SURROGATE MARKERS PREDICTING OVERALL SURVIVAL FOR LUNG CANCER: ELCWP RECOMMENDATIONS.


Correspondence: Dr Thierry BERGHMANS  
Service des Soins Intensifs & Oncologie Thoracique  
Institut Jules Bordet  
Centre des Tumeurs de l’Université Libre de Bruxelles (ULB)  
1 rue Héger-Bordet, B-1000 Bruxelles, Belgium  
Tél (32) 2 5413185  
Fax (32) 2 5343756  
e-mail: thierry.berghmans@bordet.be
The present guidelines were formulated and approved by the members of the European Lung Cancer Working Party (ELCWP) as follows:

- T Berghmans, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- Y. Bonduelle, CH Peltzer-La Tourelle, Verviers, Belgique
- I. Cs Toth, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- C. Garcia, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- V. Giner, Hospital de Sagunto, Valence, Espagne
- S. Holbrechts, Hôpital Ambroise Paré, Mons, Belgique
- J.J. Lafitte, CHRU de Lille, Hôpital Albert Calmette, Lille, France
- J. Lecomte, CHU de Charleroi, Charleroi, Belgique
- I. Louviaux, CH Peltzer-La Tourelle, Verviers, Belgique
- E. Markiewicz, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- A.P. Meert, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- M. Paesmans, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- F. Pasleau, Université de Liège, Liège, Belgique
- M. Richez, CHR St-Joseph-Warquignies, Boussu, Belgique
- M. Roelandts, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- A. Scherpereel, CHRU de Lille, Hôpital Albert Calmette, Lille, France
- J.P. Sculier, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- Ch. Tulippe, Centre Hospitalier de Mouscron, Mouscron, Belgique
- P. Van Houtte, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- C. Wachters, Institut Jules Bordet, Université Libre de Bruxelles, Bruxelles, Belgique
- V. Westeel, CHU de Besançon, Besançon, France
- P. Van Schil, UZ Antwerpen, Antwerpen, Belgium
- J. Cadranel, Hôpital Tenon, AP-HP, and Université Pierre et Marie Curie, Paris, France
Summary

The present systematic review was performed under the auspices of the European Lung Cancer Working Party (ELCWP) in order to determine the role of early intermediate criteria (surrogate markers) in determining treatment efficacy instead of survival in patients with lung cancer. Preliminarily, the level of evidence for the use of overall survival to evaluate treatment efficacy was reviewed. Nine questions were then formulated by the ELCWP. After reviewing the literature with experts on these questions, it can be concluded that overall survival is still the best criterion for predicting treatment efficacy in lung cancer. Some intermediate criteria can be early predictors, if not surrogates, for survival despite limitations in their potential application: these include time-to-progression, progression-free survival, objective response, local control after radiotherapy, downstaging in locally advanced non-small cell lung cancer (NSCLC), complete resection and pathological TNM in resected NSCLC, and a few circulating markers. Other criteria assessed in these recommendations are not currently adequate surrogates of survival in lung cancer.

Keywords: guidelines, lung cancer, surrogate, survival, systematic review

Words number: 8826
**Introduction**

The European Lung Cancer Working Party (ELCWP) is a cooperative research group actively involved in the performance of academic clinical trials in thoracic oncology for more than 30 years. Published reports on trials, performed in various stages and with various histological types of lung cancer, can be found at the ELCWP website (www.elcwp.org). The group is also involved in evidence-based medicine and has published a number of meta-analyses and systematic reviews (1), as well as guidelines on the management of small (SCLC) and non-small cell lung cancer (NSCLC) (www.elcwp.org) (2-6).

Improvement in survival or cure rates is the main therapeutic goal in lung cancer management. Other criteria of efficacy are commonly used as primary endpoints in clinical studies and randomised trials, mainly concerning disease control (duration, response) and quality of life (QOL) or symptom control. During the last decade, new techniques have emerged that have had some impact on the therapeutic approach to lung cancer, such as integration of positron emission tomography into the staging strategy or use of tissue and blood biomarkers for predicting survival prognosis or treatment efficacy. The above-mentioned intermediate criteria have been proposed by some researchers for use as primary efficacy criteria in lung cancer. Nevertheless, it is currently not clear whether these criteria can correctly predict overall survival.

