Accepted Manuscript

The accuracy of clinical staging of stage I-IIIa non-small cell lung cancer: An analysis based on individual participant data

Neal Navani, David Fisher, Jayne F. Tierney, Richard J. Stephens, Sarah Burdett, on behalf of the NSCLC Meta-analysis Collaborative Group

PII: S0012-3692(18)32607-2
DOI: https://doi.org/10.1016/j.chest.2018.10.020
Reference: CHEST 2018

To appear in: CHEST

Received Date: 23 May 2018
Revised Date: 17 August 2018
Accepted Date: 2 October 2018


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
The accuracy of clinical staging of stage I-IIIa non-small cell lung cancer: An analysis based on individual participant data

Neal Navani¹, David Fisher², Jayne F Tierney², Richard J Stephens³ and Sarah Burdett⁵ on behalf of the NSCLC Meta-analysis Collaborative Group

Affiliations

¹Lungs for Living Research Centre, UCL Respiratory & Department of Thoracic Medicine, University College London Hospital, London, UK
²MRC Clinical Trials Unit at UCL, London, UK
³(Retired) MRC Clinical Trials Unit at UCL, London, UK

Word count (not including abstract)

2241

Corresponding Author

Neal Navani - n.navani@ucl.ac.uk

Author contributions

SB and DF had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. SB, DF, JT, RS and NN contributed substantially to the study design, data analysis and interpretation, and the writing of the manuscript.

Conflict of interest

No author has a conflict of interest
Abstract (246 words)

Background

Clinical staging of NSCLC helps determine prognosis and management of patients; few data exist on accuracy of clinical staging and the impact on treatment and survival of patients. We assessed whether participant or trial characteristics were associated with clinical staging accuracy as well as impact on survival.

Methods

We used individual participant data from RCTs, supplied for a meta-analysis of pre-operative chemotherapy (+/- radiotherapy) versus surgery alone (+/- radiotherapy) in NSCLC. We assessed agreement between clinical TNM (cTNM) stage at randomization and pathological TNM (pTNM) stage, for participants in the control group.

Results

Results are based on 698 patients who received surgery alone (+/- radiotherapy) with data for cTNM and pTNM stage. 46% of cases were cTNM stage I, 23% cTNM stage II and 31% cTNM stage IIIa. cTNM stage disagreed with pTNM stage in 48% of cases, with 34% clinically understaged and 14% clinically over-staged. Agreement was not associated with age (p=0.12), gender (p=0.62), histology (p=0.82), staging method (p=0.32) or year of randomisation (p=0.98). Poorer survival in understaged patients was explained by the underlying pTNM stage. Clinical staging failed to detect T4 disease in 10% of cases and misclassified nodal disease in 38%.

Conclusions
This study demonstrates suboptimal agreement between clinical and pathological staging. Discrepancies between clinical and pathological T and N-staging could have led to different treatment decisions in 10% and 38% of cases respectively. There is therefore a need for further research into improving staging accuracy for patients with stage I-IIla NSCLC.
Background

The clinical staging of non-small cell lung cancer (NSCLC) is of paramount importance in determining a patient’s prognosis, guiding treatment decisions and defining clinical trial eligibility, as well as allowing comparison between clinical trials. Incorrect staging of NSCLC may result in inaccurate prognostic information for patients and errors in patient management. After extra-thoracic metastases have been excluded, tumor and nodal staging are critical in making treatment decisions, as patients with N0 and N1 involvement are generally candidates for surgery. Patients with ipsilateral mediastinal disease (N2) are a heterogeneous group and may be offered chemo-radiation therapy or surgery (with pre-operative or post-operative chemotherapy). Patients with contra-lateral (N3) mediastinal (or supraclavicular) nodal disease are offered chemo-radiation therapy or palliative treatment options. Therefore, clinical under-staging, i.e. staging that misses mediastinal metastases or mediastinal invasion of the primary lesion may risk the patient undergoing radical treatment of the primary lesion for no benefit. Conversely, incorrect clinical over-staging of mediastinal disease may result in surgery being denied to an otherwise operable patient. The current guidance from the Union for International Cancer Control (UICC) suggests that when there is doubt about stage, the less advanced, or lower category should be chosen.

