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)	Payment query- Interpreting the Lancet Surgical Indicators in Somaliland: A Cross-Sectional Study
AUTHORS	Dahir, Shukri; Cotache-Condor, Cesia; Concepcion, Tessa; Mohamed, Mubarak; Poenaru, Dan; Adan Ismail, Edna; Leather, Andy; Rice, Henry; Smith, Emily

VERSION 1 – REVIEW

REVIEWER	Gregory Peck Rutgers University - USA
REVIEW RETURNED	19-Aug-2020

GENERAL COMMENTS	Congratulations to all authors on this collaborative work. This cross-sectional descriptive study assesses Somaliland surgical system using the LCoGS Indicators 1-3, and 5-6. I would like to understand a bit more about how a "hospital's capacity to perform surgery" is based on the presence of an OR, because this defines the basis of your assessment for the entire country. The mere presence of an OR indicates capacity? Indicator 1 uses GIS but I do not see an explanation of this analysis. What proportion of the population owns a vehicle, what public transportation exists, if any, for this to have some relevance and context, and to address the limitations you state with this indicator? Indicator 2 uses SAO density but includes sub-specialists that likely do not perform the bellwether procedures, e.g., 3 ENT and 1 neurosurgeon. If the volume of surgery i.e., healthcare service package is based only on the Bellwethers in your assessment, then are these surgeons relevant to your assessment? In another words, are they even related. The logbooks used included what data element? i.e., a CPT code, or was it free text and then interpreted how? where is that data? For example use of an open fracture for Indicator 3? open fracture of what? all open fractures? Where is the data per region, per hospital? Indicator 4 is omitted because of the difficulty in obtaining mortality data? Is there any mortality data that can be projected or modeled to fit your Ind 1 - 6 assessment for the country or a region? This may help extrapolate out a relationship of the indicators, even if a modeled mortality indicator is used. This would strengthen the paper because the use of the indicators separately has been described in the literature quite extensively without real known use or impact if not interdependently assessed (i.e., all 6 indicators is what a surgical system assessment includes based on all six system nodes, originally described by Blake et al.). For an example of an interdependent interpretation, I recommend Hanna et al. Use of the Six Core Surgical Indicators from the Lancet Commission on Global Surgery in Colombia: A Situational Analysis. Lancet Glob Health. 2020;8(5):e699-e710. This is the latest and most comprehensive example of "non-

	modeled" and interdependence and this work is not referenced in your paper (especially when you highlight the limitations of previous work. I suggest this for the purposes of interpreting your indicators in light of one another (for example, what associations exist between the indicators and regions?) For example, why does Sahil have the lowest SAO density, yet it has the highest delivery based on your indicator definitions? How does the impact indicators translate to an average income within the geographic areas you highlight so the indicators 5 and 6 have local context as a way you send a message for first interventions to country officials? A second example is in Sahil, again with the highest delivery, but with the best impact? How do you interpret this? Consider your use of GDP (national average) and this as a significant limitation to interpretation in the poorest regions that are several standard deviations away from this average. Or is it? Can you translate the GDP to a local context, along with the indicators in that local context for more relevance to your assessment? For indicator 5, you mention an average cost of healthcare as a proxy cost? Is this the average cost of medical or surgical admissions? Is this the average cost of a particular bundle of surgical interventions? Overall - I worry a bit about the use of the word "deficit" in the light of a HIC senior author assessing an LIC system. Is there a more culturally sensitive/competent term?

REVIEWER	Isabelle Citron Queen Victoria Hospital, East Grinstead
REVIEW RETURNED	23-Aug-2020

GENERAL COMMENTS	Review: Interpreting the Lancet Surgical Indicators in Somaliland The authors performed an analysis of surgical capacity at 16 hospitals in Somaliland using six indicators defined by the Lancet Commission on Global Surgery. These indicators are designed to assess preparedness, delivery and impact of surgical care. The study is clear, well written and has a number of strengths. Firstly, the paper fulfils an unmet need for data, as there is insufficient data on surgical systems worldwide. Secondly, the paper in comprehensive in that it assesses both the public and private system and includes all surgical facilities in the country. This is particularly important given the high proportion of surgical care provided in the private sector. Thirdly the discussion of the paper is nuanced and analytical, stressing the importance of surgery as part of a national health system and UHC as opposed to a “surgery first” approach. The authors justifiably question the validity of the very indicators they used, to measure what they are purported to measure: preparedness, delivery and impact. There are a few minor suggestions which could improve the paper: - The paper states as one of its strengths its methodology for calculation of financial risk protection stating it is not modelled data. However, the explanation of the methods for this may be a little unclear as given the cost and income data is not from the primary source (patients) and not from data they collected their estimates would also be “modelled”. From the methods it seems that the authors used estimates from another paper, which in turn used estimates from another paper, which in turn used estimates from another paper from 2010 which estimated cost of C-section,