Intermediate criteria could provide some advantages by replacing final endpoints in clinical trials. They can occur more frequently or sooner than disease-specific mortality or overall survival. The term “surrogate” is generally used in an informal way to indicate that a biomarker or an intermediate endpoint can be used instead of the final endpoint of interest. A definition of surrogacy was proposed 20 years ago by Prentice (7), who has also defined operational criteria for validating the value of a candidate surrogate: 1°) a treatment should impact both on the candidate surrogate and on the true clinical endpoint; 2°) the treatment
effect on the surrogate should fully capture the treatment effect on the true endpoint. As this second condition is hard to prove and practically not suitable, other definitions have emerged. As mentioned by Buyse (8), a general consensus now exists that the validation of a surrogate endpoint can be carried out on the basis of a correlation approach: first, a correlation should be demonstrated at the individual level, that is, the surrogate endpoint has to be shown to be predictive of the true outcome; second, a correlation should be obtained at the trial level (a correlation that can be assessed through meta-analytical techniques). Nevertheless, showing surrogacy remains a difficult exercise. As we anticipated finding scarce demonstrations of surrogacy in the field of lung cancer, we searched for intermediate markers that are "simply" predictive/prognostic of the true outcome. Therefore, we avoid in the present report the use of the term surrogate/surrogacy and instead use the term "predictive", "prognostic", or "early" marker of treatment effect.

The following guidelines have been developed by the ELCWP in order to determine if intermediate criteria can be used as adequate predictive markers of overall survival in patients with lung cancer. The main target of our group is to help clinicians involved in the treatment of patients with lung cancer to integrate study results from the literature into their routine practice.

**Methodology**

A preliminary symposium held in Brussels with the participation of a panel of experts was dedicated to this topic (10th Annual Day of Thoracic Oncology, 21 March 2009, Brussels, Belgium). This led the ELCWP to make recommendations on the value of intermediate criteria for predicting survival in lung cancer. Preliminarily, the level of evidence for the use of overall survival to evaluate efficacy was reviewed. Furthermore, the guidelines aimed to answer nine questions: 1) Can the criteria of progression-free survival, time-to-progression, and disease-free survival be used to assess overall survival? 2) Is objective morphological
response to chemotherapy a valid efficacy criterion? Which evaluation criteria for response are to be used? 3) Are objective morphological response criteria and local control applicable to assess overall survival in the case of chemoradiotherapy (either sequential or concomitant)? 4) Can quality of life criteria be used to predict overall survival? Which quality of life scoring system should be applied? 5) Can downstaging after induction chemotherapy predict overall survival? Which assessment methodology has to be used? 6) Do complete resection and pathological TNM have a predictive role for survival? 7) Can metabolic response, assessed by \(^{18}\text{F}-\text{ fluorodeoxyglucose gathering on positron emission tomography (}\text{^{18}F-FDG-PET or PET-CT)}, predict the efficacy of chemotherapy? 8) Which is the place of tissue biological markers for the evaluation of overall survival? 9) Which is the place of circulating markers (molecules or cells) to assess overall survival? For each question, a definition of the intermediate criteria is provided.

The review of the literature was performed using the Ovid Medline system. This research was performed by a scientific librarian, experienced in searching for medical and scientific publications, and by physicians, experts in the treatment of thoracic neoplasms and trained in evidence-based medicine assisted by a biostatistician.

Ovid Medline was searched using the OvidSP interface. Unless otherwise stated, search terms were MeSH terms (Medical Subject Headings). MeSH terms were also combined with relevant free-text terms that were searched for in all of the fields containing text words, particularly in titles and abstracts. The PICO (population, intervention, comparator, outcome) model for clinical questions was used to isolate the concepts included in the question (9). The corresponding search criteria were translated into MeSH terms and free-text keywords (Annex 1). Completed search strategies included the “P” and “O” criteria combined with one of the possible intermediate markers listed in annex 1. Results were limited to publication years 1986 - March 2010, except when the demonstration of overall
survival required retrieval of old surgical series or randomised trials in SCLC. When the number of retrieved citations was too large, additional limits were applied according to publication types and levels of evidence, with randomised controlled trials and meta-analyses having the strongest levels of support. In a few cases, it was also necessary to reduce background noise by excluding studies focused on secondary lung cancers. Citations were exported from Medline into Reference Manager databases to allow the removal of duplicates and to facilitate the next selection by reviewers. Exported references were scanned for relevance. Articles were rejected on initial screening when titles, abstracts, and MeSH terms did not meet the inclusion criteria. The remaining articles were evaluated further for inclusion in the current review. This search was supplemented by screening the references of the selected articles and other literature known by the research team and the ELCWP members.