The emergence of techniques such as stereotactic body radiotherapy (SABR) and radiofrequency ablation (RFA) to treat early stage NSCLC in medically inoperable patients has further highlighted the importance of accurate clinical staging. Applying local non-surgical treatments without the benefit of systematic lymph node dissection runs the risk of being futile if there is clinical under-staging with unrecognized mediastinal or systemic disease.
Although the importance of accurate clinical staging is clear and the performance characteristics of individual tests in lung cancer staging are known, fewer data exist on the accuracy of clinical staging of NSCLC and how this relates to the staging techniques employed. Three studies that have been reported all show high levels of inaccurate clinical staging; however none have demonstrated the impact of erroneous staging on clinical outcome. A prospective study of 383 patients with potentially resectable NSCLC demonstrated that clinically unsuspected N2 disease was found in 14% of patients. Despite routine use of PET-CT scanning, a post-hoc analysis of 67 patients from the control arm of the MRC LU22 trial of pre-operative chemotherapy suggested that nodal staging was inaccurate in 25% (95% CI 15 – 36%) of patients who underwent PET-CT scanning and mediastinoscopy. A recently published study comparing clinical and pathological TNM data collected for 2336 patients included in the Dutch Lung Surgery Audit, showed that only 54% of patients were clinically staged accurately and no comment could be made on whether this impacted on patient survival outcomes. Thus, to investigate further, we used individual participant data (IPD) from trials supplied for a systematic review and meta-analysis of pre-operative chemotherapy in non-small cell lung cancer to assess the accuracy of clinical staging, factors that may affect inaccuracy and how inaccuracy might impact on treatment decisions and survival.

**Methods**

To be eligible for inclusion in the original IPD meta-analysis, trials should have randomized patients with NSCLC to pre-operative chemotherapy followed by surgery (+/- post-operative radiotherapy) versus surgery (+/- post-operative radiotherapy). Full details of the methods
are presented elsewhere\textsuperscript{8}. IPD were collected for fifteen eligible randomized controlled trials and included 2385 patients with non-small cell lung cancer\textsuperscript{8}. However, only data from patients from the control arm in these trials were used in this analysis, to ensure that any difference between clinical and pathological staging could not have been influenced by pre-operative chemotherapy. Included RCTs used different editions of TNM staging and these changes over time were taken into account (e-appendix 1).

Data on age, gender, clinical staging techniques, clinical TNM stage, extent of resection, pathological TNM stage, histology, performance status, treatment group and dates of randomization, last-follow-up and death were collected. We approached study investigators for permission to use these data for these analyses and for clarification where staging methods were unclear in the original trial protocol or manuscript.

Statistical analysis

To assess agreement between clinical TNM stage (cTNM) and pathological TNM stage (pTNM), a simple percentage agreement was calculated. Agreement between clinical and pathological stage was also calculated using a weighted Cohen’s kappa, which takes into account both agreement by chance and the degree of disagreement. Kappa statistics were categorised, as \(<66\%\)= low agreement, \(\geq 66\%\)= fair agreement and \(\geq 90\%\)= good agreement\textsuperscript{9, 10}.

To assess whether or not patient and trial characteristics might be associated with any cTNM staging inaccuracy age, gender, histology, year of randomisation and staging method were included in a multivariate logistic regression model. Histology was classified into adenocarcinoma, squamous, and other/unknown. Staging methods were classified as CT
scan with or without a chest X-ray or CT scan plus any other staging method, as there were insufficient data to do this in more detail. Staging method correlated strongly with year of randomization, so we only included the former in our primary analysis. However, a sensitivity analysis was also performed, where staging method was replaced with year of randomization. We generated Kaplan-Meier curves for overall survival based on patients who were clinically under-staged, clinically over-staged and for those whose cTNM and pTNM agreed, and compared these using a log-rank test, stratified by trial and subsequently also pathological stage. The accuracy of clinical T stage and nodal status were considered separately to help pinpoint which disagreements could have influenced treatment decisions.

