. •
o We did not think the morbidity data should be published as separate paper. •
o Please clarify in which way these results will be useful for general readers and public health experts. •
o Please justify the use of aggregate HDI data. Countries with high HDI includes countries with considerable social and economic disparities such as Sweden and Greece and Saudi Arabia. •
o We don't know what these differences mean, which we are not convinced are just down to differences in infrastructure, staffing or surgical standards of care. There could be differences, for example, in access (or lack of it) to pre-hospital emergency care. •
o We did not get a good picture of who was enrolled in the study. How were the centres selected? Who goes to these centres? Please provide a better description.
o The author list is very long. Please verify that all meet the criteria for authorship http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html. •
o How much adjustment was done for case mix? •
o The committee wasn’t sure of the statistics in your paper and decided formal statistical review was necessary.

I regret this review identified too many issues and the final decision was rejection therefore.

Reviewer Comments:

Reviewer: 1

Recommendation:

Comments:
The manuscript reports the results of an interesting global epidemiological study of emergency abdominal surgery stratified by United Nations development indices.

The manuscript is clearly presented and to my knowledge the first global study conducted in emergency abdominal surgery. It provides a good basis for policymakers making local or global health priorities and for clinicians optimizing care for the individual patient.
Comments:
1) Study period (page 5): An institution could collect

over as many two week periods as desired within the study period. How often did this take place and how did it affect data. In the sensitivity analysis you account for clustering of hospitals within countries, but will this account for multiple data sets from one institution? Please clarify.

2) Data (page 5): Please clarify the criteria's for excluding case report forms from analysis. Did you use the criteria reported in the protocol publication? If so, please report this in the manuscript.

3) Outcome measures (page 6): Regarding Clavien-Dindo (C-D) grade IV complications: please clarify that ''requiring critical care'' covers both patients actually admitted to intensive care and patients not admitted to critical care despite being indicated.

You state ''For this study, 24 hour mortality, 30 day mortality and the reintervention rate were selected for reporting''. Isn't it major complications (C-D grade 3-4) you are reporting, which covers both re-interventions and critical care admissions?

4) Statistical analysis (page 6): First of all I congratulate the authors that they managed to get a dataset with a high level of data completeness.

Please clarify how you handled missing data in your analyses, e.g. complete case analysis, multiple-imputation or other strategy. There is a drift toward a higher number of missing data from low HDI countries, and thereby data is not missing completely at random (71/1318= 5,4% which will add up to more if you do a complete case analysis). Please comment on bias in regard to this.

5) Results (page 8-9): You report that 9 centers and 161 patients were excluded prior to analysis. Please specify the reason(s) for exclusion e.g. duplicates, data inconsistency, not fulfilling data completeness criteria's ether in the text or in the flow chart?

Please report confidence intervals on the rates when reporting the main results (24h mortality rate, 30d mortality rate and major complication rates) as a significant p-value will only clarify if one of the three groups sticks out.

6) Discussion: Over all the discussion is well written with clear messages, but it is lacking references to previous studies.

Page 11, Line 39-40: ''Major complication rates were lowest but the proportion of surgical (versus non-surgical) reintervention highest in the low HDI centres''. Please clarify, if non-surgical covers endoscopies or interventional radiology or something else. In the result section, the surgical reintervention rates in the high HDI and in the low HDI are similar.

7) Conclusion. In the abstract the authors writes that the result strongly supports 30-day outcomes as an international benchmark. This conclusion is not commented on in the discussion section. Please comment on this in relation to the recently published standards for definitions and use of

outcome measures (Jammer I et al, Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: European Perioperative Clinical Outcome (EPCO) definitions. Eur J Anaesthesiol 2015; 32: 88-105.). In this it is outlined that 30-day mortality may not be adequate as a measure of outcome and that 90-day and/or one year mortality may be more beneficial.
8) Tables and figures: I will recommend that some of the tables and figures are moved to supplemental files only accessible via the internet e.g. table 4, table 9, table 10, figure 3 and figure 4

Minor comments:
Study design: please state that the study has been reported in regard to the STROBE statement
Page 9, line 23: P-value is missing
Table 2: Please report age as median and not mean. Please report that perforated viscus is diagnosed preoperatively and not during surgery as reported in on page 7 line 10-13
Table 9, footnote: Please write Full model instead of final model as this is the terminology used in the manuscript and previous tables

Congratulation with a very interesting study!