	for their estimation of financial costs. They then applied GDP per capita as a substitute for income with no accounting for income distribution in the country, something most of the “modelling papers” do account for as income inequality is often quite significant in this context. There are differing regional estimates of financial risk protection which represent the different proportion of OOP expenditure reported by each hospital, as the cost of surgery and poverty line/GDP per capita are based on national estimates. It is not feasible for the authors change their methodology at this stage but instead I might suggest that they avoid highlighting it as a core strength of the paper and do not state its superiority to “modelled” data. The gold standard data source for FRP would be from direct patient reports of costs and income which is then used to make national estimates. - Given the general readership of the BMJ Open, it may help to include in the introduction some supporting literature regarding the cost effectiveness of surgery and the role of surgery within a health system and UHC to underscore why this is an important area of research - Given the general readership of BMJ Open it may be helpful to the reader to include a little more context on the endorsement of these indicators for example their inclusion (4/6) in the WHO core 100 and the World Development Indicators as these indicators have a clearer mandate for collection and reporting than that of the Lancet - Figure 2- although it makes clear the geographical area within the two hour catchment it does not show the population distribution. This might add more insight to the reader to get an understanding of the density of the population in the served and underserved areas. - Some data elements from this paper have been previously published by this group. This may need to be more clearly acknowledged.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:

Congratulations to all authors on this collaborative work. This cross-sectional descriptive study assesses Somaliland surgical system using the LCoGS Indicators 1-3, and 5-6.

1. I would like to understand a bit more about how a "hospital's capacity to perform surgery" is based on the presence of an OR, because this defines the basis of your assessment for the entire country. The mere presence of an OR indicates capacity?

RESPONSE:

We thank the reviewer's astute comments regarding the definition of surgical capacity in our manuscript. We used the definition of hospital capacity to perform surgery as the presence of at least one operating room as a cutoff point during the selection criteria process. A facility assessment tends to pick up on one of the three aspects of the Donabedian framework elements – that of structure and not process and outcome. However, the presence of an operating room does not indicate a comprehensive analysis of hospital capacity. Therefore, it was only used as an initial selection criterion on which the Lancet Commission on Global Surgery (LCoGS) indicators were applied.

In the text: The **METHODS** section was updated to include more description of hospital capacity. We also added discussion regarding this limitation in the **DISCUSSION** section.

2. Indicator 1 uses GIS but I do not see an explanation of this analysis. What proportion of the population owns a vehicle, what public transportation exists, if any, for this to have some relevance and context, and to address the limitations you state with this indicator?

RESPONSE:

We agree with the reviewer's astute comments regarding the need to include additional information to the already existent explanation for the GIS analysis. We included a description of the geospatial analysis in the methods section. We also consider that the comments regarding population that owns a vehicle and use public transport are very pertinent and relevant for the patient in terms of realized access versus geographic access. However, we consider this information is irrelevant in terms of our study methodology as we set out to use of the LCoGS indicators (with all their weaknesses) to assess the Somaliland surgical ecosystem.

In the text: The **METHODS** section was updated to include additional information regarding the geospatial analysis under the ***Indicator assessment: Indicator 1*** paragraph.

3. Indicator 2 uses SAO density but includes sub-specialists that likely do not perform the bellwether procedures, e.g., 3 ENT and 1 neurosurgeon. If the volume of surgery i.e., healthcare service package is based only on the Bellwethers in your assessment, then are these surgeons relevant to your assessment? In another words, are they even related. The logbooks used included what data element? i.e., a CPT code, or was it free text and then interpreted how? where is that data? For example, use of an open fracture for Indicator 3? open fracture of what? all open fractures? Where is the data per region, per hospital?