Role of the funding source

Funded by the UK Medical Research Council MC_UU_12023/28. The sponsors of the original trials had no role in this study design, data collection, data analysis, data interpretation, or writing of the report. No IRB approval is needed.

Results

Fifteen RCTs were included in the original IPD systematic review and meta-analysis of pre-operative chemotherapy followed by surgery versus surgery alone. Nine trials (randomising 1,586 patients in total) included data on both cTNM and pTNM stage, providing 698 control-arm patients for analysis (Table 1). These RCTs accrued patients between 1987 and 2005.
Clinical staging protocols varied between the trials (Table 1). One trial [11](which recruited patients between 1987 and 1993) used a chest x-ray and mediastinoscopy only. More recent trials used CT scans and PET-CT, but no trial utilised PET-CT scanning routinely, such that only 67 patients included in the analysis underwent PET-CT. There was also variation between trials in the surgical methods used (Table 1).

Of the 698 patients included, 318 (46%) were cTNM stage I (83% of which were Ia), 160 (23%) were cTNM stage II (91% of which were IIa), and 218 (31%) were cTNM stage IIIa (Table 2). Only 2 patients were classed as cTNM stage IIIB, and were therefore not included in the regression or survival analyses. A more detailed breakdown is given in e-appendix 2.

Agreement between cTNM and pTNM staging was low (52%, weighted Cohen’s kappa=0.35 (95% CI 0.30 to 0.40) (Table 2). In 34% of cases, patients were clinically under-staged, and in 14% of cases, patients were clinically over-staged (Table 2). In the main regression analysis, age (p=0.12), gender (p=0.62), histology (p=0.82) or the staging method (p=0.32) were not significantly associated with the accuracy of cTNM staging and in a sensitivity there was no association with year of randomization (p=0.98; e-appendix 3).

Survival varied with the accuracy of cTNM staging. In particular, patients who were clinically under-staged appeared to have poorer survival than those who were clinically over-staged or those for whom cTNM and pTNM staging agreed (log-rank test stratified by trial p<0.0001; Figure 1). However, this is driven by the underlying pTNM stage (log-rank test stratified by trial and pathological stage p=0.54), which is more clearly illustrated in Figure 2. In particular, 44% of patients classed as cTNM stage I were pTNM stage II-IV, and 33% of patients classed as cTNM stage II were pTNM stage III-IV, explaining their lower survival (Figure 2).
Agreement was low between clinical and pathological T stage (65%, weighted Cohen’s kappa=0.33 (95% CI 0.27 to 0.39), Table 3) and N stage (62%, weighted Cohen’s kappa=0.42, (95% CI 0.37 to 0.48), Table 4). Specifically, clinical staging failed to detect T4 disease in 10% of patients (Table 3), and nodal disease in 19% of patients. In addition, 12% were judged erroneously to have node positive disease (Table 4).

Discussion

Results summary

We found that cTNM stage disagreed with pTNM stage in around a half of patients, and was not clearly associated with age, gender, histology, the staging method used or year of randomization. The discrepancies between clinical and pathological T-staging and N-staging could have led to different treatment decisions in 10% and 38% of cases respectively.

Strengths

To our knowledge, this is the first time IPD from major RCTs have been combined to assess the accuracy of staging in stage I-III NSCLC. Whilst the randomized controlled trials included did not intend to evaluate staging, with the agreement of those who provided the data, this novel methodology provided us with a valuable opportunity to investigate more reliably the accuracy of clinical TNM staging. We could take advantage of per protocol clinical staging and surgery and rigorous documentation of clinical and pathological TNM stage for each
patient. Also, data from randomized trials are less susceptible to the selection biases that can affect cohort studies\textsuperscript{19,20}. Using IPD has enabled us to restrict the analysis to the control arms of these trials, thus avoiding confounding by treatment received and, in particular, potential downstaging from use of pre-operative chemotherapy.

For the first time, this study also demonstrates the impact of the inaccuracy of clinical staging on patient survival outcomes. Importantly, the impact of staging accuracy on clinical decision making is also demonstrated using unselected data. The poorer survival seen in clinically understaged patients was explained by the underlying pTNM stage.