To be eligible, a study had to fulfil the following criteria: to be dedicated to the study of primary lung cancer, whatever the stage or the histology; to deal with tumour staging, at least for most of the patients included in the study, using the 4th, 5th, or 6th edition (10;11) of the International Staging System, except for small cell lung carcinoma; to assess the relationship between one of the intermediate criteria selected for the guidelines and survival, at least in univariate analysis, whether it was the primary objective of the publication or not; to be published in the English, French, German, or Dutch language literature, accessible to all co-authors. Abstracts were excluded as they cannot be expected to provide enough details to assess methodology and/or survival information. However, we allowed the consideration of unpublished studies if detailed results and methodology allowing adequate interpretation for the current guidelines were available, for example, an extensive oral presentation on the website of a congress.

The recommendations were graded according to the Grades of Recommendation Assessment, Development, and Evaluation (GRADE) system (12) (Table 1). The
methodology is summarised in Table 1. Selection of the publications was made according to this system. The best level of evidence was systematically looked for. The search was first limited to meta-analyses and randomised controlled trials; if none was available, we looked for case-control, cohort, or prospective studies; finally, case reports, case series, and expert opinion were considered. We graded the recommendations as strongly or weakly for or against using an intervention. When no data were available to answer a question, we decided that the recommendation was against the use of the intervention.

A draft was written and discussed during three workshops of the ELCWP (March 28 and October 16, 2009, and April 16, 2010) by all the co-authors who reviewed the evidence found in the literature. Thereafter, the guidelines were reviewed by the ELCWP during a meeting held in Brussels in November 2010 and approved by the panel of experts. The authors included medical oncologists, thoracic surgeons, pneumologists, radiation oncologists, and methodologists involved with the management of thoracic malignancies (target audience) as recommended by the AGREE collaboration group for the evaluation of guideline quality (www.agreecollaboration.org). Remarks and suggestions were then integrated into the final version of the guidelines. The level of evidence, quality of evidence, and grade of recommendation are listed in Table 8.

**Preliminary question: The use of overall survival to evaluate treatment efficacy**

According to the American Food and Drug Administration (13), survival is defined as the duration between registration or randomisation in a study and death from any cause (measured in intent-to-treat). Cure is more difficult to define. The World Health Organisation (WHO) proposed a definition based on survival duration, in the group of patients with the disease, in comparison to that observed in a general population with comparable demographic criteria but without the disease (14). Another option is to consider that the patient is cured
when he survives in good health for a sufficient length of time comparable to that of a group of patients with known low probability of recurrence.

The assessment of survival as a marker of treatment efficacy in lung cancer patients can be evaluated in five instances: resectable NSCLC, unresectable non-metastatic NSCLC, advanced and metastatic NSCLC, limited and extensive disease SCLC. This evaluation is based on the review of the literature performed by the ELCWP (2-6) (www.elcwp.org).

a. For obvious ethical reasons, no randomised trials have been performed comparing surgery alone to another medical treatment or no therapy in operable patients with resectable NSCLC. The level of evidence that surgery improves overall survival and curability is based on historical data from retrospective and prospective surgical series (15). In accordance with other guidelines, surgery is the treatment of choice for stage I and II diseases and remains an option for stage III NSCLC. Randomised trials and meta-analyses have demonstrated that adjuvant and, to a lesser extent, induction cisplatin-based chemotherapy - as well as UFT (Uracil-Tegafur) in Japanese patients - prolonged survival in comparison with surgery alone, at least in stage II and III NSCLC (16;17).

b. The evidence of the curative role of radiotherapy alone for unresectable non-metastatic NSCLC comes from control arms of randomised trials of treatment with irradiation alone or by extrapolation from series of inoperable patients with resectable NSCLC (stages I and II). A 5-year survival of 4-10% was documented (18;19). The addition of chemotherapy to curative radiotherapy, administered either sequentially or concomitantly, yielded improved survival and cure rates in comparison with those for radiotherapy alone. This evidence came from randomised trials and meta-analyses (20-25).

c. At least 15 randomised trials and 5 meta-analyses confirmed that survival improvement is obtained with first-line chemotherapy in comparison with best palliative care alone in advanced and metastatic NSCLC (20;26-29). Two randomised trials demonstrated
that salvage chemotherapy with docetaxel or erlotinib also improved survival in comparison with palliative care (30;31).

