Peer review file
Article information: http://dx.doi.org/10.21037/jtd-21-390

Reviewer A

Comment 1:
I have only minor suggestions:
In 16.7% systematic lymph node dissection was performed and in 16% clinical staging revealed N1 or N2 disease. Have you staged all patients with PET/CT or invasive mediastinal staging to minimise the risk of undetected N2 disease in patients who are planned for lobe specific lymph node dissection?

Reply 1:
We appreciate the reviewer's point. We performed PET/CT in all patients and endobronchial ultrasound-guided transbronchial needle aspiration or mediastinoscopy if necessary to minimise the risk of undetected N2 disease. We have added the above to the Imaging and Lymph Node Dissection of the Methods section (P4Ln19-).

Comment 2:
Was there an association between the adenocarcinoma subgroups and lymph node involvement?

Reply 2:
pN1 is 102 cases (8.4%) and pN2 is 120 cases (9.9%) in the adenocarcinoma subgroups (n=1216). An association between the adenocarcinoma subgroups and lymph node involvement was not strong.

Comment 3:
Did you analyze the pattern of local recurrence and distant recurrence, this may also reveal important further information on how to deal with the knowledge of lymph node involvement in lung cancer patients.

Reply 3:
We appreciate the reviewer's instructive suggestion. There were 326 recurrences, of which local recurrences were 168 cases and distant metastases were 145 cases (13 cases were unknown). Lymph node involvement included 85 local recurrences and 88 distant metastases. Our data did not reveal any features of lymph node involvement and recurrence pattern. We have added the above to the Results section (P5Ln22-).

Comment 4:
In order to classify the survival results more information on adjuvant or neo-adjuvant chemotherapy and radiation therapy would be important.

Reply 4:
We performed adjuvant chemotherapy in 500 patients (31.9%). We have added the above to the Results section (P5Ln21-). At that time, we did not perform neoadjuvant chemotherapy in clinical practice. PORT for pN2 was also not a standard treatment in Japan.

Reviewer B

Comment 1:
I absolutely agreed that the number of LN involvement really matters in the prognosis of NSCLC, but the IASLC already proposed a new N staging descriptor, reflecting the quantity of LN involvement in addition to the anatomical location. I do not quite understand why the authors would have to coin a new N descriptor, even though they could simply validate the IASLC-proposed N staging system using their database with a good number of patients. If they could understand me with regards to this point, I think that this manuscript deserves to be published by the JTD.

Reply 1:
We appreciate the reviewer's point. The IASLC-proposed N staging system is quite simple and reflects the prognosis well. We strongly agree with pN2, but our data showed that there was no prognostic difference in pN1 patients with 1, 2, and 3 mLNs. This may be because the hilar lymph nodes are densely packed in a small area. The IASLC data showed a significant difference in prognosis between 1 and 2 or more mLNs, but our data showed a prognostic difference between 3 and 4 or more, therefore it is worth reporting. As we wrote in the limitation, it may be because it is difficult to accurately determine the number of mLNs. We have added the above to the second paragraph of the Discussion section (P7Ln7-).

Reviewer C

Comment 1:
The term "complete resection" should be defined and, after that, to clarify if it exists some differences between institutions involved in this multicenter database.

Reply 1:
We appreciate the reviewer's point. Complete resection is defined as no residual tumor, either macroscopically or microscopically. The definition is the same at each institution. We have added the above to the Patients of the Methods section (P3Ln24-).

Comment 2:
Authors mentioned that they review 2662 patients who underwent R0 resection. The term "R0 resection" should be defined and clarify if these patients had R0 resection or also R0 resection + complete resection. I recommend the following paper to use for these definitions:

Reply 2:
We appreciate the reviewer's point, which we consider is also important. R0 resection includes uncertain resection, which means that a dissection of three mediastinal and hilar-intrapulmonary nodal stations, so that the final specimen includes at least six lymph nodes, are not met. In our study, 69 cases met the above conditions. We have added the above to the Patients of the Methods section (P3Ln25-).

Comment 3:
Authors excluded patients with "limited node dissection". This term should be defined and especially which is the minimum number of lymph nodes and nodal stations are accepted for this study.

