

Peer review file

Article information: <http://dx.doi.org/10.21037/jtd-20-2251>

Reviewer A

Comment 1. There is no standardization of techniques, therefore the aggregate comparison might not be appropriate

Reply 1: Thank you for these comments. No standardization of techniques could be both a limitation of our study and advantage. Like in reality, preferred techniques differ from surgeon to surgeon. We also intended to compare VATS vs TORA generally. Different techniques could be used in both groups (i.e. TORA and VATS). Even one of the reviewers suggested, that lack of standardization in limitations of our study could be considered as advantage “Though inclusion of many surgeons [thus techniques – author note] may be perceived as a limitation (lines 328-330), it also improves the generalizability of your findings.” On one hand we could have an analysis of one standardized technique (homogenous population, but in our opinion artificial setting) and on the other hand we could have a general analysis (heterogenous, but in our opinion more practical, more real). However, at general every institution recommends Major Lymph Node Dissection, and removal at least 6 LNs (3 N1 and 3 N2 – accordingly to ESTS). And as lymphadenectomy technique impacts our study results directly, we could say that our study was relatively standardized in that field.

Changes in the text: none

Comment 2. There are already several meta-analysis and systematic review on the topic, so I did not understand what your article add to the international community, perhaps it would be better published in a national polish surgical journal to inform all the local thoracic surgical group.

Reply 2: Despite many articles and even meta-analysis on this topic, proper lymphadenectomy is still an unsolved problem in surgery of NSCLC. There are no unanimous recommendations regarding minimal extent of lymphadenectomy. Please, look at these 2 American studies:

- Osarogiagbon RU, Allen JW, Farooq A, Wu JT. Objective review of mediastinal lymph node examination in a lung cancer resection cohort. J Thorac Oncol.

2012;7(2):390-396. doi:10.1097/JTO.0b013e31823e5e2d

- Osarogiagbon RU, Yu X. Nonexamination of lymph nodes and survival after resection of non-small cell lung cancer. *Ann Thorac Surg.* 2013;96(4):1178-1189. doi:10.1016/j.athoracsur.2013.05.021

Issue seems to be solved, but at the same time, American Surgeons with top-shelf healthcare and top-shelf methods and top-shelf tools seem to struggle in such important topic (i.e. Forty-five percent of resections were reported by surgeons as MLND, 8% RS, and 48% NS. None met pathology criteria for MLND; Thirteen percent of all resections (18% of node negative resections) were pNX). The quality of lymphadenectomy is important issue worldwide. As it is not solved, we need more and more data, and more and more studies to finally solve it.

Also, fact that there are a lot of studies regarding this topic, does not necessarily mean that there is nothing else to study in this field. This topic is important and thus there are many studies regarding it. As it is unsolved, there will be many more. We provided extra text in the introduction to explain our rationale for the study.

Changes in the text: New paragraph in Introduction section with our rationale for the study: Page 4, Lines 77-83. Reference list was updated.

Comment 3. An important part in counting lymph nodes is that, often, especially during vats surgery, they are removed in pieces rather than as a whole or simply sampled. Can you explain whether lymph nodes were sampled in all institutions or how they were counted

Reply 3: Major lymph node dissection (MLND) was used almost everywhere to our knowledge. Every thoracic center uses this technique. We have no knowledge of programmes or trials to use selective sampling or lobe specific sampling in other institutions. Pathological reports were used as data in our study. Also, every Institution in Poland recommends use of ESTS recommendations of removal at least 6 LNs (3 of N1 and 3 of N2). Lymph node fragmentation is another example of unsolved problems

in thoracic surgery of lung cancer.

Changes in the text: none

Comment 4. As pet was not done in all patients how can you be sure that the staging was accurate?

Reply 4: Preoperative staging of NSCLC is another unsolved problem. cN->pN upstaging happens relatively often in lung cancer. PET seems to lower that risk notably. But the upstaging problem still exists. Also, it is expensive diagnostic tool and not easily accessible in many countries. In Poland as you can see it was rarely accessible at beginning and partially accessible in later time. In countries with poorer healthcare systems this rate may be even worse. This is just the reality we and some others live in. The CT is the most important and established diagnostic tool. Even in American studies based on SEER database, not every patient had undergone PET in their populations.

Changes in the text: none

Comment 5. Why did you include the learning curve of VATS in Poland? I understand the desire to increase the number of cases, but I think you should compared an established technique with another.

Reply 5: The learning curve was not a main aspect of our study. It did not affect distant results significantly, therefore in our opinion it was correct to include them in analysis. The learning curve only affected the early complications. That's why we decided to present the complications in subgroups. One could wonder why our complications were so high, so we decided to clarify it with learning curve subgroups.