\(d\). The effect on survival of chemotherapy for first-line treatment of SCLC was first observed in old randomised trials comparing alkylating agents with inert compounds, which showed improved survival with cyclophosphamide monotherapy (32). Currently, multiple drug administration has become the standard chemotherapy. In this instance, two meta-analyses demonstrated that cisplatin- and/or etoposide-based chemotherapy is more effective and yielded better survival than other regimens (33;34). In limited disease SCLC, fifteen randomised trials and three meta-analyses demonstrated that the addition of radiotherapy to chemotherapy increased survival in comparison with chemotherapy alone (35-37). There has been only one randomised trial comparing topotecan with palliative care for salvage treatment, with improved survival with chemotherapy (38).

Overall survival has the advantage of being non-ambiguous (the patient is dead or alive). Assessing the distribution of survival duration in a group of patients makes use of probability theory and not simple mortality rate. The main disadvantages of this are the impact of subsequent treatment on overall survival and, in the case of long-term survival, the follow-up duration and the risk of loss to follow-up.

*An improvement in overall survival or rate of cure is widely accepted as an important goal of lung cancer therapy and is generally considered as the leading criterion of efficacy of a treatment in lung cancer patients*

Subsequent questions (these questions are not listed in any particular hierarchical order)

**Question 1:** Can the criteria of progression-free survival, time-to-progression, and disease-free survival be used to assess overall survival?
Progression-free survival (PFS) is defined as the duration between registration or randomisation in a study and the first objective aggravation of the disease or death. Time-to-progression (TTP) is the time from study enrolment to the first date of disease progression. Disease-free survival (DFS) is the period between registration or randomisation in a study and tumour recurrence or death. This criterion is mainly used in the adjuvant and surgical settings when no more target lesions are available (39).

To date, two meta-analyses have been published assessing the predictive value of TTP on overall survival (40;41) in lung cancer patients. A pooled analysis of selected randomised controlled trials was orally presented during the 2008 ASCO (American Society of Clinical Oncology) meeting (42). Although not published in full, it was selected for the present guidelines because detailed results and methodology were available on the ASCO website. The results of these three studies are reported in Table 2. The three studies dealt with advanced NSCLC treated with first-line chemotherapy in randomised controlled trials. In the two published studies (40;41), a statistically significant correlation between TTP and overall survival was detected, although it was of limited amplitude. In Hotta’s study (41), the median TTP ratio accounted for less than half the variability (41%) of the median survival ratio. To predict a statistically significant improvement in the survival time at the p value of 0.05, an increase in the median TTP of 1.8 months must be observed for trials with 750 patients, and 2.2 months and 3.3 months for those with 500 and 250 patients, respectively (40). Some evidence for a relationship between PFS and overall survival is provided in the last mentioned study (42): a 30% reduction in the risk of progression assessed by PFS predicts a statistically significant effect on overall survival in patients with advanced NSCLC treated with first-line docetaxel- or vincalcaloid-containing regimens. It must be pointed out that this positive association between improved PFS and further increase in survival was recently not
confirmed in randomised trials comparing chemotherapy to EGFR tyrosine kinase inhibitors (TKI), despite a high hazard ratio for PFS favouring the TKI treatment (43-45).

We did not find meta-analyses or specific studies with, as their primary objective, the assessment of PFS, TTP, or DFS as an intermediate marker for overall survival in resectable NSCLC treated by surgery with or without chemotherapy, in unresectable NSCLC treated by chemoradiotherapy, in NSCLC receiving salvage chemotherapy, and in SCLC.

PFS, TTP, and DFS have potential advantages in prospective studies by reducing the number of patients to be accrued, and the follow-up and study durations, as well as in alleviating the affect of subsequent treatment on the study endpoint. There are numerous disadvantages related to the parameters used: Which is the best criterion determining objective progression? How do you consider disagreements on evaluation criteria among investigators? How do you manage the absence of blinding in interpretation of results? Most importantly, all these parameters are highly dependent on the duration between two disease assessments and the type of assessment, which must be independent of the treatment arm.

Recommendations

1) **Time-to-progression is an intermediate marker for overall survival in advanced NSCLC treated with first-line chemotherapy**

Level of evidence: meta-analyses; quality of evidence: 3; grade of recommendation: Fs

2) **Progression-free survival is a potential intermediate marker for survival in the setting of docetaxel or vinca-alkaloid first-line regimens**

Level of evidence: meta-analyses; quality of evidence: 2; grade of recommendation: Fw

Statements

*More evidence is needed to confirm that TTP or PFS and disease-free survival are intermediate criteria predicting overall survival in the following settings: in patients with resectable NSCLC treated by surgery with or without chemotherapy, with unresectable*
NSCLC treated by chemoradiotherapy, with NSCLC treated by salvage chemotherapy, and with SCLC. Disease-free survival cannot be recommended as an intermediate marker for overall survival owing to the lack of data.