Limitations

Over time the trials included here used increasingly sophisticated staging methods, but surprisingly, a significant improvement in accuracy was not seen. However, many of the staging methods utilised in the included trials may now be considered sub-optimal\textsuperscript{21}. Earlier studies employed CT scanning and mediastinoscopy while the most recent trial used additional PET-CT, but none used endosonography. Despite this, our staging accuracy results are remarkably similar to those from the audit of the quality of staging in Dutch patients\textsuperscript{7} which included routine use of PET-CT and endosonography and included patients from January 2013-December 2014. Indeed, of the patients included in our analysis that did undergo PET-CT, a quarter of cases were still understaged and this is discussed elsewhere\textsuperscript{6}. While PET-CT or endosonography was not routinely utilized in the trials included in this meta-analysis, this practice reflects current American College of Chest Physicians’ guidance\textsuperscript{22} for patients with stage 1A disease which does not recommend the use of PET or...
endosonography. Although it is difficult to generalise, assuming the trial population reflects routine practice, the data here suggest that 44% of patients with clinical stage 1 disease might have more advanced disease diagnosed post-operatively. A further limitation is that intra-operative pathological staging protocols may have varied and are unlikely to be as comprehensive as currently recommended\(^{23}\). However, incomplete pathological staging would only serve to reduce the extent of nodal staging inaccuracy.

Context

The advent of stereotactic radiotherapy and radiofrequency ablation for the treatment of early stage NSCLC has highlighted the importance of accurate nodal staging. These newer techniques are used for the treatment for early stage lung cancer but, in contrast to surgery, do not provide pathological staging information. In a study of relapse of NSCLC following stereotactic radiotherapy or surgery, there were twice as many recurrences in local lymph nodes in patients undergoing stereotactic radiotherapy compared to surgery\(^{24}\), emphasizing the importance of accurate nodal staging prior to SABR.

When surgery is undertaken and pathological staging is available, prior invasive mediastinal sampling may take on less significance if we assume that surgery followed by adjuvant chemotherapy is at least as effective as chemo-radiation. When considering stage II and III disease, inaccurate clinical staging may reduce the efficacy of surgery by failing to detect multi-station N2 or N3 disease. For patients undergoing radical radiotherapy, imprecise clinical staging can result in an incorrect radiation field.
The most likely explanation for the low level of accuracy of clinical staging for patients with operable NSCLC is the sensitivity of the diagnostic tools employed. Patients being considered for treatment with curative intent typically undergo CT and PET-CT imaging as well as mediastinal sampling when required. Using a 10mm short axis cut-off for significance of mediastinal nodes, the sensitivity of CT scanning in detecting mediastinal metastases is 55%. PET-CT has a sensitivity of 77-81% and may vary according to brand of scanner and histology. In a systematic pooled analysis of 9267 patients, mediastinoscopy had a sensitivity of 78%. Overstaging may occur with PET-CT unless current guidelines [22] are adhered to and PET positive findings are clarified by invasive sampling. More recently the introduction of endobronchial and endoscopic ultrasound has improved the clinical staging of patients with NSCLC, resulting in a reduction in futile surgery and potentially increased survival when employed routinely for patients with stage I-III disease.

Implications

These findings have implications for the care of patients with NSCLC, as well as appropriate selection of suitable patients for inclusion in clinical trials. Under-staging the T stage may mean that the patient undergoes surgery without the surgeon knowing the full extent of the primary disease, which may result in an incomplete resection. 10% of patients in our analysis were found to have previously unexpected T4 disease. Erroneous nodal staging in patients without metastatic disease can similarly result in inappropriate treatment decisions, which can significantly impact on patient outcomes. Patients with nodal disease undetected by clinical staging methods may undergo futile surgery (or SABR) whereas chemo-radiotherapy may have been the preferred initial treatment of clinicians and patients
with full knowledge of nodal involvement. Conversely, if clinical staging overestimates the extent of nodal disease (114 (15%) of patients in this meta-analysis) then this may mean patients are denied potentially curative surgery. The data for this analysis were obtained from patients in controlled clinical trials, generally from centers with lung cancer expertise. Therefore, clinical staging accuracy in the wider population could well be worse.

