Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-2025-118

Reviewer Comments

In this brief report, the authors perform a cognitive task analysis to develop a consensus for steps of a robotic right lower lobectomy, and then determine level of importance. Seven 'expert' surgeons were interviewed for this study. 32 unique steps were developed, 87% determined to be important by at least 5 of the 7 surgeons.

Comment 1: Thank you for including what defined an expert in this study. It would be interesting if you could include how many surgeons nationally have performed >200 robotic lobectomies (Intuitive may have this data) to further confirm that it is an appropriate cutoff. Further, how did you decide on these specific 7 surgeons? Was a national request made to participate if criteria were met, or did the authors just reach out to colleagues/friends who they knew had experience? (Which may lead to inherent bias, as close colleagues may perform the procedure the same way.)

Reply 1: We thank the reviewer for their insightful review of our article. While we do not have access to that specific data, there have been several studies that have shown proficiency to be achieved after 50-60 cases, for which the surgeons were far beyond this cutoff. We have included two useful citations elucidating this (https://pubmed.ncbi.nlm.nih.gov/31175871/, https://pmc.ncbi.nlm.nih.gov/articles/PMC8575838/). Furthermore, these were expert robotic surgeons who agreed to participate and volunteer their time with the understanding that this map will contribute to future machine learning in robotic lobectomy and was not a national request. However, like any expert consensus, a national request is not always made. We have expanded our limitation sections to address these points.

Changes in the text: We have added the following in our methods (page 4, lines 78-79), see underlined: Seven board-certified expert thoracic robotic surgeons (> 200 robotic lobectomies, <u>which is significantly greater than the 50-60+ cases required to achieve profiency[9,10]</u>) from four academic institutions were recruited. We have also added the following in our limitations section (page 7, lines 146-147), see underlined: Second, our convenience sample was based on high-volume robotic surgeons from academic institutions <u>who agreed to participate and volunteer their time with the understanding that this map will contribute to future machine learning in robotic lobectomy instead of a national request</u>, which may not be generalizable to other practice settings.

Comment 2: Based on responses to Phase E, it seems like there were at least 2 surgeons who do not 'fissure dive,' and likely perform their RLL from an inferior to superior approach. I understand that there was no formal discussion between participants, but this is something that is important to elucidate, and the technical map for a fissure first approach vs bottom-up approach is quite different, and will make results to this kind of survey very confusing.

Reply 2: We thank the reviewer for this important point. We have adapted our map to account for this fissure-last approach. Please see the new map for the changes.

Changes in text: We have adapted our figure.

Comment 3: I'm unclear how you developed the procedural map. Several steps were included that did not reach any sort of consensus. I understand that you included the red star, but isn't that a bit disingenuous? Test bronchial stump under water only had 1 surgeon mention it in interview, and only 1 say it's important (likely same person), but somehow it made it onto the map.

Reply 3: We appreciate the reviewer's concerns regarding the inclusion of certain steps that did not achieve consensus. The rationale for this is that the goal of the procedural map was to serve as a structured annotation tool for video-based machine learning models. Including fewer common steps ensures that the algorithms are exposed to variations in surgical practice to better improve generalizability. We tried to balance this competing demand with our consensus map by visually distinguishing steps that did not meet consensus with an asterisk to allowed users of the map to differentiate between widely accepted and less common techniques. This project also shows the real-world surgical variability that is seen even among high-volume expert surgeons, and we hope is useful.

Changes in text: We added the following in our discussion (page 6, lines 136-138), see underlined: However, there were several operative steps that did not reach consensus <u>that were included in the map (visually distinguished with red asterisk) to show the real-world surgical variability that is seen even among high-volume expert surgeons and to serve as a structured annotation tool for video-based machine learning models</u>.