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

Please cite this article as: Navani N, Fisher D, Tierney JF, Stephens RJ, Burdett S, on behalf of the NSCLC Meta-analysis Collaborative Group, The accuracy of clinical staging of stage I-IIIa non-small cell lung cancer: An analysis based on individual participant data, *CHEST* (2018), doi: https://doi.org/10.1016/j.chest.2018.10.020.

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

Neal Navani[1], David Fisher[2], Jayne F Tierney[2,] Richard J Stephens[3] and Sarah Burdett[2] on behalf of the NSCLC Meta-analysis Collaborative Group

Affiliations

[1]Lungs for Living Research Centre, UCL Respiratory & Department of Thoracic Medicine, University College London Hospital, London, UK

[2] MRC 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-IIIa 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)[1] 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[2] (SABR) and radiofrequency ablation[3] (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[4], a post-hoc analysis of 67 patients from the control arm of the MRC LU22[5] 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[6]. A recently published study comparing clinical and pathological TNM data collected for 2336 patients included in the Dutch Lung Surgery Audit[7], 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[8], 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[8]. IPD were collected for fifteen eligible randomized controlled trials and included 2385 patients with non-small cell lung cancer[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, ≥66%= fair agreement and ≥90%=good agreement[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[5,11-18] (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[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[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[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[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[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%[22]. PET-CT has a sensitivity of 77-81%[25] and may vary according to brand of scanner and histology. In a systematic pooled analysis of 9267 patients, mediastinoscopy had a sensitivity of 78%[22]. 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[26,27] and potentially increased survival[28] 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

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, Kashiwas-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

1. Brierley JD, Gospodarowicz MK, Wittekind C. *TNM Classification of Malignant Tumours.* Vol 8. Oxford: Wiley-Blackwell; 2016.

2. Timmerman R, Paulus R, Galvin J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA.* 2010;303(11):1070-1076.

3. Lencioni R, Crocetti L, Cioni R, et al. Response to radiofrequency ablation of pulmonary tumours: a prospective, intention-to-treat, multicentre clinical trial (the RAPTURE study). *Lancet Oncology.* 2008;9(7):621-628.

4. Cerfolio RJ, Bryant AS, Ojha B, Eloubeidi M. Improving the inaccuracies of clinical staging of patients with NSCLC: a prospective trial. *Annals of Thorac Surgery.* 2005;80(4):1207-1213.

5. Gilligan D, Nicolson M, Smith I, et al. Preoperative chemotherapy in patients with resectable non-small cell lung cancer: results of the MRC LU22/NVALT2/EORTC 08012 multicentre randomised trial and update of systematic review. *Lancet.* 2007;369(9577):1929-1937.

6. Navani N, Nankivell M, Stephens RJ, et al. Inaccurate clinical nodal staging of non-small cell lung cancer: evidence from the MRC LU22 multicentre randomised trial. *Thorax.* 2010;65(5):463.

7. Heineman DJ, Geert ten Berge M, Daniels JM, et al. The quality of staging non-small cell lung cancer in the Netherlands: Data from the Dutch Lung Surgery Audit. *Ann. Thorac. Surg.* 2016;102:1622-1629.

8. NSCLC Meta-analysis Collaborative Group. Preoperative chemotherapy for non-small cell lung cancer: a systematic review and meta-analysis of individual participant data. *Lancet.* 2014;383:1561-1571.

9. McMinn AM, van Sluijs EMF, Harvey NC, et al. Validation of a maternal questionnaire on correlates of physical activity in preschool children. *International Hournal of Behavioral Nutrition and Physical Avtivity.* 2009;6:81.

10. Portney LG, Watkins MP. *Foundations of clinical research: Applications to practice.* 2nd edition ed. Upper Saddle River, New Jersey, USA: Prentice-Hall; 2000.

11. Roth JA, Fosella F, Komaki R, et al. A randomized trial comparing perioperative chemotherapy and surgery with surgery alone in resectable stage IIIa non-small cell lung cancer. *J. Natl. Cancer Inst.* 1994;86(9):673-680.

12. Depierre A, Milleron B, Moro-Sibilot D, et al. Preoperative chemotherapy followed by surgery compared with primary surgery in resectable stage I (except T1N0), II and IIIa non-small cell lung cancer. *J. Clin. Oncol.* 2002;20(1):247-253.

13. Splinter TA, van Putten JW, Meuzalaar J, Smit EF, Kho GS, Groen HJ. Randomized multicentre phase II study of chemotherapy followed by surgery versus surgery alone in stage I and II non-small cell lung cancer (NSCLC). *Proc Am Soc Clin Oncol.* 2000;19:495.

