Dear Editor and Reviewers,

Thank you for a very thorough review of our manuscript, your comments and suggestions have been very helpful in improving the manuscript. We have addressed all of the comments and provided a detailed response below.

On behalf of all the authors,

Martin Gerdin Wärnberg, MD, PhD Department of Global Public Health Karolinska Institutet Stockholm, Sweden

## Response to Editor’s Comments:

1. Please revise the title of your manuscript to include the setting. This is the preferred format of the journal.

We have revised the title to include the setting. The new title is “Feasibility of a Cluster Randomised Trial on the Effect of Trauma Life Support Training: A Pilot Study in India”.

1. Please revise the ‘Strengths and limitations of this study’ section of your manuscript (after the abstract). This section should contain up to five short bullet points, no longer than one sentence each, that relate specifically to the methods. The novelty, aims, results or expected impact of the study should not be summarised here.

The strengths and limitations section now focuses on the methods, and have been revised in the light of the comments from the reviewers.

1. Please ensure that you have fully discussed the methodological limitations of the study in the Discussion section of the main text.

We have revised the discussion to include the methodological limitations highlighted by the reviewers.

## Response to Reviewer 1, Dr. Feroze Sidhwa, San Joaquin General Hospital:

“We found that the ATLS ® and PTC arms had lower 30-day mortality compared to the PTC and standard care arms. This finding could hint towards a potential effect of training physicians in trauma life support, but it is important to note that this pilot study was not powered to detect any differences in outcomes. The arms differed considerably in sample size, with the ATLS ® arm having the smallest sample size. This difference most likely resulted from the randomisation process with a small number of heterogeneous clusters, and this heterogeneity highlights the importance of taking varying cluster sizes into account in the design of the full scale trial.”

If this is the case then these outcomes simply should not be reported. Since this is a feasibility trial the major concern should be assessing missing data, which it seems to be, but this statement will be controversial and may distract from the otherwise important and better-supported conclusions. I would suggest removing discussion and reporting of mortality, and instead simply report on whether or not mortality data was missing, and if so was it differentially missing in the different arms of the study. If the study is not powered to detect a difference and no measures of statistical significance were calculated then this is not useful information to the reader.

The above is the only concern I had when I chose “no” for “Are the discussion and conclusions justified by the results?”

Thank you for pointing this out and we agree that the reporting of the mortality outcomes is problematic for a feasibility trial. We have decided to report mortality outcomes descriptively in the paper, given comments from other reviewers, but revised the discussion to focus on the missing data and to avoid any conclusions about a potential effect of the training.

## Response to Reviewer 2, Dr. Wietske H.W. Ham, University of Applied Sciences:

Thank you for submitting this relevant paper and the impressive work behind it. It was an honor to review the paper. Here are some suggestions and comments, which I hope will be helpful in further developing the paper.

Thank you for your comments, they were very helpful. We have revised the manuscript in the light of your comments.

The aim of the study is to gain insight into the feasibility of a larger study. At the same time, more results are presented than necessary to answer this question. This is not clear to the reader.

We agree that the manuscript, and the supplementary material, includes more results than necessary to fulfil the aim of the study. As you have noticed, we struggled to find a balance between adhering to the protocol and reporting the results that were most relevant in a transparent manner. We have now attempted to address this by clearly stating that we deviated from the original protocol in reporting the pilot, to focus on the feasibility of the full-scale trial. We have therefore limited the number of results that we present in the main text of the manuscript considerably.

It is unclear to me whether patients who ultimately did not regain consciousness were also included in this study. This should be described more clearly.

We have clarified that patients who did not regain consciousness were included based on the consent by their representative.

It also seems that, given the research period, the study sites were included sequentially; if so, the order and period should be described.

That is correct, data collection started in each cluster as they received the necessary approvals. The information about the study period in each cluster has been added as supplementary material.

It is not sufficiently clear to me how the statistical methods used contribute to answering the research question, namely the feasibility of an RCT and determining an effect size (and other measures, which are too vague).

We have clarified how we used descriptive statistics to summarize patient characteristics and feasibility outcomes. We now describe the methods used for the additional analyses according to the original protocol in the supplementary materials.

It seems that in the description of this manuscript, the pilot outcomes are mixed with the outcomes of the trial itself, which is confusing. This is likely due to the way it is written (it needs to be better explained and specified): the authors describe results, but also the outcomes. It is unclear in the current writing how the outcomes contribute to answering the research question.