Changes in the text: New sentence in Results – Complications section: “Also, early and late periods did not differ in long term results significantly. Thus, we decided to analyze long term results of these subgroups collectively.” Page 10, Lines 210-212.

Reviewer B

Comment 1. Complications should be defined in the Materials and methods section. For example, how long did “prolonged” of prolonged intubation and prolonged air leakage mean? How about “bleeding”? It is so difficult to evaluate the frequency of the complications in this cohort.

Reply 1: Thank you for your accurate suggestion. Indeed, we should define what exactly those complications mean. Therefore we explained each complication's definition in the materials and methods section.

Changes in the text: New sentence in material and methods section: Surgery complications included in analysis were as follows: hemorrhage (>800ml), prolonged air leakage (5-7 days), atrial fibrillation, pneumonia, prolonged intubation (>7 days), bronchial fistula. Page 7, Lines 135-137

Comment 2. Considering the study period, the percentage of the VATS-L group is so small in comparison with the TORA-L group. Any comments? Are there any tendencies that more recent patients underwent more VATS-L? Or large centers performed more VATS lobectomies in this study period as the authors described in the limitation section of their discussion?

Reply 2: Thank you for your comment. Our data material also covers the period when VATS lobectomy was initiated in Poland. Thus, as with any new surgical technique, in the beginning period VATS-L were performed only in large centers. With all due respect, but in many highly cited studies when comparing new technique vs regular one, the number of "new cases" is usually much lower than in our study. In other words, although there is a disproportion and percentage of the VATS-L group may seem low, we think that the overall number of VATS-L cases is sufficient enough for data usage in this paperwork. As for now, VATS lobectomies are increasingly done in many surgical centers in all over the Poland.

Changes in the text: none

Comment 3. Please provide more convincing discussion about the results of lobectomy-related survival. If the authors want to explain the rate of stage 1A1 as they did in their discussion, please provide the data by the sub-group analysis.

Reply 3: Thank you for your suggestion. We extended our discussion about the results of lobectomy-related survival, according to your request. Due to organization, administrative etc. reasons we are unable to provide additional statistical analyses or modifications.

Changes in the text: New sentences thorough Discussion (Survival part) Page 14 Lines 309-314

Comment 4. According to Table 5 and Figure 2, p value was 0.002, but the text showed $p < 0.002$ in the abstract (Page 2, Line 36) and in the results (Page 8, Line 225).

Reply 4: Thank you for noticing our miscalculation mistake. We changed the “ $p < 0.002$ ” in the abstract and results for the “ $p = 0.002$ ” as it should be.

Changes in the text: Changed accordingly to reviewer suggestions.

Comment 5. In the abstract, is “Multifactorial analysis” correct? (Page 2, Line 38)

Reply 5. Indeed, as you correctly noticed this expression is misleading, so we changed that for “multivariable analysis”.

Changes in the text: Changed accordingly to reviewer suggestions.

Comment 6. In the Materials and methods section, “2014” should be “2015”? (Page 5, Line 109)

Reply 6: Yes, sorry for a typo- it should be “2015”, as it covers the data collection period. We changed it according to your suggestion.

Changes in the text: Changed accordingly to reviewers suggestions.

Comment 7. In the Materials and methods section, “polish” should be “Polish”. (Page 5, Line 121)

Reply 7: We changed small letter for a big one in the word “Polish” as you recommended.

Changes in the text: Changed accordingly to reviewer suggestions.

Comment 8. In the Materials and methods section and Table 3, “ $p = 0.000$ ” is correct? It would be better to write “ $p < 0.001$ ”. (Page 5, Line 179)

Reply: In accordance to your comment we changed the “ $p = 0.000$ ” for the “ $p < 0.001$ ” in materials and methods section and Table 3.

Changes in the text: Changed accordingly to reviewer suggestions (both text and Table 3).

Reviewer C

Comment 1. The work needs to be proofread for grammatical and syntax issues

Reply 1: We have proofread the manuscript again, but because the reviewer gave no examples, we are unclear about what changes are requested. Also, as only Reviewer C reported language issues (3 other reviewers did not report it), we decided not to seek for professional English Revision. We asked our colleague, who is better English speaker to proofread the article.