**Question 2: Is objective morphological response to chemotherapy a valid efficacy criterion? Which evaluation criteria for response are to be used?**

The first definition of objective response was proposed by Zubrod in 1960. Currently, three systems are available (Table 3). All are based on expert opinions without adequate statistical methodology applied for choosing the cut-off determining the different categories of response. The World Health Organisation (WHO) system was published in 1979 (14). The South West Oncology Group (SWOG) used its own definition, close to that of the WHO, in its studies (46). RECIST (Response Evaluation Criteria In Solid Tumours) was developed in 2000 to be used in clinical studies (47). The main differences between RECIST and WHO/SWOG criteria were the following: use of unidimensional instead of bidimensional measurement, limitation in the number of target lesions (maximum five per organ with an overall maximum of ten lesions), and different cut-offs for response and progression. A new version of RECIST has recently been published (48). The main modifications of version 1.1 are that a maximum of two lesions per organ are considered and that a complete response on lymph nodes is defined as a node no more than 10 mm in short axis.

Seven meta-analyses have been published to date that assess, as the primary objective of the study, the relationship between objective response to chemotherapy and survival (40;49-54). Objective response and/or disease control (objective response plus stable disease) predict improvement in overall survival in the following situations: advanced NSCLC receiving first-line or salvage chemotherapy and extensive disease SCLC treated by first-line chemotherapy (Table 4). In three meta-analyses, the impact of response on overall survival
was quantified. A statistically significant (p < 0.05) increase in overall survival could be observed for an increase in response rate of 18% requiring 750 patients to be included in randomised trials. This value increases to 21% and 30% if the number of patients is 500 or 250, respectively (40). In two other studies, each 1% increase in response rate resulted in an improvement in median survival time of 0.26 months in patients receiving tyrosine kinase inhibitor (54) or 0.07 months with salvage chemotherapy (51). When specified, the evaluation criteria for response were most often those of the WHO and RECIST. However, no formal comparison between the two systems has been performed for lung cancer patients. We also looked at large retrospective studies pooling prospective trials performed for lung cancer (55-58) (Table 5). Objective response was also found to be a significant predictor for better survival in advanced and metastatic NSCLC (55-57) and in SCLC (58), whatever the stage of the disease, treated by chemotherapy.

No meta-analyses or randomised trials having the evaluation of response as an intermediate criterion for survival were available in the following clinical settings: induction chemotherapy for operable NSCLC, sequential or concomitant chemoradiotherapy for locoregionally advanced NSCLC, and limited disease and recurrent SCLC. Better survival was found in patients with NSCLC presenting with an objective response to induction chemotherapy before surgery or to treatment for locoregionally advanced NSCLC, induction chemotherapy before radiotherapy, chemoradiotherapy, or radiotherapy alone (59-64). The level of evidence was from analyses of secondary objectives of randomised trials published on each topic (Table 5) and phase II studies (65-73). We found only one study in SCLC showing that patients with complete response had longer survival than those with partial response or no change, those progressing under first-line treatment with chemotherapy or chemoradiotherapy having the worst prognosis (74).
There are few formal comparisons allowing determination of which evaluation criteria are to be used. In the RECIST publication (47), a comparison between RECIST 1.0 and WHO criteria did not find any significant difference for assessing objective response or progression, whether considering the whole population or only lung cancer patients. Further studies have been performed in patients with lung cancer, which are summarised in Table 6. Overall, there was no significant difference between the two systems of evaluation for response determination (47;75-79).

Recommendations

1) **Objective response is an intermediate criterion for overall survival in advanced NSCLC treated with first-line or salvage chemotherapy**

   Level of evidence: meta-analyses and randomised trials; quality of evidence: 4; grade of recommendation: Fs

2) **Objective response is an intermediate criterion for overall survival in extensive disease SCLC treated with first-line chemotherapy**

   Level of evidence: meta-analyses and retrospective studies; quality of evidence: 2; grade of recommendation: Fw

3) **Objective response is a potential intermediate criterion for overall survival in operable NSCLC treated with induction chemotherapy and in locoregionally advanced NSCLC treated by chemotherapy and radiotherapy**

   Level of evidence: retrospective studies; quality of evidence: 2; grade of recommendation: Fw

Statements

More evidence is needed to confirm that objective response is an intermediate criterion predicting overall survival in the following settings: operable NSCLC treated with induction chemotherapy, locoregionally advanced NSCLC and limited disease SCLC treated by chemotherapy and radiotherapy.
WHO and RECIST 1.0 criteria appear to have the same efficacy in predicting objective response in locoregionally advanced and metastatic NSCLC. On the basis of the available data, it is not possible to determine the best response criterion to be used in other settings related to lung cancer management, the equivalence of WHO and RECIST criteria having to be confirmed in larger studies.