Reviewer: 2

Recommendation:

Comments:
For the Editor
This is a valuable study, it should be published but it is too long and risks spoiling its main messages by the inclusion of morbidity outcome as well as mortality.
Recommend: reinvite submission of a shorter paper with mortality as the outcome measure but not the major morbidities.
Once paper accepted on mortality I would look favourably on receiving one on morbidities and personally I would be willing to review it.
If this paper is accepted there should certainly be an accompanying editorial. I would be willing to write this although there are many other appropriate alternative surgeons.
Pages 15-24 of this draft is extreme despite this being a global paper. I think there is a difference in what is perceived as a right of authorship in the different countries but I believe this list can only be put up on a website and the main authors need to find an appropriate way of managing this.
Authorship and Acknowledgments: TOO MANY PEOPLE!! I do not believe that everyone listed has contributed to the paper though they may be the bedcard on-call consultant for the

cases. That does not constitute a right to be listed as an author. I realise there are 357 insitutions but I am not sure there needs to be one chosen from every institution and this list needs to be trimmed. However, hopefully the BMJ can find some way of managing this because global collaboration in studies is really important.
Tables must be simplified and not so many.
Figure 1 is 4 figures all of which repeat information in tables. Allow one at most.

For the Authors
Relevance: This is a topic that is of international importance - the global outcomes of laparotomy.
There are considerable challenges in addressing this topic because of differences in casemix, presentation, and resources available for treatment.
Methodology: The methodology has resulted in submitting institutions with enthusiasts providing the laparotomy data. Although this is likely to result in some selection bias (probably towards better outcomes), it is also reassuring that the data submission by local champions will result in accurate and complete data for the two week period (s) selected. Thus the many views into the outcomes of the different hospitals in different settings are likely to be authentic. I therefore approve of the design and congratulate the authors for selecting this method.
It is my opinion that despite having a self-selected source of emergency laparotomies the 10,000 laparotomies mean this is a significant number, their findings should not be ignored and although there may be other ways to reassess emerg laparotomy outcomes in individual countries this was a pretty good way to attempt it globally.
Statistical methods look appropriate
Presentation and analysis of Results:
The question I asked myself over the past two weeks about this study is whether the authors have tried to pack too much in to this paper. For example, the paper would be simpler if they only presented their data on mortality as a single outcome, stratify its rate by various factors and by duration after surgery as they have done. The combination of mortality and morbiditiy outcomes has a tendency to confuse, because morbidities (eg intestinal fistula) that are not survivable in LICs - eg the need for TPN, may well be survivable in HICs.
As it turns out they had no space for morbiidity in the abstract in any case.
Another point about morbidity and the interventions (I support their use of Clavien-Dindo and limiting comps to Grades 3 and 4) is that endoscopy and radiological interventions are largely only going to be available in HICs or selected hospitals in MICs. - this morbidity data is superfluous to the main messages of this paper about mortality.
Discussion

If morbidity left out is a good length
Tables: 10 tables are too many: Table 8 could be out
Table 3 is critical: The conditions treated are somewhat dominated by acute appendicitis. I realise this is normal that acute appendicitis is the commonest emergency but then when one remembers there are 357 institutions why are there so few cholecystectomies and acute diverticular disease?
Why are there so few hernias in the LICs? Similarly Female reproductive causes - tubo-ovarian sepsis is normally really common in LICs.
Why are the causes of peritonitis - incl diverticular disease, perforation of the intestine, not almost as common as bowel obstruction? Peritonitis (?primary) as a cause got 8 out of over 10,000 cases.
The total numbers should be at the bottom of this table for each column.
Table 4: Simply by system and put more of the low number procedures as "others"
Suggest this table could be halved.
Table 5 on ASA is good. clear, and punchy. Message easy to understand.
Table 6: I think there are too many factors included. I would suggest limiting to 10.
Table 6 & 7 I believe these could be combined, reduce the information and as the number of cases is the same (column 2) for each have the 24hr and 30 day mortality in the next two columns and then present the stats for the 8-10 most important factors.
We do not want too many significant risk factors - risk stratification needs to be kept simple. Again they could make this the topic of another paper as it is common to look at 30 or 40 factors before deciding on the best half dozen.
Table 9 is an example of combining the two for hierarchical logistic regression but remove the column on morbidity if this is being omitted from a revision.
Table 10 - I would leave this out of this paper. Far too long and too many factors.
Figures: Figure 1 is 4 figures. Far too much info. Not necessary, or only use % trauma figure. But this is all a repeat of what is in tables.
I like and approve of figs 2-4.
Authors and collaborators and acknowledgments:
Pages 15-24 of this draft is extreme despite this being a global paper. I think there is a difference in what is perceived as a right of authorship in the different countries but I believe this list can only be put up on a website and the main authors need to find an appropriate way of managing this.
I do not believe that everyone listed has contributed to the paper though they may be the bedcard on-call consultant for the cases. That does not constitute a right to be listed as an author. I think one or at the most two people need to be chosen from every institution and this list needs to be trimmed.