Reply 3:
Limited node dissection means ND0-1 without mediastinal lymph node dissection. In our study, the minimum number of lymph nodal stations is 6. However, the minimum number of lymph nodes is

two, since the dissection area contains only fat and may not include lymph nodes. We have added the above to the Patients of the Methods section (P4Ln1-).

Comment 4:
Although the type of lymphadenectomy performed in this study is defined I recommend to add a reference to describe these techniques based on the international guidelines.

Reply 4:
We have added ESTS guidelines as reference 17 (P4Ln25-).

Comment 5:
Regarding overall survival, I understand that is not disease-specific survival rate, isn't it? This point it should be specified.

Reply 5:
We apologize for the confusion. Overall survival in our paper is not disease-specific survival. Overall survival was defined as the period beginning with surgery to the date of death regardless of disease or survival confirmation. We have added the above to the Statistical Analysis of the Methods section (P5Ln11-).

Comment 6:
Authors describe that multivariate analyses is performed but it is necessary to define the variables used in the univariate analysis.

Reply 6:
In the multivariable analyses, we included the covariate variables of age, sex, smoking history, histologic type, pathological T descriptor, and mLNs after checking for explanatory variables that are closely related to each other. We have added the above to the Statistical Analysis of the Methods section (P5Ln14-).

Comment 7:
In my opinion, to response the aim of this work "number of mLNs predict prognosis" is not necessary so many Kaplan Meier's curves. I recommend to use only those curves with maximal statistic differences.

Reply 7:
Since we think it is reasonable to point out that there are many figures of Kaplan-Meier curve, we deleted Supplemental Figure 1 and explained it only in the text. We have added the above to the Rusults section (P6Ln3-).

Comment 8:
Regarding those cases with pN1, the type of pN1 analysed is not described. Are they hilar, sublobar, lobar nodes...?

Reply 8:
Regarding pN1, lymph node metastasis to hilar (#10), interlobar (#11), lobar (#12), segmental (#13), and subsegmental (#14) was observed in 16 (9.2%), 44 (25.4%), 74 (42.8%), 26 (15.0%), 4 cases

(2.3%), respectively (5 cases were unknown). We have added the above to the Rusults section (P5Ln29-).

Comment 9:
Thanks to the authors for being so honest with the limitations of this study but, in my opinion, there are some limitations that can cause an important bias for this study especially in the type of lymphadenectomy performed (such as a low number of systematic nodal disection).

Reply 9:
We appreciate the reviewer's point, which we consider is important. In our study, the small number of systematic lymph node dissection may cause bias. We have added the above to the limitation of the Discussion section (P8Ln17-).

Comment 10:
The teaching of the authors based on the results: for example, if the number of mLNs counts for the prognosis, there is something in the clinical practice for changing? for example, increase the accuracy of presurgical methods to select the best candidates for surgery..

Reply 10:
We appreciate the reviewer's instructive suggestion. If the number of mLNs predicts prognosis, it is necessary to accurately diagnose the number of mLNs or to puncture multiple lymph nodes via EBUS or mediastinoscopy before surgery. This will help identify groups with poor prognosis and actively treat them. We have added the above to the conclusion of the discussion section (P9Ln6-).

Comment 11:
2) Taking into account the important limitations of this study (retrospective, database was not designed for this aim, diferent types of lymphadenectomies...) it should be necessary to add that it is necessary a well-designed controlled trial to certify the results reported by the authors.

Reply 11:
To clarify its prognostic impact, we need prospective data with a uniform patient background. We have added the above to the conclusion of the Discussion section (P9Ln8-).

Reviewer D

Comment 1:
Well written. Would recommend as you utilized p-value in first line of results, to add p-values when reported differences in outcomes further on. If word limit allows would recommend to add in HR and 95% CI.

Reply 1:
We appreciate the reviewer's point. We have added p-values to the abstract section (P2Ln12-).

Comment 2:
First statement should have a reference.
Line 63-65 You state multiple researchers but provide one reference only.

Reply 2:
We have added a reference to first statement in the introduction section (P3Ln3-) and 5 references to Line 63-65 (P3Ln12-).