**Conclusions**

The results of this analysis highlight some flaws in the clinical care of patients with NSCLC and emphasize the need for further research into techniques for improving staging accuracy for patients with stage I-III NSCLC.

**Acknowledgements**

We thank all the patients who took part in all the trials included in these analyses. Publication is on behalf of the Non-Small Cell Lung Cancer Collaborative Group

**NSCLC Meta-analysis Collaborative Group**

*Project Management Group*

Sarah Burdett, Larysa HM Rydzewska, Jayne F Tierney - MRC Clinical Trials Unit at UCL, London, UK

Anne Auperin, Thierry Le Chevalier, Cécile Le Pechoux, Jean-Pierre Pignon - Gustave-Roussy, Villejuif, France

*International Advisory Group*
Navani – Accuracy of staging in NSCLC

Rodrigo Arriagada, Karolinska Institutet, Stockholm, Sweden; Gustave-Roussy, Villejuif, France

David H Johnson, University of Texas, Southwestern Medical Center, Dallas, TX, US

Jan van Meerbeeck, MOCA-Thoracic Oncology, University Hospital Antwerp, Belgium

Mahesh KB Parmar, MRC Clinical Trials Unit at UCL, London, UK

Richard J Stephens, MRC Clinical Trials Unit at UCL, London, UK (retired)

Lesley A Stewart, Centre for Reviews and Dissemination, York, UK

Collaborators who supplied individual participant data

Paul A Bunn, University of Colorado Cancer Centre, Aurora, Colorado, USA; Bertrand Dautzenberg, Service de Pneumologie et Réanimation, Groupe Hospitalier Pitié-Salpêtrière, Paris, France; David Gilligan, Addenbrooke’s Hospital, Cambridge, UK; Harry Groen, Universitair Medisch Centrum Groningen, Groningen, The Netherlands; Aija Knuuttila, Helsinki University Central Hospital, Helsinki, Finland; Eric Vallieres, Swedish Cancer Institute, Seattle, WA 98104; Rafael Rosell, Catalan Institute of Oncology, Hospital Germans Trias i Pujol, Barcelona, Spain; Jack Roth, University of Texas MD Anderson Cancer Centre, Houston, Texas, USA; Giorgio Scagliotti, University of Turin, San Luigi Hospital, Torino, Italy; Masahiro Tsuboi, National Cancer Center Hospital East, Kashiwanoha, Kashiwa-shi, Japan; David Waller, Glenfield Hospital, Leicester, UK; Virginie Westeel, Centre Hospitalier Universitaire, Besançon, France; Yi-Long Wu, Guangdong Lung Cancer Institute, Guangdong General Hospital and Guangdong Academy of Medical Sciences, Guangzhou, PR China; Xue-Ning Yang, Guangdong Lung Cancer Institute, Guangdong General Hospital & Guangdong Academy of Medical Sciences, Guangzhou, PR China
References


Navani – Accuracy of staging in NSCLC


Figure legends

Table 1: Characteristics of included trials

Table 2: Agreement between clinical and pathological TNM stage data

Table 3: Agreement between clinical and pathological of T stage data

Table 4: Agreement between clinical and pathological nodal status data

Figure 1: Kaplan-Meier curves for overall survival for all trial data combined, by agreement of clinical TNM staging with pathological TNM staging

Figure 2: Kaplan-Meier curves for overall survival in clinically staged 1, 2 and 3 patients, by agreement of clinical TNM staging with pathological TNM staging

e-appendix 1: Comparison of TNM staging systems

e-appendix 2: Flowchart describing clinical and pathological agreement, clinical over staging and clinical under staging