The cut-off point to define response/progression, as proposed by the WHO and the RECIST systems, needs validation assessment.

In the context of maintenance therapy, disease control rate (response plus stable disease) needs validation assessment.

**Question 3: Are objective morphological and metabolic response criteria and local control applicable to assess overall survival in the case of chemo-radiotherapy (either sequential or concomitant)?**

There are many problems when defining response by conventional criteria (WHO and/or RECIST) due to pneumonitis and fibrosis modifications induced by thoracic irradiation, which must be differentiated from residual neoplastic tissue. As these changes are time-related and increase by the use of radio-sensitising chemotherapeutic agents, timing of evaluation after treatment could influence response assessment.

The role of new evaluation techniques like PET(-CT) was assessed in a few small retrospective studies. In 73 NSCLC patients treated by radical radiotherapy or chemoradiotherapy, there was poor agreement between PET and CT response assessment (Kappa 0.35) using respectively EORTC and WHO criteria. In multivariate analysis, only PET response was associated with survival (p < 0.0001) (80). In a second study including 50 patients treated with induction chemoradiotherapy, followed by surgery in 37 cases, change in standardized uptake value (SUV) on PET-CT after induction treatment was not related to survival (p = 0.75) (81). Lastly, SUVmax after induction chemoradiotherapy predicted
survival in operated patients with NSCLC. The median survival was longer for the patients with SUVmax < 4 than in those with SUV more than 4, namely, 56 months versus 19 months (p < 0.001) (82).

Local control is another commonly used approach to undertake intrathoracic response assessment. In a consensus expert conference held in Bruges in 1993, the proposals for defining lung cancer local control after thoracic irradiation were complete disappearance of all radiographic abnormalities by chest film CT and residual radiographic abnormality assessed by chest CT at 3 and 6 months, which then remains stable for an additional 6 months or more. This definition has never been the topic of a validation study.

Literature on this question is relatively disappointing. We found only a few retrospective studies or secondary analyses of randomised trials assessing the association between local control and survival in lung cancer. Birch et al. observed for limited disease SCLC that survival was longer for patients achieving local control after chemoradiotherapy than for those who did not (83). In a small retrospective study including limited and extensive disease SCLC, Cox et al. had the same observation (84). Dosoretz et al. showed that local control resulted in improved survival in irradiated medically inoperable NSCLC (85). In a retrospective study of RTOG randomised trials, local control was associated with significantly better survival (86). However, the ELCWP could not confirm these results in a phase III trial comparing chemotherapy to chemotherapy followed by radiotherapy in stage III unresectable NSCLC (87). A significant increase in local control duration was obtained in the combination arm: median of 158 weeks versus 31 weeks, with respective two-year rates of 57% and 24% (p = 0.0007). This better local control duration did not result in a statistically significant increase in survival duration: median of 54 weeks (95% CI: 43-73) versus 42 weeks (95% CI: 35-51), with two-year survival rates of 22% (95% CI: 11-33) and 18% (95% CI: 8-28).

**Recommendations**
1) Although commonly used, conventional criteria for response assessment of primary lung
tumour treated by (chemo)radiotherapy cannot be used as intermediate criteria for survival.
Level of evidence: retrospective studies; quality of evidence: 1; grade of recommendation: Aw

2) Local control, for which the definition has to be clarified, is a possible intermediate
criterion for overall survival.
Level of evidence: retrospective studies; quality of evidence: 1; grade of recommendation: Fw

Statement
The roles of conventional WHO or RECIST criteria, of metabolic response criteria (PET
imaging), and of local control as intermediate criteria in the context of lung cancer treated by
(chemo)radiotherapy have to be assessed in prospective studies.

Question 4: Can quality of life criteria be used to predict overall survival? Which
quality of life scoring system has to be applied?