Conclusion (in abstract): The second sentence of the conclusion is not quite right. I do not think it is prognosis but rather casemix. I do not think they have done the work to assess prognosis in different settings and countries. Prognosis depends on things like ASA and other factors but this is not the focus of this paper.

Reviewer: 3

Recommendation:

Comments:
Statistical Review
The authors report on an international survey of the practice and outcomes of emergency abdominal surgery across a variety of countries. The authors should be congratulated on completing this project.

However I have substantial concerns regarding the validity and reporting of their statistical analysis.
1. First the paper is titled ''variations in the practice and outcome'' but the current paper largely focuses on variation in outcome. It would be interesting for the authors more carefully characterise and describe the differences in practice. Some of the detail of this is given in the tables, but there is little summary in the text. The story that these data tell will be of interest. A greater level of granularity is needed in the data presentation to achieve this.

2. The abstract clearly implies that 24 hr mortality increases three-fold which does not change with adjustment. This is not what the adjusted estimates indicate in the tables in the paper, where adjustment for the full model substantially reduces the gradient observed. See below for concerns about these analyses.

3. The sample size calculation in the paper refers to information in the protocol. The summary in the paper (that 900 per HDI comparison provides considerable precision) differs considerably from what is written in the protocol (which is based on statistical power to detect a difference in survival based on different treatment practices).

4. Also I note that the aims of the project as described in the protocol differ considerably from that given in this paper. There is no clear aim stated that you would look to compare outcome based on HDI.

5. The authors use multivariable models to ''adjust'' for differences in the prognosis of patients when making comparisons in the mortality rates between the HDI groups. The validity of any such case mix adjustment relies on the

prognostic factors being identifiable, measured and available in the dataset. The analysis will also assume linearity and multiplicative effects when no interactions are included. The greater the degree of diversity in case mix the less likely these assumptions will be met.

6. Data presented in tables 2, 3 and 4 reveals substantial heterogeneity in the characteristics of patients between HDI groups, as well as the diagnoses, and procedures undertaken. Given the magnitude of these differences (and the small sample sizes in some of the categories) many would consider it highly optimistic to hope that any model would be able to adjust for the differences in prognosis that exist observed. I would suggest that the findings from these adjusted analyses cannot be considered reliable and should be downplayed in the paper. There is some reflection of this in the degree to which the estimates of the effects shift substantially with inclusion of increasing numbers of confounders in the baseline and full models.

7. As flagged up with the abstract, why do the authors not comment on the results from the second fully adjusted model in the text? The focus is on comparing the unadjusted model with the baseline model but do not mention the findings from the full model. The significance of the HDI differences disappears on 24 hr mortality when the full model is used. The abstract states that adjustment did not change these gradients which is not what is observed in the tables.

8. As an alternative approach the authors might alternatively consider looking to see whether they identify a single procedure where they can make comparisons between HDI groups. If a group can be identified where it can reasonably be argued that all key prognostic variables have been measured, and where there is adequate data, they might more reasonably be able to assess whether the differences associated with HDI persist. One example might be appendectomy. Even then it would be very important to be able to characterise the differences in patient presentation between centres. I note that they have attempted to look at some groups in this way already.

9. What is the impact of the differential selection of patients between centres? The proportions sampled from each country differ significantly. In the High HDI, the UK contributed more than 10 times as patients than any other country. Can the results for the high HDI category thus be thought to be representative, given all that is known about differences in health services? The lack of a structured sample frame to ensure representativeness of the sample is a major weakness of this study.

10. The presentation of the tables would be improved if percentages were given which facilitated comparison between

the columns rather than between the rows. For example, Table 2