14. Nagai K, Tsuchiya R, Mori T, et al. A randomised trial comparing induction chemotherapy followed by surgery with surgery alone for patients with stage IIIa N2 non-small cell lung cancer. *J. Thorac. Cardiovasc. Surg.* 2003;125:254-260.

15. Mattson KV, Abratt RP, ten Velde G, Krofta K. Docetaxel as neoadjuvant therapy for radically treatable stage III non-small cell lung cancer: a multinational randomised phase III study. *Ann. Oncol.* 2003;14:116-122.

16. Pisters KM, Vallieres E, Crowley JJ, et al. Surgery with or without preoperative paclitaxel and carboplatin in early-stage non-small-cell lung cancer: Southwest Oncology Group Trial S9900, an intergroup, randomized, phase III trial. *J. Clin. Oncol.* 2010;28(11):1843-1849.

17. Wu Y-L, Gu L-J, Weng Y-M, Feng W-N, Cheng C. Neo-adjuvant chemotherapy with docetaxel plus carboplatin for non-small cell lung cancer. *Ann. Oncol.* 2002;13(Suppl 5):140.

18. Yang X, Wu Y, Gu L, et al. A randomized trial comparing neoadjuvant gemcitabine plus carboplatin or cisplatin followed by surgery with surgery alone in Clinical Stage IIIA non-small-cell lung cancer (NSCLC). *Lung Cancer.* 2005;49(Suppl. 2):S288, P-645.

19. Lopez-Encuentra A, Garcia-Lujan R, Rivas JJ, Rodriguez-Rodriguez J, Torres-Lanza J, Varela-Simo G. Comparison between clinical and pathologic staging in 2994 cases of lung cancer. *Ann. Thorac. Surg.* 2005;79(3):974-979.

20. Stiles BM, Servais EL, Lee PC, J.L. P, Paul S, Altorki NK. Point: Clinical stage IA non-small cell lung cancer determined by computed tomography and positron emission tomography is frequently not pathologic IA non-small cell lung cancer: the problem of understaging. *J. Thorac. Cardiovasc. Surg.* 2009;137(1):13-19.

21. Postmus PE, Kerr KM, Oudkerk M, et al. Early and locally advanced non-small cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 2017;28 (Suppl 4):iv1-iv21.

22. Silvestri GA, Gonzalez AV, Jantz MA, et al. Methods for staging non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest.* 2013;143(5 Suppl):e211S-250S.

23. Lardinois D, P. DL, Van Schil P, et al. ESTS guidelines for intraoperative lymph node staging in non-small cell lung cancer. *European Journal of Cardio-thoracic Surgery.* 2006; 30:787-792.

24. van den Berg LL, Klinkenberg TJ, Groen HJ, Widder J. Patterns of Recurrence and Survival after Surgery or Stereotactic Radiotherapy for Early Stage NSCLC. *J Thorac Oncol.* 2015;10(5):826-831.

25. Schmidt-Hansen M, Baldwin DR, Zamora J. FDG-PET/CT imaging for mediastinal staging in patients with potentially resectable non-small cell lung cancer. *JAMA.* 2015;313(14):1465-1466.

26. Annema JT, van Meerbeeck JP, Rintoul RC, et al. Mediastinoscopy vs endosonography for mediastinal nodal staging of lung cancer: a randomized trial. *JAMA.* 2010;304(20):2245-2252.

27. Korevaar DA, Crombag LM, Cohen JF, Spijker R, Bossuyt PM, Annema JT. Added value of combined endobronchial and oesophageal endosonography for mediastinal nodal staging in lung cancer: a systematic review and meta-analysis. *Lancet Respiratory Medicine.* 2016;4(12):960-968.

28. Navani N, Nankivell M, Lawrence DR, et al. Lung cancer diagnosis and staging with endobronchial ultrasound-guided transbronchial needle aspiration compared with conventional approaches: an open-label, pragmatic, randomised controlled trial. *The Lancet. Respiratory medicine.* 2015;3((4):):282-289.