Changes in the text: Minor changes thorough the manuscript Page 12 Lines 255-266
Page 14 Lines 316-317. Also, missing commas thorough manuscript

Comment 2. Cancer specific survival as an outcome variable should be presented

Reply 2: We want to thank you for this suggestion and undoubtedly cancer specific survival seems to be more accurate in order to better describe our patients. However, our study is based on data from Polish Lung Cancer Study Group and unfortunately this rate cannot be incorporated in our paper because of incompleteness of data in a database. As we mentioned in Materials&Methods when patient is lost to f/u we can still assess the OS by asking government institution for dead/alive status in national database of ID numbers. We do not have ability to confirm the reason of death. Even if we had access to such information it could be impossible to still include more data of e.g. death reason because when patient is lost to f/u and does not visit the doctors, it is very likely that his death would be described as natural or due to generic reason (e.g. cardiac arrest, shock etc.) Nowadays, at thoracic centers there are hired assistants that input the data in the Database. Other doctors are asked to do the same on their behalf (e.g. outpatient oncologic care, or primary care), but they may not have the resources or funds to do so. So, unfortunately many data might be missing. Also, due to organization, administrative etc. reasons we are unable to provide additional statistical analyses or modifications.

Changes in the text: none

Comment 3. How many patients were lost to f/u? We need some measure of this or compliance with f/u

Reply 3: Thank you for another good suggestion. We omitted some data, so we added this information to clarify. We hope that new changes, which were made accordingly to your tentative, are clear now. Similar like in IASLC studies – we also excluded patients without full data needed for the study. As we mentioned above in Comment 2. lost to

f/u is a minor issue, as we have ability to assess the OS when lost to f/u occurs.

Changes in the text: “However, we excluded 167 patients from the study, because of incomplete data (mostly), loss of clinical information and insufficient period of follow-up. “ was added to Materials and Methods Section Page 6 Lines 121-123

Comment 4. The authors should stay away from global statements about the effectiveness of one procedure vs another and not hypothesize regarding reasons in concluding paragraphs. Just report the findings so as to not mislead the readers.

Reply 4: Indeed, we overestimated the conclusions here. We understand that some of the statements could be considered as our assumptions, but we tried to confirm every of this hypothesis in literature. Also, as reviewer did not provide any examples, we are unclear about what changes are requested beside the last part of conclusion paragraph. Please elaborate.

Changes in the text: We deleted the last part of conclusion as requested Page 16: Lines 348-349.

Comment 5. The authors need to comment on the inherent selection bias in the two groups. The improved outcome in the VATS group could easily explain this

Reply 5: It was mentioned briefly in discussion. We added extra sentence for clarity.

Changes in the text: “Surgeons may have been biased as e.g. more percentage of IA1 stage was present in VATS-L group on the left side.” Was added Page 15 Lines 328-329

Comment 6. We need some clarification regarding clinical and pathological stage throughout the manuscript

Reply 6: Our study is based on pathological-surgical database; thus, stages are pathological ones. Everywhere throughout the manuscript.

Changes in the text: none

Comment 7. The pathology issue the authors highlight in the limitations regarding number of nodes and fragments of nodes is a major issue. The manuscript misleads in terms of its vocabulary as it relates to this. The authors need to find some other nomenclature of the term "lymph node" throughout the body of work so as to not

mislead the audience as this is a major issue not just in Poland, but internationally.

Reply 7: As you said Lymph Node fragmentation is unsolved problem worldwide. We mentioned about it in limitations of our study. If reader analyses our paper thoroughly, he will acknowledge this issue. We are unable to distinguish Lymph Node vs Lymph Node fragment based on data from our database.

Changes in the text: none

Comment 8. How can the authors not cite Nwogu et al Ann Thor Surg 2015 Feb;99(2):399-405?

Reply 8: We totally agree that the paper of Nwogu et al. should be cited, because of its impact on treatment decisions and such a sizable population. We added this multi-institutional study to our references. Thank you for remind us about its results.

Changes in the text: Nwogu was cited in text of discussion Page 13, Line 296. Reference list was updated (ref no. 31).

Comment 9. The concluding sentence of the abstract (as with other statements) overstates the findings.

Reply 9: The second part of the last sentence can be overstated, but our aim was to encourage scientists to dedicate time and in our opinion this problem needs to be looked into further. We decided to delete this part of sentence.

Changes in the text: Last sentence of conclusion of abstract was modified. Page 2 Line45-46

Reviewer D

Introduction:

Comment 1. Multiple single institution (eg, Merrit, Ann Thorac Surg, 2013), US national cancer registry (eg, Medbery, J Thorac Oncol, 2016) and meta-analyses (Zhang, Ann Thorac Surg, 2016) have already examined lymph node yields of VATS vs open lobectomy. In addition, a highly cited meta-analysis (Whitson, Ann Thorac Surg, 2008) has also evaluated survival after VATS vs open lobectomy. Indicate what piece of information is missing from this literature and how you wish to add to it.