Comment 3:
If not present in introduction, would be worth mentioning in discussion that societal guidelines are also not uniform in sampling requirements (i.E ACS recommends minimum of 10 LN this was based on SEER and NCDB data as they did not have stations, while NCCN recommendations to sample all stations.

Reply 3:
We appreciate the reviewer's instructive suggestion. As you said, there is no unified recommendation for lymph node sampling among the guidelines. We quoted the guidelines for ESTS and NCCN. We have added the above to the discussion section (P8Ln20-).

Comment 4:
Indent first paragraphs in multiple sections.

Reply 4:
We indented first paragraphs in all sections.

Comment 5:
Did any of the patients undergo pre-operative invasive LN staging with mediastinoscopy or EBUS when appropriate? This should be noted one way or another as potentially some of these patients would have undergone induction therapy, and if so it would be important to report neoadjuvant rates. It is standard if you suspect N1 or N2 to access ahead of time and likely given neoadjuvant.

Reply 5:
We appreciate the reviewer's point. We performed EBUS or mediastinoscopy if necessary to minimise the risk of undetected N2 disease. We did not perform neoadjuvant chemotherapy in clinical practice at that time. We have added the above to the Imaging and Lymph Node Dissection of the Methods section (P4Ln19-).

Comment 6:
Furthermore, no mention is made of additional variables collected which include baseline variables later referenced.

Reply 6:
We collected clinicopathological variables, which included age, sex, smoking history, carcinoembryonic antigen (CEA), clinical stage, surgical procedure, lymphadenectomy, resected lymph nodes, metastatic lymph nodes, histologic type, pathological stage, pleural invasion, pulmonary metastasis, lymphatic vessel invasion, blood vessel invasion, EGFR mutation, adjuvant therapy, and prognosis. We have added the above to the Patients of the Methods section (P4Ln7-).

Comment 7:
One presumes this was limited to lobectomy.

Reply 7:

We excluded patients who underwent wedge resection/segmentectomy. Surgical procedure in our study includes lobectomy (98.9%) and pneumonectomy (1.1%).

Comment 8:
Do the authors note whether procedure was open or VATS as this may impact survival to a degree?

Reply 8:
Unfortunately, this multicenter database does not collect information on surgical approaches. At that time, hybrid VATS or thoracotomy was mainly performed.

Comment 9:
Furthermore, there is no mention of what stage these patients were or T or M descriptor. A T1N1 and T4N1 will clearly have different survival. As I read the tables, the authors include these factors therefore this should be mentioned in the methods.

Reply 9:
Although M descriptor was all M0, various T descriptors were included, which could affect the prognosis. In other words, all cases were cT1a-4N0-2M0. We have added the above to the Patients of the Methods section (P4Ln11-).

Comment 10:
When analyzing Na-Nc you mention prognosis but as you allude to in discussion these were not uniform for survival and recurrence and this different should be noted rather than blanket prognosis. I had to go back to the figures multiple times to get at the conclusion.
While I appreciate the cox models showing that these differences were present, would it not strengthen the purpose of the paper to add a final model using the other co-variates while examining but for nodal metastasis use pN0 (vs. PNa, PNb, PNc) to see if your conclusions hold valid after adjusting for other factors. In the models, again was neoadjuvant or adjuvant used as this Is likely to impact OS and RFS.

Reply 10:
We appreciate the reviewer's instructive suggestion. RFS and OS are certainly not uniform and should be interpreted with caution. In our study, we performed adjuvant chemotherapy in 500 patients (31.9%). Since our cases were between 2010 and 2016, only conventional cytotoxic agents were used. Therefore, we do not believe that they have a significant impact on prognosis. Since many cases with mLNs receive adjuvant therapy, it is difficult to compare pN0 with pN1 and pN2 without adjuvant therapy. We have added the above to the Discussion section (P8Ln6-).

Comment 11:
1."the pN2 182 (mLNs = 1) group is a different population in patients with pN2 disease, and the number of 183 mLNs may be less important in the pN2 (mLNs ≥ 2) group"
This was confusing to me. First the extent of metastasis (was still different between 1 and 2+) in both cox models (but after 2 there is no difference and is confusing but it sounds like your saying extra after 2 is not important and this should just be stated more clearly).