RESPONSE:

We thank the reviewer for the comment regarding the use of the indicator 2 and 3 in our study. The indicator 2 of LCoGS, workforce density, is defined as the number of specialist surgical, anesthetic, and obstetric physicians who are working, per 100 000 population. This definition neither excludes sub-specialties nor restricts only providers that only perform bellwether procedures. Likewise, indicator 3, surgical volume, is not restrictive to bellwether procedures. Therefore, we believe our

manuscript is accurate in describing the surgical system in Somaliland according to the LCoGS indicators 2 and 3. We also thank the reviewer for the clarifying questions regarding data collection. Information on data collection was already cited from our previous studies. However, we decided to address the reviewer's questions by adding specific information regarding the surgical logbooks and clarifying that we included all types of open fracture procedures in the data collection.

In the text: In the **METHODS** section, under ***Indicator assessment: Indicator 3*** paragraph, we added specific information regarding the surgical logbooks and types of open fractures included in the data collection.

4. Indicator 4 is omitted because of the difficulty in obtaining mortality data? Is there any mortality data that can be projected or modeled to fit your Ind 1 - 6 assessment for the country or a region? This may help extrapolate out a relationship of the indicators, even if a modeled mortality indicator is used. This would strengthen the paper because the use of the indicators separately has been described in the literature quite extensively without real known use or impact if not interdependently assessed (i.e., all 6 indicators is what a surgical system assessment includes based on all six system nodes, originally described by Blake et al.).

RESPONSE:

We agree with the reviewer that a full surgical system assessment would include all six system nodes. However, mortality data specifically for Somaliland is very hard to accurately obtain. Most national level data combines Somaliland with Somalia, which likely has very different mortality data and overall surgical health systems. Thus, we did not feel comfortable including this indicator in the paper.

5. For an example of an interdependent interpretation, I recommend Hanna et al. Use of the Six Core Surgical Indicators from the Lancet Commission on Global Surgery in Colombia: A Situational Analysis. Lancet Glob Health. 2020;8(5):e699-e710. This is the latest and most comprehensive example of "non-modeled" and interdependence and this work is not referenced in your paper (especially when you highlight the limitations of previous work. I suggest this for the purposes of interpreting your indicators in light of one another (for example, what associations exist between the indicators and regions?) For example, why does Sahil have the lowest SAO density, yet it has the highest delivery based on your indicator definitions? How does the impact indicators translate to an average income within the geographic areas you highlight so the indicators 5 and 6 have local context as a way you send a message for first interventions to country officials? A second example is in Sahil, again with the highest delivery, but with the best impact? How do you interpret this? Consider your use of GDP (national average) and this as a significant limitation to interpretation in the poorest regions that are several standard deviations away from this average. Or is it? Can you translate the GDP to a local context, along with the indicators in that local context for more relevance to your assessment?

RESPONSE:

We thank the reviewer for the excellent suggestion of a current example of interpretation of the interdependence of the LCoGS indicators in a nation-wide assessment. We revised the suggested article and used it as a guide to enhance the discussion. We also cited the article in the limitations. The reviewer's questions regarding the specific case of Sahil were also addressed and embedded in the discussion.

In the text: We addressed the reviewers' suggestions and questions in the **DISCUSSION** section (paragraph 3 and 2) and the **LIMITATIONS** (paragraph 1).

6. For indicator 5, you mention an average cost of healthcare as a proxy cost? Is this the average cost of medical or surgical admissions? Is this the average cost of a particular bundle of surgical interventions?

RESPONSE:

We thank the reviewer for this clarifying question. It is the cost per surgical procedure based on a previous study. We mention this in the Methods section but can clarify more if unclear.

8. Overall - I worry a bit about the use of the word "deficit" in the light of a HIC senior author assessing an LIC system. Is there a more culturally sensitive/competent term?

RESPONSE:

We thank the reviewer's astute comments regarding the use of a more culturally sensitive term to address a low-income country's surgical system. We changed the word 'deficit' or 'deficient' by 'limitation', 'limited' or 'unmet goal' throughout the manuscript. **Figure 1** was also updated to reflect these changes.

In the text: The words' deficit' or 'deficient' were replaced throughout the manuscript.

Reviewer #2:

The authors performed an analysis of surgical capacity at 16 hospitals in Somaliland using six indicators defined by the Lancet Commission on Global Surgery. These indicators are designed to assess preparedness, delivery and impact of surgical care. The study is clear, well written and has a number of strengths. Firstly, the paper fulfils an unmet need for data, as there is insufficient data on surgical systems worldwide. Secondly, the paper in comprehensive in that it assesses both the public and private system and includes all surgical facilities in the country. This is particularly important given the high proportion of surgical care provided in the private sector. Thirdly the discussion of the paper is nuanced and analytical, stressing the importance of surgery as part of a national health system and UHC as opposed to a "surgery first" approach. The authors justifiably question the validity of the very indicators they used, to measure what they are purported to measure: preparedness, delivery and impact.