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

Figure2: 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

Trial	Total patients randomised	Patients randomised to the control arm	Patients that provided clinical and pathological data	Accrual Period	Staging System (TNM)*	Staging Method	Surgical Protocol
MD Anderson(USA) 1994(11)	60	32	32	87-93	4	Chest x-ray	One or more positive nodal stations allowed. Patients with left lung tumors and paratracheal lymph node metastases excluded
MIP-91(France)(12, 29)	355	176	170	91-97	4	Chest x-ray, CT	Mediastinal node dissection and node sampling were left to the discretion of the surgeon
Netherlands 2000(13)	79	40	37	91-99	4	CT and mediastinoscopy	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.
JCOG 9209 (Japan)(14)	62	31	31	93-98	4	CT	Surgery was either lobectomy, bilobectomy, or pneumonectomy along with systematic mediastinal lymph node dissection.
Finland 2003(15)	62	32	23	95-99	4	CT	'Local surgery'
MRC LU22(UK)(5)	519	261	194	97-05	5/6	Bronchoscopy, mediastinoscopy and CT, PET	At cervical mediastinoscopy, the following lymph node stations will, wherever possible, be sampled: 2R, 2L, 4R, 4L, 7
SWOG S9900 (USA)(16)	354	174	170	99-04	5/6	Chest x-ray and CT	All accessible hilar (level 10) lymph nodes must be dissected …A complete

							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.
China 2002(17)	55	23	20	99-04	5/6	Chest x-ray, CT, bronchoscopy and abdominal ultrasound	Surgery consisted of radical lung resection and systematic mediastinal lymph node dissection
China 2005(18)	40	21	21	99-04	5/6	Chest x-ray, CT, bronchoscopy and abdominal ultrasound	Lobectomy or pneumonectomy with systematic lymph node dissection

* For details of TNM Staging systems, see Appendix 1

TNM stage	pI	pII	pIIIa	pIIIb	pIV	**Total**
cI	177 (25.4%)	72 (10.3%)	44 (6.3%)	22 (3.2%)	3 (0.4%)	**318 (45.6%)**
cII	40 (5.7%)	67 (9.6%)	32 (4.6%)	16 (2.3%)	5 (0.7%)	**160 (22.9%)**
cIIIa	32 (4.6%)	28 (4.0%)	116 (16.6%)	30 (4.3%)	12 (1.7%)	**218 (31.2%)**
cIIIb	0	0	0	2 (0.3%)	0	**2 (0.3%)**
cIV	0	0	0	0	0	**0**
Total	**249 (35.7%)**	**167 (23.9%)**	**192 (27.5%)**	**70 (10.0%)**	**20 (2.9%)**	**698 (100%)**

Clinically overstaged

Clinically understaged

T stage	pT1	pT2	pT3	pT4	**Total**
cT1	34 (4.9%)	16 (2.3%)	3 (0.4%)	7 (1.0%)	**60 (8.6%)**
cT2	35 (5.0%)	360 (51.6%)	69 (9.9%)	40 (5.7%)	**504 (72.2%)**
cT3	7 (1.0%)	42 (6.0%)	60 (8.6%)	23 (3.3%)	**132 (18.9%)**
cT4	0	0	0	2 (0.3%)	**2 (0.3%)**
Total	**76 (10.9%)**	**418 (59.9%)**	**132 (18.9%)**	**72 (10.3%)**	**698 (100%)**