e-appendix 3: Multivariate logistic regression; Factors that may predict staging agreement
<table>
<thead>
<tr>
<th>Trial</th>
<th>Total patients randomised</th>
<th>Patients randomised to the control arm</th>
<th>Patients that provided clinical and pathological data</th>
<th>Accrual Period</th>
<th>Staging System (TNM)*</th>
<th>Staging Method</th>
<th>Surgical Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD Anderson (US A) 1994(11)</td>
<td>60</td>
<td>32</td>
<td>32</td>
<td>87-93</td>
<td>Chest x-ray</td>
<td></td>
<td>One or more positive nodal stations allowed. Patients with left lung tumors and paratracheal lymph node metastases excluded</td>
</tr>
<tr>
<td>MIP-91(France)(12, 29)</td>
<td>355</td>
<td>176</td>
<td>170</td>
<td>91-97</td>
<td>Chest x-ray, CT</td>
<td></td>
<td>Mediastinal node dissection and node sampling were left to the discretion of the surgeon</td>
</tr>
<tr>
<td>Netherlands 2000(13)</td>
<td>79</td>
<td>40</td>
<td>37</td>
<td>91-99</td>
<td>CT and mediastinoscopy</td>
<td></td>
<td>Mediastinal lymph node exploration was encouraged: for right-sided lesions, this included 2R, 4R, 7, 8, 9. For left-sided lesions, this included 4L, 5, 6, 7, 8, 9.</td>
</tr>
<tr>
<td>JCOG 9209 (Japan)(14)</td>
<td>62</td>
<td>31</td>
<td>31</td>
<td>93-98</td>
<td>CT</td>
<td></td>
<td>Surgery was either lobectomy, bilobectomy, or pneumonectomy along with systematic mediastinal lymph node dissection.</td>
</tr>
<tr>
<td>Finland 2003(15)</td>
<td>62</td>
<td>32</td>
<td>23</td>
<td>95-99</td>
<td>CT</td>
<td></td>
<td>‘Local surgery’</td>
</tr>
<tr>
<td>MRC LU22(UK)(5)</td>
<td>519</td>
<td>261</td>
<td>194</td>
<td>97-05</td>
<td>Bronchoscopy, mediastinoscopy and CT, PET</td>
<td></td>
<td>At cervical mediastinoscopy, the following lymph node stations will, wherever possible, be sampled: 2R, 2L, 4R, 4L, 7</td>
</tr>
<tr>
<td>SWOG S9900 (USA)(16)</td>
<td>354</td>
<td>174</td>
<td>170</td>
<td>99-04</td>
<td>Chest x-ray and CT</td>
<td></td>
<td>All accessible hilar (level 10) lymph nodes must be dissected ...A complete</td>
</tr>
</tbody>
</table>
Mediastinal lymph node sampling should be performed... for right-sided lesions, this includes 2R, 4R, 7, 8 and 9. For left-sided lesions, this includes 4L, 5, 6, 7, 8 and 9.

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Number</th>
<th>Age</th>
<th>Year of Diagnosis</th>
<th>Chest Imaging and Bronchoscopy</th>
<th>Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>China 2002(17)</td>
<td>55</td>
<td>23</td>
<td>20</td>
<td>99-04</td>
<td>5/6</td>
<td>Chest x-ray, CT, bronchoscopy and abdominal ultrasound</td>
</tr>
<tr>
<td>China 2005(18)</td>
<td>40</td>
<td>21</td>
<td>21</td>
<td>99-04</td>
<td>5/6</td>
<td>Chest x-ray, CT, bronchoscopy and abdominal ultrasound</td>
</tr>
</tbody>
</table>

* For details of TNM Staging systems, see Appendix 1
<table>
<thead>
<tr>
<th>TNM stage</th>
<th>pI (Percent)</th>
<th>pII (Percent)</th>
<th>pIIIa (Percent)</th>
<th>pIIIb (Percent)</th>
<th>pIV (Percent)</th>
<th>Total (Percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cl</td>
<td>177 (25.4%)</td>
<td>72 (10.3%)</td>
<td>44 (6.3%)</td>
<td>22 (3.2%)</td>
<td>3 (0.4%)</td>
<td>318 (45.6%)</td>
</tr>
<tr>
<td>cII</td>
<td>40 (5.7%)</td>
<td>67 (9.6%)</td>
<td>32 (4.6%)</td>
<td>16 (2.3%)</td>
<td>5 (0.7%)</td>
<td>160 (22.9%)</td>
</tr>
<tr>
<td>cIIIa</td>
<td>32 (4.6%)</td>
<td>28 (4.0%)</td>
<td>116 (16.6%)</td>
<td>30 (4.3%)</td>
<td>12 (1.7%)</td>
<td>218 (31.2%)</td>
</tr>
<tr>
<td>cIIIb</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2 (0.3%)</td>
<td>0</td>
<td>2 (0.3%)</td>
</tr>
<tr>
<td>cIV</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>249 (35.7%)</td>
<td>167 (23.9%)</td>
<td>192 (27.5%)</td>
<td>70 (10.0%)</td>
<td>20 (2.9%)</td>
<td>698 (100%)</td>
</tr>
</tbody>
</table>