We agree that this was confusing and have therefore attempted to streamline the paper further by presenting mortality descriptively along with the other patient characteristics, and only report the feasibility outcomes in the outcomes section.

The flow diagram does not clearly show how many participants did not participate or were missed. The design of the study itself seems to have been conducted properly.

Thank you for pointing this out. We have added the number of patients who were lost to follow up in each arm to the flow diagram.

In the discussion, deviations from the protocol are mentioned and referred to in the appendix, while these data are actually appropriate to discuss (and share) in the article itself, since it is a pilot.

We have revised the manuscript to describe the deviations where they are most relevant, and retained a list in the supplementary materials for completeness.

I find the number of appendices in the appendix very extensive, which makes it lose its purpose, and I also wonder to what extent it contributes to answering the research question. Parts of the protocol, which are now included in the appendix, should, in my opinion, be incorporated into the main paper.

We agree that the supplementary materials are very extensive. We included these to transparently report the results of the analyses specified in the original protocol, but in response to your comment we have now published the majority of the supplementary materials as online additional files, and cite these in the manuscript. We have attempted to include parts of the supplementary materials in the main text where appropriate, in response to your earlier comments.

## Response to Reviewer 3, Dr. Sheila Sprague, McMaster University:

Thank you for the opportunity to review the manuscript titled “Feasibility of a Cluster Randomised Trial on the Effect of Trauma Life Support Training: A Pilot Study”. This manuscript describes an important trial with the potential to save many lives at hospitals across India. There are many sections in this manuscript which are unclear. The overall structure of the manuscript also needs to be improved upon. Detailed comments are below.

Thank you for your thorough review. We have made major revisions in response to your comments, and we hope that the manuscript is now clearer.

Introduction - Provide clear objectives for the both the pilot phase and the definitive trial phase

We have revised the objective of the pilot study to also include the aim of the full scale trial. The revised objective is “to assess the feasibility of a cluster randomised controlled trial comparing the effect of ATLS® and PTC with standard care on outcomes in adult trauma patients.”.

Methods - Describe what information is being collected on standard of care at each hospital as it varies.

We did not collect in-depth data on the standard care at each hospital, and we have added a note in the methods section to clarify this. We have also added this as a limitation of this pilot study, and suggested that the full scale trial should include a more detailed assessment of the standard care at each hospital.

* Provide additional details on ATLS and PTC training interventions.

We have added more details on the ATLS and PTC training interventions to the methods section.

* Include details on the cluster selection.

We have revised the methods section to clarify that the seven hospitals were a convenience sample of hospitals in India that fulfilled the inclusion criteria, and with existing connections to the research team.

* The authors should comment on the sample size for a pilot study. For example, why did they select seven clusters? They should also comment on the number of residents and patients required to allow for a meaningful analysis of the feasibility outcomes.

We have added the rationale for the number of clusters and the number of units per cluster to the methods section. We did not conduct a formal power calculation, but deemed that the anticipated sample sizes would be sufficient to evaluate the feasibility outcomes.

* The reason for the one-month lead in is unclear. It is unclear how this data will be analyzed.

We included the one month lead to be able to evaluate the feasibility of comparing patient outcomes both as absolute differences between the intervention phases, and as differences in change from baseline.

* Using sealed envelopes for randomization increases the risk of selection bias, tampering, and accidental loss, potentially compromising allocation concealment. Authors should have used the computer-generated randomization system at time of randomization for each site.

We recognise that the use of sealed envelopes for randomisation increases the risk of bias and errors, and we have added this as a limitation of this pilot study.

* For blinding, the authors should state that it was not possible to blind those involved. It is unclear if those analyzing the data were blinded, as this is possible.

We have added that we did not blind the data analysts in this pilot study, but recommend doing so in the full scale trial.

* The follow-up period for each participant is unclear. The methods state that data were collected for four months. The methods also mention a 3-day mortality outcome.

The data collection period in each hospital was intended to be four months, and then patients were followed up for one month. However, the actual data collected periods varied slightly between clusters, and we have clarified this in the methods section.

* Success criteria and threshold details for the feasibility outcomes should be clearly outlined in the methods.

We did not specify success criteria for the feasibility outcomes, and we have added this to the methods section.