Reply: Despite many articles and even meta-analysis on this topic, proper

lymphadenectomy is still an unsolved problem in surgery of NSCLC. There are no unanimous recommendations regarding minimal extent of lymphadenectomy. Please, look at these 2 American studies:

- Osarogiagbon RU, Allen JW, Farooq A, Wu JT. Objective review of mediastinal lymph node examination in a lung cancer resection cohort. *J Thorac Oncol.* 2012;7(2):390-396. doi:10.1097/JTO.0b013e31823e5e2d
- Osarogiagbon RU, Yu X. Nonexamination of lymph nodes and survival after resection of non-small cell lung cancer. *Ann Thorac Surg.* 2013;96(4):1178-1189. doi:10.1016/j.athoracsur.2013.05.021

Issue seems to be solved, but at the same time, American Surgeons with top-shelf healthcare and top-shelf methods and top-shelf tools seem to struggle in such important topic (i.e. Forty-five percent of resections were reported by surgeons as MLND, 8% RS, and 48% NS. None met pathology criteria for MLND; Thirteen percent of all resections (18% of node negative resections) were pNX). The quality of lymphadenectomy is important issue worldwide. As it is not solved, we need more and more data, and more and more studies to finally solve it.

Also, fact that there are a lot of studies regarding this topic, does not necessarily mean that there is nothing else to study in this field. This topic is important and thus there are many studies regarding it. As it is unsolved, there will be many more.

Changes in the text: New paragraph in Introduction section with our rationale for the study: Page 4, Lines 77-83. Reference list was updated.

Comment 2. State your hypothesis

Reply: We preferred to ask general question e.g. “What is the impact” rather than state hypothesis. Please note that some reviewers commented that we are too eager with statements, conclusions and hypotheses. We decided not to change our general intention “to assess the number of lymph nodes removed and the overall survival (OS) during VATS and open thoracotomy (TORA) lobectomies depending on the location of lobectomy.

Changes in the text: None

Methods:

Comment 1. "VATS" means different things to different people. The proper definition involves nonrib spreading incisions. How did you define "VATS?"

Reply 1: Thank you for your comment. The full definition of a surgical technique is "video-assisted thoracic surgical non-rib spreading lobectomy (VNSSL)". We added this information to the operative technique's section of the manuscript.

Changes in the text: "The surgical technique of VATS was non-rib spreading incisions." Page 6, Line 130

Comment 2. If a VATS lobectomy was converted to open at some point in the operation, was it still labeled as "VATS."

Reply 2: Thank you kindly for your question. Indeed, we should have included this information in the text earlier. During the surgery, in a case of a technique switch from VATS lobectomy to open lobectomy (TORA-L) it was counted as open lobectomy. To be more precise we had 30% surgery conversions during the early period and 18% in the late period according to the learning curve. We also added this missing information to the manuscript in the complication's section.

Changes in the text: "During the surgery, in a case of a technique switch from VATS-L to TORA-L it was counted as open lobectomy. The conversion rate was: 30% during the early period and 18% in the late period." Was added Page 10 Lines 206-208

Comment 3. How did you take into consideration the learning curve for new VATS surgeons?

Reply 3: Sadly, we were unable to follow new surgeons into the analysis, due to lack of such data. It was more like "new VATS institution" rather than particular surgeon. Also, as teacher-surgeons and their institutions have more experience, the new resident-surgeons should have less complications because of proper supervision and adequate teaching of the method. So, we believe, that it had less impact on our study in comparison to major impact of the whole new institution learning the technique and care.

Also, if you are interested, please have a look at this study - the importance of

Institution-experience is shown.

32. Lee PC, Kamel M, Nasar A, et al. Lobectomy for Non-Small Cell Lung Cancer by Video-Assisted Thoracic Surgery: Effects of Cumulative Institutional Experience on Adequacy of Lymphadenectomy. *Ann Thorac Surg* 2016;101:1116-22.

Changes in the text: none

Comment 4. Technically, you performed multivariable, not multivariate (line 145), modeling. Describe which variables were evaluated for inclusion in this model and whether you used forward or backward selection.

Reply 4: Yes, thank you for correcting our mistake. We changed word “multivariate” for “multivariable” in the manuscript. Please have our statistician reply attached:

“The selection of predictive variables was performed on the base of the series of univariate Cox models for every variable. As significant we assume these variables for which p-value in corresponding univariate Cox model was below 0.05. These selected variables were included in the multivariate Cox model: age, sex, stage, cardiac infarction, illability of the circulatory system, kidney failure, CODP, coronary disease, pathological T descriptor, pathological M descriptor, the extent of pulmonary resection, surgical approach and pathological N stage.