We identified some data in the literature related to this topic, although there are many
papers suggesting a prognostic role for baseline QOL data (88-90). Several scoring systems
exist and have been validated with the availability of modules dedicated to lung cancer
symptoms and treatment side effects. Those which have been studied in association with the
predictive value of QOL on other outcomes (response to chemotherapy, progression-free
survival, survival) are the EORTC quality of life core questionnaire (30 items) complemented
by the lung cancer module LC13 (91;92) and the Functional Assessment of Cancer Therapy –
Lung (FACT-L) with a module specific for the symptoms of the disease (LCSS) (93). Before
looking at this predictive value, it is relevant to define what is meant by a response at the
QOL level. Osoba (94) defined an increase in the overall EORTC QOL score of more than 10
points as a response, a decrease of more than 10 points as a progression, and intermediate
results as stabilisation, on the basis of concordance between QOL score evolution and a
subjective question. Cella (95) has proposed a definition of response for two outcomes for the LCSS score (from 0 to 28, with increasing values for increasing QOL) and the TOI (Trial Outcome Index, which is the sum of LCSS and functional and physical wellbeing scales, from 0 to 84). For the first outcome, an increase of 2-3 points is needed and, for the TOI, an increase of 5 to 7 points is required. Definitions of stabilisation and progression are derived from these thresholds. This proposal resulted from the analysis of a clinical trial and used the association between QOL evolution and response to chemotherapy or PFS.

One report on a meta-analysis (96) suggests that an association between QOL and radiological response exists (positive correlation), but this meta-analysis was not dedicated to lung cancer and suffered from a serious selection bias: out of more than 350 studies identified by the literature search, only 21 could be included in the quantitative review.

Eton (97) has shown, for patients with advanced lung cancer, using the database of a clinical trial and the FACT-L scoring system, that classification into 4 groups (I: high baseline FACT-L score with improvement; II: high baseline with deterioration; III: low baseline with improvement; IV: low baseline with deterioration) has prognostic value for time-to-progression and overall survival. Group I had the best results, followed by group II, group III, and group IV (p < 0.001).

Cella (98) studied the QOL evolution among patients treated in the IDEAL 2 trial testing gefitinib as second-line treatment for NSCLC. The patients with a QOL response (> 2 point improvement on the LCSS scale) had a significantly prolonged survival compared with the other patients (median of 13 months compared with 5 months after a landmark of 8 weeks, p<0.001). Data from IDEAL 1 point in the same direction (99). In the INTACT trials looking at the addition of gefitinib to a chemotherapy regimen in the first-line setting, the association between QOL and antitumoral response has been studied using the EORTC QOL questionnaire. The associations with different scales (overall scale, physical scale, pain,
cough, dyspnea) and response were all significant. However, we do not have any available data about the value of QOL response for progression-free survival or overall survival (100;101).

**Recommendation**

*Response at the QOL level cannot be recommended as an intermediate criterion for overall survival owing to a lack of robust data.*

Level of evidence: retrospective studies; quality of evidence: 1; grade of recommendation: Aw

**Statement**

*The association between response at the QOL level and long-term outcomes, as suggested in some instances (recurrent NSCLC treated with gefitinib), needs further investigation including adequate determination of cut-off points defining a response and of the most appropriate questionnaire.*

**Question 5: Can downstaging after induction chemo(radio)therapy predict overall survival? Which assessment methodology has to be used?**

Downstaging can be defined as a shift to a lower disease stage after a treatment. In NSCLC, studies assessing the prognostic role of downstaging on survival after induction chemotherapy have been performed at the mediastinal (N stage) or the TN level. The literature is generally heterogeneous. Inclusion criteria are quite different from one study to another according to stage, mediastinal evaluation (clinical or pathological definition), type of treatment (induction chemotherapy or chemoradiotherapy before surgery; non-surgical treatment only), and type of downstaging (N or TN). Most of the publications are case series or retrospective analyses of phase II studies of small sample size. We found only two phase III trials in which these questions were retrospectively analysed. The first trial compared the role of surgery to radiotherapy in stage IIIA pathological N2 NSCLC treated by induction
chemotherapy (102). In the surgery arm including 154 randomised patients, mediastinal downstaging was an independent prognostic factor for survival in multivariate analysis (p = 0.04). In the second trial (103), comparing chemoradiotherapy alone to chemoradiotherapy followed by surgery in stage III pathological N2 NSCLC, mediastinal downstaging in the surgical group (164 patients) was associated with better survival in univariate analysis with median survival times for pathological N0 and N1-3 of 34.4 months and 26.4 months, respectively (p < 0.0001).