* The outcomes for the definitive trial (clinical outcomes) need to be clearly listed and described.

We now specify that we collected data on potential outcomes for the full scale trial, including mortality within 30 days, in-hospital mortality, complications and health related quality of life.

* The statistical analysis section is unclear. It should list the descriptive analyses plans and then the analyses used to assess the feasibility objectives. Feasibility pilot studies are not powered to detect meaningful differences between groups. Rather their goal is to determine whether a definitive trial is practical by assessing feasibility metrics. The clinical outcomes (the outcomes for the definitive trial) could be reported descriptively, but the use of comparative statistics (differences in final values, differences from baseline, confidence intervals, and subgroup analyses) to assess these outcomes is not appropriate during the pilot phase and potentially misleading, especially without a formal sample size calculation being conducted.

Thank you for pointing this out. We acknowledge that the estimation of potential effect sizes is not appropriate for this pilot study, and we have revised the aim of the study to reflect this and report this change as a deviation from the original protocol. We now report that we used descriptive statistics to analyse the feasibility outcomes and report clinical outcomes descriptively.

* The statistical analyses section also describes inconsistent times for outcome comparisons. The intervention arms measure change from baseline over one month pre-training vs. three months post-training, while the control arm measures change over a four-month period. For the definitive trial, the timing of assessment should be consistent across all arms to avoid biased comparisons.

We agree.

Results - Add a description of each cluster.

We have added a description of each cluster as Supplementary Table S1.

* Comment on why the number of residents and patients varied across the treatment groups.

We have clarified that the number of residents trained in each intervention cluster depended on the size of the units, and that the number of patients included per cluster varied between clusters depending on the volume of patients at each hospital.

* A few key prognostic factors differ across the treatment groups (age, ISS).

Thank you for pointing this out and we now highlight this in the results section.

* Consider reporting overall enrollment rates (e.g. number of participants enrolled per month). There is a paragraph in the discussion on low enrollment, but it is not listed in the results.

We have added a figure to the result’s section, reporting the number of patients included per month per cluster. Notably, the process of preparing this figure allowed us to identify a data collection error in two clusters, and we have added this to the methods section. We have also included this in the discussion as a limitation of this pilot study and an important lesson for the full scale trial.

* The results of feasibility outcomes should be reported clearly in a table, with a column indicating whether the feasibility outcome was successfully achieved.

We have decided not to report the feasibility outcomes in a table, because some of them, for example the comparison of directly observed data and data extracted from medical records, were assessed qualitatively. We have however removed the description of the clinical outcomes from the outcomes section to avoid confusion.

* Mortality should only be reported once in the results section.

Mortality is now reported together with the other patient characteristics in the results section. We have removed the table comparing clinical outcomes from the results section.

* Report the resident outcomes and the participant outcomes separately.

We now report the resident outcomes and the participant outcomes separately.

* List the protocol deviations in the results section, not in the discussion.

We now report the protocol deviations where relevant in the methods section, and include a list as supplementary material for completeness and easy reference.

* It is unclear the importance and relevance of the before training after training data (Table S1).

We included the one month before training and three month after training data to be able to evaluate the feasibility of comparing patient outcomes both as absolute differences between the intervention phases, and as differences in change from baseline. We have attempted to clarify this in the methods section.

* The EQ-5D should be scored as opposed to showing the results for each question.

Currently, there is no Indian value set available for the EQ-5D-3L, and we have therefore not scored the data. We now explain this in the methods section.

* Simplify the outcomes tables to focus on the key findings. Remove the yes/no and only list the yes.

We have removed the majority of the outcome tables from the supplementary materials and published them as online additional files, referenced in the manuscript.

* The absolute difference versus relative difference does not add to the analyses and could be removed.

We have removed the absolute difference versus relative difference from the main results section, but report them in the additional online files as this analysis was part of the original protocol.

* As this a pilot study, outcome data should be reported descriptively in a clearer manner.

We have attempted to clarify the reporting of the outcome data in the results section.

* There are too many outcomes.

We agree and suggest that the full scale trial should focus on the most relevant outcomes.

Discussion - Remove all results from the discussion and place them in the results section, then discuss the results in the discussion section.

We now only discuss results already reported in the results section in the discussion.

* Discuss the implication of adding the seventh site in more detail.

We have clarified that the seventh site was included for pragmatic reasons.