1. There are a few minor suggestions which could improve the paper. The paper states as one of its strengths its methodology for calculation of financial risk protection stating it is not modelled data. However, the explanation of the methods for this may be a little unclear as given the cost and income data is not from the primary source (patients) and not from data they collected their estimates would also be "modelled". From the methods it seems that the authors used estimates from another paper, which in turn used estimates from another paper from 2010 which estimated cost of C-section, for their estimation of financial costs. They then applied GDP per capita as a substitute for income with no accounting for income distribution in the country, something most of the "modelling papers" do account for as income inequality is often quite significant in this context. There are differing regional estimates of financial risk protection which represent the different proportion of OOP expenditure reported by each hospital, as the cost of surgery and poverty line/GDP per capita are based on national estimates. It is not feasible for the authors change their methodology at this stage but instead I might suggest that they avoid highlighting it as a core strength of the paper and do not state its superiority to "modelled" data. The gold standard data source for FRP would be from direct patient reports of costs and income which is then used to make national estimates.

RESPONSE:

We agree with the reviewer's astute comments about not including the calculation of financial costs as a strength of this manuscript. Therefore, we deleted our attempt to collect data on financial cost from the core strengths at the beginning of our manuscript.

In the text: We updated the **Strengths and limitations of the study** section to reflect the deletion of the methodology for collecting financial cost data as a core strength in this study.

2. Given the general readership of the BMJ Open, it may help to include in the introduction some supporting literature regarding the cost effectiveness of surgery and the role of surgery within a health system and UHC to underscore why this is an important area of research

RESPONSE:

We agree with the reviewer's suggestion about including supportive information on cost-effectiveness of surgery as this would help to underscore why surgery is an important area of research. Therefore, we added the mentioned supporting literature in the introduction.

In the text: We updated the **INTRODUCTION** section (first paragraph) to reflect the incorporation of supporting literature regarding the cost-effectiveness of surgery.

3. Given the general readership of BMJ Open it may be helpful to the reader to include a little more context on the endorsement of these indicators for example their inclusion (4/6) in the WHO core 100 and the World Development Indicators as these indicators have a clearer mandate for collection and reporting than that of the Lancet.

RESPONSE:

We agree with the reviewer's wise comment since adding information on how the LCoGS indicators have been endorsed would help the readership to have a better context of the global importance of these indicators. Therefore, we have addressed this suggestion by adding more information regarding the inclusion of 4 out of 6 LCoGS indicators in the WHO core 100 indicators and World Development indicators.

In the text: We updated the **INTRODUCTION** section (first paragraph) to reflect the incorporation of information on the endorsement of the LCoGS indicators in the WHO core 100 indicators and World Development indicators.

4. Figure 2, although it makes clear the geographical area within the two hour catchment it does not show the population distribution. This might add more insight to the reader to get an understanding of the density of the population in the served and underserved areas.

RESPONSE:

We agree with the reviewer's astute comments about including information on population distribution since it would provide more insight to the reader regarding the population in the served and underserved areas.

In the text: We updated the **FIGURE 2** section to reflect the incorporation of the population distribution and density at the regional level and alongside the service area. We also included these changes in the **METHODS** section (**under *Indicator assessment: Indicator 1***) and add a brief mention of the relation of population density and hospital density in the **DISCUSSION** section (first paragraph).

5. Some data elements from this paper have been previously published by this group. This may need to be more clearly acknowledged.

RESPONSE:

We agree with the reviewer's suggestion. Therefore, we included more information regarding our previous studies to more clearly acknowledge that some data elements have been previously published.

In the text: We have updated the **METHODS (under data collection)** to clearly state some data elements have been previously published by our team.

VERSION 2 – REVIEW

REVIEWER	Isabelle Citron Royal London Hospital, Barts Health NHS trust
REVIEW RETURNED	18-Oct-2020

GENERAL COMMENTS	Many thanks to the authors for systematically taking on board my suggestions from the previous review. In particular the revised Figure 2 adds significant depth. I have no further comments.