- Clinically overstaged
- Clinically understaged
<table>
<thead>
<tr>
<th>T stage</th>
<th>pT1</th>
<th>pT2</th>
<th>pT3</th>
<th>pT4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>cT1</td>
<td>34 (4.9%)</td>
<td>16 (2.3%)</td>
<td>3 (0.4%)</td>
<td>7 (1.0%)</td>
<td>60 (8.6%)</td>
</tr>
<tr>
<td>cT2</td>
<td>35 (5.0%)</td>
<td>360 (51.6%)</td>
<td>69 (9.9%)</td>
<td>40 (5.7%)</td>
<td>504 (72.2%)</td>
</tr>
<tr>
<td>cT3</td>
<td>7 (1.0%)</td>
<td>42 (6.0%)</td>
<td>60 (8.6%)</td>
<td>23 (3.3%)</td>
<td>132 (18.9%)</td>
</tr>
<tr>
<td>cT4</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>2 (0.3%)</td>
<td>2 (0.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>76 (10.9%)</td>
<td>418 (59.9%)</td>
<td>132 (18.9%)</td>
<td>72 (10.3%)</td>
<td>698 (100%)</td>
</tr>
</tbody>
</table>

- Clinically overstaged
- Clinically understaged
<table>
<thead>
<tr>
<th>Nodal status</th>
<th>pN0</th>
<th>pN1</th>
<th>pN2</th>
<th>pN3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>cN0</td>
<td>259 (37.1%)</td>
<td>74 (10.6%)</td>
<td>57 (8.2%)</td>
<td>1 (0.1%)</td>
<td>391 (56.0%)</td>
</tr>
<tr>
<td>cN1</td>
<td>56 (8.0%)</td>
<td>67 (9.6%)</td>
<td>29 (4.2%)</td>
<td>0</td>
<td>152 (21.8%)</td>
</tr>
<tr>
<td>cN2</td>
<td>28 (4.0%)</td>
<td>19 (2.7%)</td>
<td>104 (14.9%)</td>
<td>4 (0.6%)</td>
<td>155 (22.2%)</td>
</tr>
<tr>
<td>cN3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>343 (49.1%)</td>
<td>160 (22.9%)</td>
<td>190 (27.2%)</td>
<td>5 (0.7%)</td>
<td>698 (100%)</td>
</tr>
</tbody>
</table>

- Clinically overstaged
- Clinically understaged
TNM-stage

Survival over time for different TNM stages:
- Clinical under-staging
- Agreement
- Clinical over-staging

Analysis time (years)
Clinical TNM stage
Total N=696

<table>
<thead>
<tr>
<th>Clinical TNM stage</th>
<th>Stage I</th>
<th>Stage II</th>
<th>Stage III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>318 (46%)</td>
<td>160 (23%)</td>
<td>218 (31%)</td>
</tr>
</tbody>
</table>

Pathological stage I
Pathological stage II
Pathological stage III
Pathological stage IV

<table>
<thead>
<tr>
<th>Analysis time (years)</th>
<th>Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1.00</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Clinical under-staging: 44% | 33% | 19%
Agreement: 56% | 42% | 53%
Clinical over-staging: 0% | 25% | 28%
CT – Computed tomography
IPD – Individual participant data
MRC – Medical Research Council
NSCLC – Non-small cell lung cancer
PET-CT - Positron emission tomography–computed tomography
RCT – Randomised controlled trial
RFA - Radiofrequency ablation
SABR - Stereotactic body radiotherapy
UCL – University College London
UICC - Union for International Cancer Control
**e-Table 1.**
Comparison of TNM staging systems