Reply 11:
We apologize for the confusion. In our study, 4 or more mLNs in pN1 and 2 or more mLNs in pN2 were the cutoff value for each N descriptor. We believe that pN2 group with 1 mLN is originally a population with a good prognosis. The fact that the number of mLNs did not affect the prognosis in

pN2 (mLNs ≥ 2) group means that the anatomical location of mLNs (in other words, metastasis to the N2 region) is more important than the number of mLNs in pN2 (mLNs ≥ 2) group. We have added the above to the Discussion section (P7Ln19-).

Comment 12:
2. when doing the combination curves (while you showed earlier that pN1 1-3 and 4+ were similar) the combination are limited. You also state "The reason is that the pN2 group with 1 mLN suggest skip N2 181 metastasis without N1 involvement, which indicates a good prognosis (16,17)." But your results suggest that those with pN2 2+ was the worst group with no N1 disease, isn't this the exact opposite. (are these also not skip?). I think a lot of this confusion has to do with your use of prognosis to simplify, as OS was perhaps not as statistically bad between pNb and pNc (despite a clinically relevant difference in OS), but you showed RFS was worse (the reader really has to dig for this)

Reply 12:
We apologize for the confusion. Previous reports have shown that pN2 group with 1 mLN has a good prognosis, as did our study. OS may be affected by treatment after recurrence. We have added the above to the Discussion section (P7Ln28-).

Comment 13:
-furthermore 1. Why did you use 1-3 for group A but then 3-4 for group B. this does not allow a uniform conclusion. 2. if the goal is to say N1 has worse outcome, but N1 + N2 is worse (these patients would be classically just listed as N2. What would be helpful is to also say this is worse then pN2 (with only 1 LN) so you should pay attention to N1 in these cases, because right now we could just be seeing the impact of any pN2 disease. but once there is 2+ pN2 disease this trumps everything in terms of recurrence (based on cox model and not KM) thereby it is important to know the # of pN2 as well. To me adding that would support your conclusion unless I'm missing something

Reply 13:
We apologize for the confusion. Do you mean group A as pNa? pNa, pNb, and pNc represent the pN1 group with 1–3 mLNs, pN1 group with ≥ 4 mLNs plus pN2 group with 1 mLNs, and pN2 group with ≥ 2 mLNs, respectively. To show that pNb has a worse prognosis than pNa and a better prognosis than pNc, we have grouped pN1 group with ≥ 4 mLNs and pN2 group with 1 mLNs into a group called pNb. pN2 group included 160 patients, of which pN2 group with 1 mLNs included 27 patients. We have added the above to the Results section (P6Ln9-).

Comment 14:
-the 7th and 8th edition font is different from the rest of the paper

Reply 14:
We thank the reviewer for pointing the mistake. This has been corrected (P8Ln25-).

Comment 15:
-would add to discussion that your data suggest that guidelines such as NCCN rather than ACS that recommend zones of sampling are better than minimum as you show not just number but station matter

Reply 15:

We appreciate the reviewer's point. As reviewer said, the NCCN guideline recommend one or more nodes sampling from all mediastinal stations. We have added the above to the Discussion section (P8Ln21-).

Comment 16:
-should be addressed that it seems no patient underwent invasive pre-op sampling despite clinical suspicion, in practice if you suspected N2 this patient would have undergone induction therapy for likely IIIA disease and lack of this or details of anyone receiving neoadjuvant if indicated are not necessarily in line with standard practice.

Reply 16:
We performed endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) or mediastinoscopy if necessary to minimise the risk of undetected N2 disease. At that time, we did not perform neoadjuvant chemotherapy in clinical practice. We have added the above to the Methods section (P4Ln19-).

Reviewer E

Comment 1:
Can the authors please explain their Methods section better? Is their count of metastatic lymph nodes based only on radiology CT/PET scan results and clinical, pre-treatment diagnosis or not?

Reply 1:
We added multiple sentences to the Methods section to make it easier for the readers to understand. We did not count metastatic lymph nodes before surgery. We think it is difficult to recognize the number of mLNs using preoperative images. If a modality that can recognize the boundaries of lymph nodes appears in the future, preoperative evaluation will be possible.