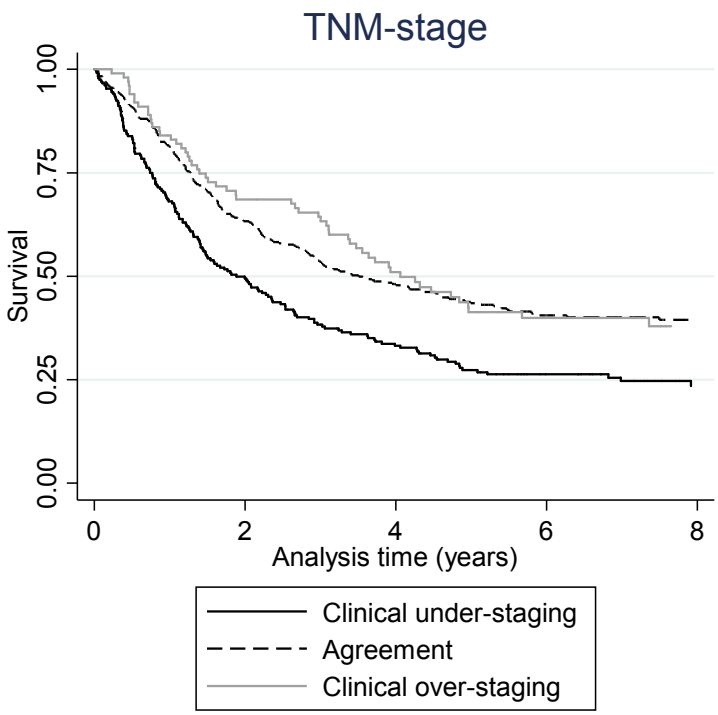
Clinically overstaged

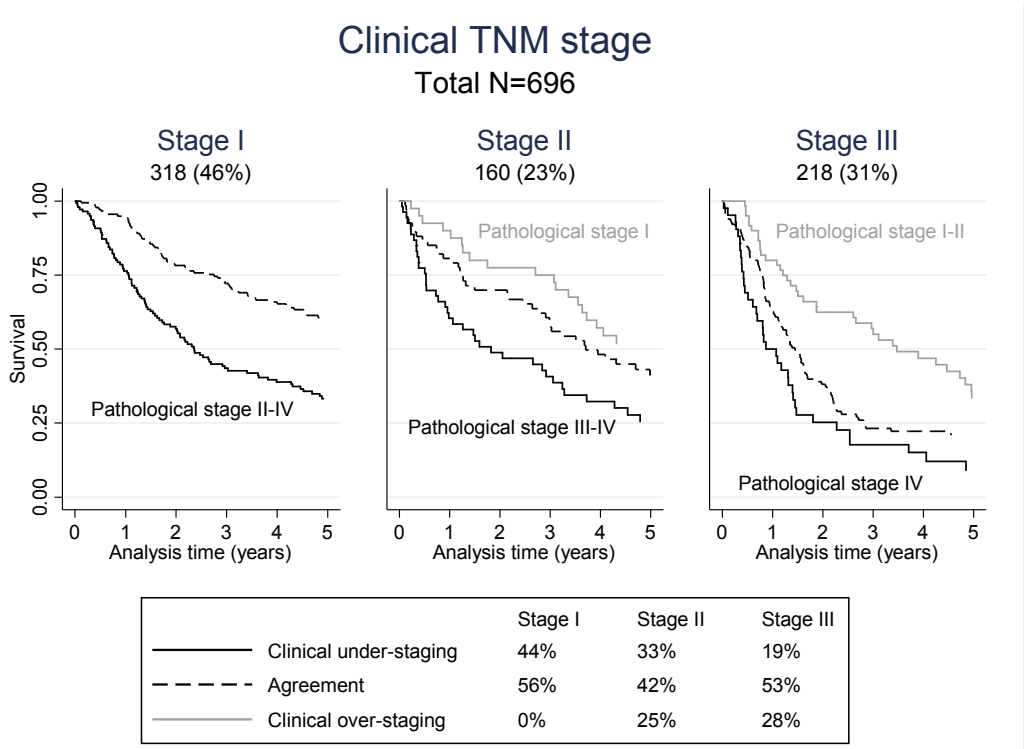
Clinically understaged

Nodal status	pN0	pN1	pN2	pN3	**Total**
cN0	259 (37.1%)	74 (10.6%)	57 (8.2%)	1 (0.1%)	**391 (56.0%)**
cN1	56 (8.0%)	67 (9.6%)	29 (4.2%)	0	**152 (21.8%)**
cN2	28 (4.0%)	19 (2.7%)	104 (14.9%)	4 (0.6%)	**155 (22.2%)**
cN3	0	0	0	0	**0**
Total	**343 (49.1%)**	**160 (22.9%)**	**190 (27.2%)**	**5 (0.7%)**	**698 (100%)**