We did not perform backward or forward selection. “

Changes in the text: We changed accordingly to reviewer suggestion (line 149, Page 7; Lines 223-224, Page 10)

Results:

Comment 1. Simple state the median month (and range) of follow-up. Line 151.

Reply 1: We provided it in months. (line 153, Page 7) Changes in the text: We provided it in months. (line 153, Page 7)

Comment 2. Any postoperative complication after lobectomy for stage I NSCLC leads to worse long term survival rates (Rueth, *Ann Surg*, 2011). How did you adjust for the impact of complications in your multivariable survival model?

Reply 2: The learning curve did not affect distant results significantly, therefore in our opinion it was correct to include them in analysis. The learning curve only affected the

early complications. That's why we decided to present the complications in subgroups. One could wonder why our complications were so high, so we decided to clarify it with learning curve subgroups.

Changes in the text: Lines 208-210, Page 10. New sentence was introduced "Also, early, and late periods did not differ in long term results significantly. Thus, we decided to analyze long term results of these subgroups collectively. "

Comment 3. Perform a 90-day landmark analysis to show consistency of your results on overall survival.

Reply 3: With all due respect but 30-day landmark analysis as shown in our tables is sufficient enough from a study perspective. Moreover, there are scientific papers indicate the superiority of 30-day landmark over the 90-day landmark from a surgeon perspective. Also, due to organization, administrative etc. reasons we are unable to provide additional statistical analyses or modifications.

Changes in the text: 30 day description of post/op mortality was added to Table 3

Comment 4. Tables 2 and 3 were a bit overloaded with information. Perhaps include them as supplementary tables

Reply 4: Thank you very much for this suggestion. Yes, we agree, that both tables contain too much information and many variables, but we are of the view that every information could be necessary to understand our results in the main text. Table 3 is indicated at least 5 times in main text, so we decided to include only table 2 as supplementary file as you recommended.

Changes in the text: We included table 2 as supplementary table and revised entire article to change numbers of other tables, eg. Table 3 is table 2 now etc.

Comment 5. Add p-values, number at risk and 95% confidence intervals to your Kaplan-Meier curves.

Reply 5: Due to organization, administrative etc. reasons we are unable to provide additional statistical analyses or modifications.

Changes in the text: none

Comment 6. For Table 4, add the time period for "early" and "late" years.

Reply 6: Thank you for your suggestion. We added the time periods into the table 4.

Changes in the text: Time periods were described in table 4

Comment 7. Table 6 is confusing. Does each line denote a separate model? If not, the reference groups are unclear. Tumor size or T stage should be added to this model.

Reply 7: Please have our statistician reply attached:

“Each line denotes a separate variable (or particular category for categorical variables) and corresponding hazard ratio in the multivariate Cox model.

This multivariate Cox model was performed with variables selected from univariate Cox analysis (response to the previous question).

For legibility of table, we presented HR only for selected variables. T stage was included in this model. “

Changes in the text: none

Discussion:

Comment 1. Though VATS had a lower lymph node yield, was it "good enough?" In other words, what is the added value of a median improvement in 2 more lymph nodes with a thoracotomy? A finding of a lymph node yield of nearly 11 is above the Commission of Cancer standard in the United States (10 nodes) and is above at least one threshold published in the literature (Ludwig, Chest, 2005).

Reply 1: In other studies ~2 less lymph nodes in VATS is similar to our study. As OS was better in VATS it is acceptable (VATS vs TORA point of view). Given the lymphadenectomy it is acceptable also (10 LNs is the most frequent value in studies from Ludwig to many other works). However, there are some studies that indicate need for higher number of LNs (e.g. 16 as also in Ludwig work, where he indicated that 16 is best value. But above 16 was not significant). So, we can say that from lymphadenectomy point of view this also is also acceptable, but it might have some impact on distant results. Also, as we (and other studies) have no ability to distinguish lymph node from lymph node fragment, we must acknowledge that in reality this number of LNs retrieved might be lower.

Changes in the text: According to many studies these are acceptable values for proper lymphadenectomy was added to discussion Line 254, Page 12

Comment 2. Though inclusion of many surgeons may be perceived as a limitation (lines 328-330), it also improves the generalizability of your findings.

Reply 2: Thank you for this comment. Yes, we agree that it could be not only a limitation, but also improves the generalizability of our findings. We decided to add one sentence, which is mentioned in your comment.

Changes in the text: “But it also improves the generalizability of our findings.” was added line 334, page 15.