* Discuss if there are methods to improve participant follow-up for the definitive trial.

We now highlight the challenges of following up patients after discharge in the discussion and suggest that the full scale trial should focus on in-hospital mortality as the primary outcome.

* Discuss if there are methods to reduce missing data for the definitive phase.

We have added a suggestion to the discussion to improve the completeness of the data extracted from the medical records.

* Comment on the ethics of the full trial. Is it ethical to have sites continue with usual care after seeing the pilot study results?

We have added a sentence on the ethics of the full trial.

* Consider a stepped wedge design where all clusters start with standard of care and then transition into an intervention.

Thank you for this suggestion and we agree that a stepped wedge design would be a good approach for the full scale trial. We now include this as a suggestion for the full scale trial in the discussion.

* Use the CONSORT statement extension for pilot and feasibility trials (<https://www.equator-network.org/reporting-guidelines/consort-2010-statement-extension-to-randomised-pilot-and-feasibility-trials/>) and the CONSORT statement extension for cluster randomised trials (<https://www.equator-network.org/reporting-guidelines/consort-cluster/>).

We have attempted to align our reporting with these guidelines.

## Response to Reviewer 4, Dr. Kyle Dack, University of Bristol:

This is a well written paper, with transparent and thorough methodological reporting, and a reliable statistical approach. The purpose of the analysis was to inform a decision on the feasability of a full trial, and as such the results are interpreted with appropriate caution.

I have only minor comments about certain study design aspects which I found unclear - particularly the observed/extracted data comparison.

Thank you for reviewing the manuscript!

Methods: 1. Intervention – how long was the training programme, and who was it delivered by? Was the training performed in stages, e.g. to different cohorts of residents within the hospital, or all at once? The first paragraph stated that there is a 1 month observation phase followed by a 3 month intervention phase, but its not clear to me if this means the intervention continued over 3 months, or if the intervention was applied at the beginning of this phase, followed by ~3 months of patient follow-up. Further intervention in the study design and intervention sections could clarify this.

We have clarified that the training was conducted at the beginning of the three month intervention phase.

1. Variables – more information on the injury severity score would be useful to readers, because I am unsure how this variable was planned to be used (it is not an outcome), and how precisely it was calculated. A reference is provided, but a high-level overview of the approach (including the range of potential scores) would help readers understand this measurement better.

We have added a description of the ISS to the methods section.

1. “Differences in distributions of observed and extracted data” – I am unsure how observed data differs from extracted data – these terms were not previously explained or defined. Does it mean the observation period vs the 3 month intervention period? “distributions” is also somewhat unclear, histogram distributions, or a specific tendency statistic?

We have attempted to clarify what we mean by distributions (summary statistics) and directly observed data, in the section where we describe the feasibility outcomes.

1. Statistical analysis: “We used an empty generalised linear mixed model to estimate the intracluster correlation coefficient” which outcomes does this apply to?

We have moved the description of this analysis to the supplementary materials, and clarified that it applies to all cause 30-day and in-hospital mortality.

1. The analysis involved comparing the 1st month and 2nd-4th months, but it is not described how outcomes will be summed – e.g. consent rate, I imagine this the total consent rate in month 1 vs months 2-4? It would be helpful to directly confirm this in the text.

We have attempted to clarify that feasibility outcomes were not assessed as differences between the observation and intervention phases, but summarised across the entire period.

Results 6. Table 1. Missing counts are reported for some but not all measurements. Does the absence of a count indicate complete data? If so, a footnote stating this would help readers.

We have added a footnote to Table 1 to clarify that the absence of a count indicates complete data.

1. Table 2. As mentioned in a prior comment it is unclear to me how these two groups are defined. Also, both have an N of 55, assuming these are patients it means there is a total N of 110, less than the total of 376. Why do the numbers differ?

We only extracted data from a convenience sample of patients to reduce the workload of the project officers collecting data. We have attempted to clarify this in the methods section and in the table caption.

1. Table S1. Suggest clarifying that this is patient characteristics in the title.

We have added “patient” to the title.

1. Table S5. Certain cells are blank – a footnote should explain what this means (no observations?)

Yes, that is correct. Empty cells indicate that there were no observations. We have now moved the majority of the tables from the supplementary materials and instead published them as online additional files, referenced in the manuscript. These files include a note to clarify that empty cells indicate that there were no observations.