Clinically overstaged

Clinically understaged





	Stage I	Stage II	Stage III
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

CHEST® Online Supplement

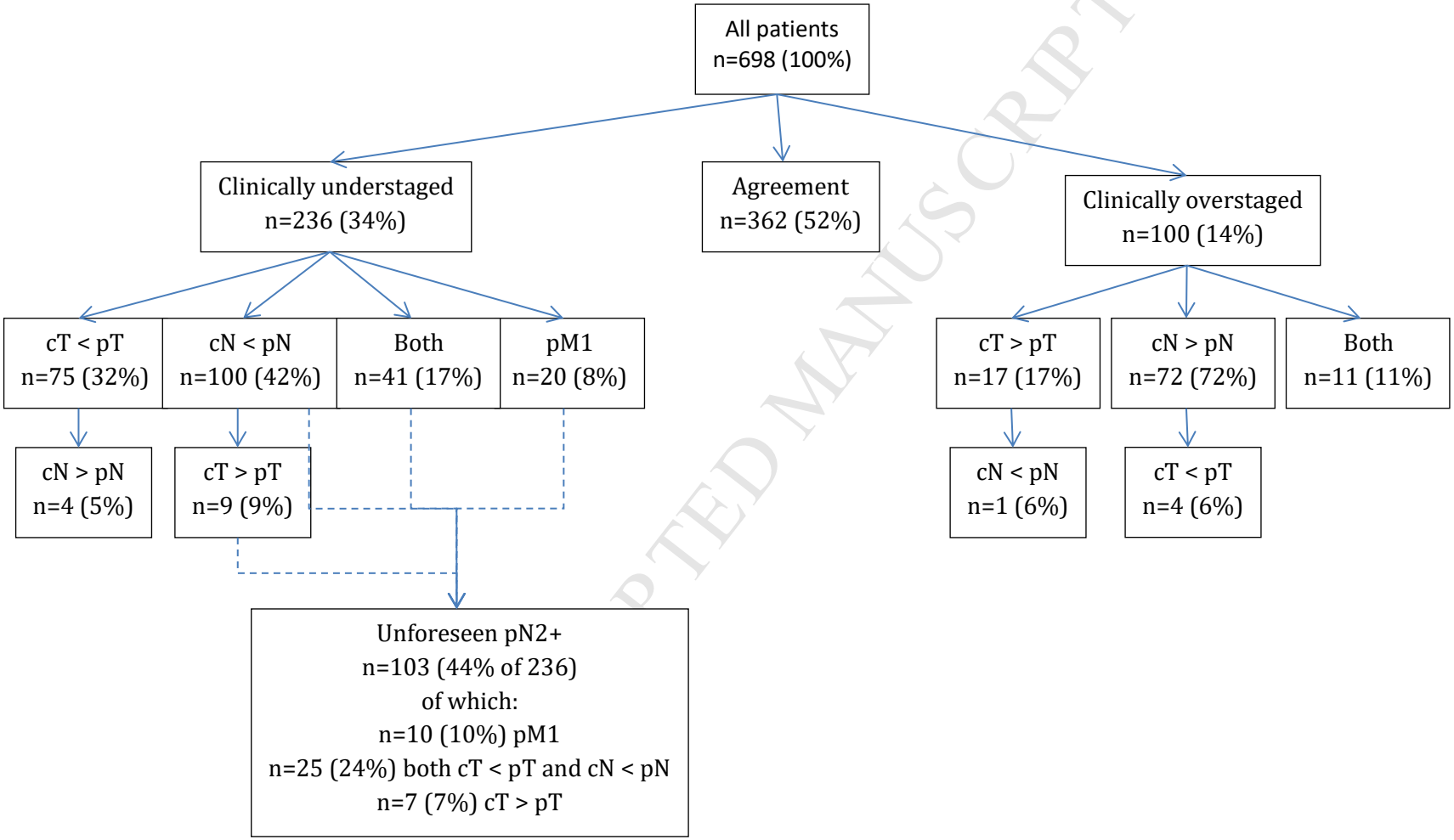
e-Table 1.

Comparison of TNM staging systems

4th edition (1987)(29)		5th edition (1997)(30), 6th edition (2002)(31)	
Stage I	T1,N0,M0 T2,N0,M0	Stage IA Stage IB	T1,N0,M0 T2,N0,M0
Stage II	T1,N1,M0 T2,N1,M0	Stage IIA Stage IIB	T1,N1,M0 T2,N1,M0 T3,N0,M0
Stage IIIA Stage IIIB	T1,N2,M0 T2,N2,M0 T3,N0/1/2,M0 anyT,N3,M0 T4,anyN,M0	Stage IIIA Stage IIIB	T1,N2,M0 T2,N2,M0 T3,N1/2,M0 anyT,N3,M0 T4,anyN,M0
Stage IV	anyT, anyN, M1	Stage IV	anyT,anyN,m1

CHEST® Online Supplement

e-Figure 1. Flowchart describing clinical and pathological agreement, clinical over staging and clinical under staging



CHEST® Online Supplement

e-Table 2.

Multivariate logistic regression; Factors that may predict staging agreement

Predictor	**TNM stage**	
	χ^2 (df)	**p-value**
Histology	0.40 (2)	0.82
Staging method	1.01 (1)	0.32
Age	2.48 (1)	0.12
Gender	0.24 (1)	0.62
Overall*	**4.22 (5)**	**0.52**

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

Sensitivity analysis with staging method replaced with year of accrual:

Predictor	**TNM stage**	
	χ^2 (df)	**p-value**
Histology	0.48 (2)	0.79
Year of randomisation	0.00 (1)	0.98
Age	2.55 (1)	0.11
Gender	0.19 (1)	0.66
Overall*	**3.21 (5)**	**0.67**