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage I</strong></td>
<td><strong>Stage IA</strong></td>
</tr>
<tr>
<td>T1,N0,M0</td>
<td>T1,N0,M0</td>
</tr>
<tr>
<td>T2,N0,M0</td>
<td>T2,N0,M0</td>
</tr>
<tr>
<td><strong>Stage II</strong></td>
<td><strong>Stage IIA</strong></td>
</tr>
<tr>
<td>T1,N1,M0</td>
<td>T1,N1,M0</td>
</tr>
<tr>
<td>T2,N1,M0</td>
<td>T2,N1,M0</td>
</tr>
<tr>
<td><strong>Stage IIIA</strong></td>
<td><strong>Stage IIIB</strong></td>
</tr>
<tr>
<td>T1,N2,M0</td>
<td>T1,N2,M0</td>
</tr>
<tr>
<td>T2,N2,M0</td>
<td>T2,N2,M0</td>
</tr>
<tr>
<td>T3,N0/1/2,M0</td>
<td>T3,N1/2,M0</td>
</tr>
<tr>
<td>anyT,N3,M0</td>
<td>anyT,N3,M0</td>
</tr>
<tr>
<td>T4,anyN,M0</td>
<td>T4,anyN,M0</td>
</tr>
<tr>
<td><strong>Stage IIIB</strong></td>
<td><strong>Stage IV</strong></td>
</tr>
<tr>
<td>anyT, anyN, M1</td>
<td>anyT,anyN,m1</td>
</tr>
</tbody>
</table>

Online supplements are not copyedited prior to posting and the author(s) take full responsibility for the accuracy of all data.
e-Figure 1. Flowchart describing clinical and pathological agreement, clinical over staging and clinical under staging

All patients  
n=698 (100%)

Clinically understaged  
n=236 (34%)

- cT < pT  
n=75 (32%)  
- cN < pN  
n=100 (42%)  
- Both  
n=41 (17%)  
- pM1  
n=20 (8%)

- cN > pN  
n=4 (5%)  
- cT > pT  
n=9 (9%)

Clinically overstaged  
n=100 (14%)

- cT > pT  
n=17 (17%)  
- cN > pN  
n=72 (72%)  
- Both  
n=11 (11%)

- cN < pN  
n=1 (6%)  
- cT < pT  
n=4 (6%)

Agreement  
n=362 (52%)

- cT < pT  
n=75 (32%)  
- cN < pN  
n=100 (42%)  
- Both  
n=41 (17%)  
- pM1  
n=20 (8%)

- cN > pN  
n=4 (5%)  
- cT > pT  
n=9 (9%)

Unforeseen pN2+  
n=103 (44% of 236)  

- of which:  
n=10 (10%) pM1  
n=25 (24%) both cT < pT and cN < pN  
n=7 (7%) cT > pT
### e-Table 2.
Multivariate logistic regression; Factors that may predict staging agreement

<table>
<thead>
<tr>
<th>Predictor</th>
<th>TNM stage</th>
<th>$\chi^2$ (df)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histology</td>
<td></td>
<td>0.40 (2)</td>
<td>0.82</td>
</tr>
<tr>
<td>Staging method</td>
<td></td>
<td>1.01 (1)</td>
<td>0.32</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>2.48 (1)</td>
<td>0.12</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>0.24 (1)</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>Overall</strong>*</td>
<td></td>
<td><strong>4.22 (5)</strong></td>
<td><strong>0.52</strong></td>
</tr>
</tbody>
</table>

"Overall" compares the model with all covariates entered to the null model

Sensitivity analysis with staging method replaced with year of accrual:

<table>
<thead>
<tr>
<th>Predictor</th>
<th>TNM stage</th>
<th>$\chi^2$ (df)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histology</td>
<td></td>
<td>0.48 (2)</td>
<td>0.79</td>
</tr>
<tr>
<td>Year of randomisation</td>
<td></td>
<td>0.00 (1)</td>
<td>0.98</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>2.55 (1)</td>
<td>0.11</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>0.19 (1)</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Overall</strong>*</td>
<td></td>
<td><strong>3.21 (5)</strong></td>
<td><strong>0.67</strong></td>
</tr>
</tbody>
</table>