Comment 2:
The last few sentences of their Discussion indicate that their count of metastatic lymph nodes is instead post treatment, the conclusion of which are already known. But if their analysis is based on pre-treatment, then their analysis is much more useful.

Reply 2:
It is difficult to recognize the number of mLNs using preoperative images. Therefore, it is unknown whether pathological stages can be applied to clinical nodal stages in lung cancer. The above is one of the limitations in our study.

Comment 3:
The authors' conclusion has already been confirmed by the eighth edition of the AJCC guidelines for lung cancer N-staging based on pathology analysis, that the progressive worsening of survival is correlated with the number of involved lymph node stations. But the AJCC eighth edition did not revise the N-stage to incorporate the number of lymph nodes because their findings were based only on pathology and not on pre-treatment clinical staging. AJCC discusses this on page 436, of the Cancer staging manual, section titled, "Rules for classification, Clinical Classification", Springer publishers, copyright 3rd printing, 2017. AJCC's Table 36.2 is the recommended future N-stage subclassification that includes the number of lymph nodes.
Their multivariate analysis results are troubling.

Recommendation: The authors did cite that there are proposed N-stage refinements in their Discussion section, but they may want to re-write their introduction, conclusion and discussion to also address AJCC's proposed Table 36.2 in the eighth edition which was published in 2017. Then the author's analysis and manuscript will be more timely.
Reply 3:
We appreciate the reviewer's instructive suggestion. Since we think AJCC proposal is important, we quoted AJCC Cancer Staging Manual. IASLC suggested that the number and station of mLNs might affect prognosis. Future subdivisions of the N descriptor will need to be combined with the number of lymph nodes. We have added the above to the conclusion of the Discussion section (P9Ln3-).

Comment 4:
The authors wrote, in their "Imaging and Lymph Node Dissection" section that high resolution CT and PET scans were used to determine if there was any suspicious activity (swelling and FDG uptake) around the lymph nodes. Then they "performed systematic lymph node dissection" – was this lymph node dissection after treatment after surgical resection of the primary tumor, or during clinical diagnosis? If it is post treatment, then their conclusion is similar to that of the eighth edition AJCC. It was somewhat confusing that the manuscript referred to pN1, pN2, pN3 patients, but used cN1 and cN2 patients in the Methods section.

Reply 4:
HRCT and PET were performed to accurately assess the presence or absence of lymph node metastases (not the number of metastatic lymph nodes) before surgery. If they revealed cN2, chemoradiation therapy was performed in some cases. The main focus of this study is pathological nodal stage, and clinical nodal stage is not examined. We apologize for the confusion. Lymph node dissection was done at the same time as surgical resection of the primary tumor.

Comment 5:
Other issues:
Although the total number of patients is large, 1567, the number of patients having any lymph node activity (stage N1 and higher) is smaller. There are only about 173+160 = 333 cases with non-zero lymph node metastasis. It is very hard to get more statistics, but can they request data from elsewhere?

Reply 5:
We performed PET in all patients and endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) or mediastinoscopy if necessary to minimise the risk of undetected N2 disease. Therefore, the frequency of unexpected lymph node metastases may be low. Our multicenter database is regularly brushed up to provide very accurate data. Therefore, unfortunately, it is difficult to increase the number of cases in this study.

Comment 6:
It is troubling that mathematically, the authors found the number of metastatic lymph nodes did NOT associate with survival in their multivariate analysis. The authors do acknowledge that their statistics are low. Can the authors explain better how they set up their multivariate analysis, did they drop out their input variables one by one to test their association results, etc? With low statistics, maybe they should just do 1 versus multiple mLNs, a bi-variate analysis for the 333 cases, and see what kind of results they get? Or zero versus non-zero mLNs to test their software setup.

Multivariate analysis can sometimes give very tricky results because mathematically the software can focus in on some numeric discrepancies that are not medically important.

Reply 6:
We also think this is a difficult problem. When we tried to drop out their input variables one by one, the number of metastatic lymph nodes was not associated with OS in multivariate analysis, probably because of the small number of cases and post-recurrence treatment. With the advent of immune checkpoint inhibitors, it is possible that post-recurrence treatment has been successful and that survival is longer than before. The fact that there was no significant difference in OS between the groups in Figure 2B may be related to the fact that the number of metastatic lymph nodes was not associated with OS in the multivariate analysis.

Reviewer F

Comment 1:
Methods section/Imaging and Lymph node dissection: did the authors perform any invasive mediastinal staging, especially for cN2 patients?

Reply 1:
We performed PET in all patients and endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) or mediastinoscopy if necessary to minimise the risk of undetected N2 disease. We have added the above to the Methods section (P4Ln19-).

Comment 2:
Methods section/statistical analysis: Once suspected by imaging, was recurrence confirmed by histology? How could imaging only proof recurrence?

Reply 2:
The follow-up examinations included a physical assessment, chest radiography, chest CT, and blood tests for the tumor marker carcinoembryonic antigen. PET and brain magnetic resonance imaging were performed in cases where the patient complained of neurological symptoms. Confirmation of recurrence were determined either radiologically or histologically. We have added the above to the Methods section (P5Ln9-).

Comment 3:
Results section: lobe specific dissection was performed in cN0 patients only, however 10.2% of patients were pN2 (4% only cN2) and 78.8% pN0 (cN0 84%). The authors should comment on the 6% mismatch and nodal upstaging in light of the chosen practice of lymph node dissection.

Reply 3:
We appreciate the reviewer's point. As you said, regarding pN2, upstaging of 6% from cN2 was observed. Although detailed preoperative examinations were performed, lymph node metastasis that cannot be detected by preoperative imaging were found. We have added the above to the Results section (P5Ln26-).

Comment 4:
If patients with neoadjuvant treatment were excluded, this should be noted in the mat/meth section.

Reply 4:
At that time, we did not perform neoadjuvant chemotherapy in clinical practice. We have added the above to the Methods section (P4Ln4-).

Comment 5:
In pN2 subgroup: does >2mLNs stand for 1 metastatic N1 node plus 1 metastatic N2 node or 0 metastatic N1 nodes and 2 metastatic N2 nodes? Did the authors take the distribution of metastatic nodes into account (single station affected, multiple stations affected)?

Reply 5:
pN2 with 2 mLNs included both 1 metastatic N1 node plus 1 metastatic N2 node and 0 metastatic N1 node and 2 metastatic N2 nodes. The former was 10 patients and the latter was 5 patients. It is certainly an important point, but due to the small number of patients in this study, we could not make a comparison.

Comment 6:
The effect of adjuvant therapy would be interesting and should be commented or put into multivariate analysis as a parameter.

Reply 6:
In our study, we performed adjuvant chemotherapy in 500 patients (31.9%). Since our cases were between 2010 and 2016, only conventional cytotoxic agents were used. Therefore, we do not believe that they are likely to have a significant impact on prognosis. Since many cases with mLNs receive adjuvant therapy, it is difficult to compare pN0 with pN1 and pN2 without adjuvant therapy. We have added the above to the Discussion section (P8Ln6-).

Comment 7:
The same with EGFR positive patients that were supposed to have inferior response to chemotherapy. The high number of patients tested positive is a good opportunity to analyze the prognostic impact of mLN in this subgroup.

Reply 7:
There were 140 EGFR-positive and 217 EGFR-negative patients who received adjuvant therapy. Of the former, 44 (31.4%) relapsed, and of the latter, 65 (29.9%) recurred. EGFR mutation in the pN1 (1–3 mLNs), pN1 (mLNs $\geq$ 4), pN2 (mLNs = 1), and pN2 (mLNs $\geq$ 2) were present in 26, 5,11, and 36 patients, respectively. Unfortunately, due to the small number of patients, it is difficult to compare with the log rank test.

Comment 8:
Table 3: the authors should comment on the significant prognostic impact in patients with “non-adenocarcinoma” histology (p 0.004, HR 2.433) and discuss the different results in N1 patients.

Reply 8:
We appreciate the reviewer’s point. Patients with adenocarcinoma may have a better prognosis than those with non-adenocarcinoma because the former have the opportunity to use new drugs such as tyrosine kinase inhibitor when they relapsed. In addition, patients with squamous cell carcinoma may be associated with interstitial pneumonia, and it may be difficult to receive effective treatment at the time of recurrence. We have added the above to the Discussion section (P8Ln1-).