Mediastinal downstaging in NSCLC has been assessed by non-invasive (CT, PET scan), minimally invasive (oesophageal [EUS] or endobronchial [EBUS] endoscopic procedures), and invasive procedures (re-mediastinoscopy, VATS). In its 2007 guidelines, the European Society of Thoracic Surgeons (ESTS) (104) did not recommend a particular procedure for restaging but recommended that cyto-histological proof be obtained by a minimally invasive or invasive procedure depending on the availability and specific expertise of staff at the centre. The literature on mediastinal restaging was recently reviewed (105). Sensitivity, specificity, false-negative, and false-positive rates for mediastinal CT restaging were on average 63%, 70%, 31%, and 34%, respectively. Respective values for PET scan were 63%, 70%, 26%, and 34%. Combined PET-CT seems more adequate, with sensitivity of 62-77%, specificity of 88-92%, false-negative rate of 20-25%, false-positive rate of 7-25%, and an accuracy of 79-83% (106;107). Few studies have assessed the role of minimally invasive procedures for mediastinal re-staging. EUS was evaluated in three very small studies, EBUS in one, and TBNA (transbronchial needle aspiration) in another one. Average sensitivity and false-negative rates were 84% and 14%, respectively. Re-mediastinoscopy is technically challenging due to adherence and fibrosis. In a few studies, the sensitivity of this technique was 63%, with 100% specificity and accuracy of 84-93%. False-negative rate was 22%. VATS was assessed in a small series of 70 patients (108). The procedure was
unsuccessful in 17 patients owing to fibrosis. Reported sensitivity, specificity, and negative predictive value were 75%, 100%, and 76%, respectively.

Recommendation

*Although not strictly demonstrated, there are numerous assertions from subgroup analyses of randomised trials and retrospective studies that mediastinal downstaging and, to a lesser extent, TN downstaging are associated with better survival in locally advanced NSCLC treated by induction chemotherapy or chemoradiotherapy before surgery.*

Level of evidence: retrospective studies; quality of evidence: 2; grade of recommendation: Fw

Statements

*The association between downstaging after induction chemotherapy (or chemoradiotherapy) and long-term outcomes should be confirmed and needs further investigation in adequately powered prospective studies.*

*PET-CT is the recommended non-invasive technique for mediastinal restaging, being more precise than CT scan, which appears insufficiently accurate in this setting. When pre-operative pathological persistent mediastinal neoplastic infiltration has to be confirmed, minimally invasive or invasive techniques are feasible, each one appearing with similar efficacy. The choice of the technique will depend on the local expertise and the initial technique used for demonstrating mediastinal invasion.*

**Question 6: Do complete resection and pathological TNM have a predictive role for survival?**

The basis of the definition of complete resection is the Union Internationale Contre le Cancer (UICC) residual tumour classification (R classification) (109). R0 category consists of the absence of residual tumour in the primary site, lymph nodes, and distant sites, which appears insufficient in lung cancer, and is mainly for the mediastinal definition of complete resection. The Complete Resection subcommittee of the International Association for the
Study of Lung Cancer (IASLC) Staging Committee, after an extensive review of the literature, proposed the definition of complete resection based on an international consensus (110): “Complete resection requires all of the following: free resection margins proved microscopically; systematic nodal dissection or lobe-specific systematic nodal dissection; no extracapsular nodal extension of the tumour; and the highest mediastinal node removed must be negative. Whenever there is involvement of resection margins, extracapsular nodal extension, unremoved positive lymph nodes, or positive pleural or pericardial effusions, the resection is defined as incomplete. When the resection margins are free and no residual tumour is left, but the resection does not fulfil the criteria for complete resection, there is carcinoma in situ at the bronchial margin, or positive pleural lavage cytology, the term uncertain resection is proposed”. R1 is defined as a persistence of microscopic tumour tissue and R2 as macroscopic incomplete resection.

The level of evidence of the prognostic role of R0 in NSCLC is relatively low. There have been many studies derived from retrospective analyses of phase II and surgical series, mainly including stage IIIA-N2 NSCLC. In most of them, R0 was a predictive factor for survival in either univariate or multivariate analysis. We found only one phase III trial in which R0 was evaluated in multivariate analysis as a secondary endpoint (102). In 154 stage IIIAN2 NSCLC patients receiving induction chemotherapy followed by surgery, R0 was an independent prognostic factor for survival with median survival times of 24 (R0) versus 12 months (R1/R2). For a few large series, including pathological stage I-IV NSCLC patients, incomplete resection was reported as a poor prognostic factor for survival in multivariate analysis. Involved resection margins were associated with an increased risk of death (HR 1.49; p < 0.001) among 3211 patients from a Norwegian registry (111). In another surgical series of 836 patients, an HR of 8.2 (p < 0.001) was observed in R1/R2 in comparison with
R0 cases (112). Doubling or tripling of the risk of death was observed in a series of 446 patients (p < 0.01) (113).

The IASLC reported the largest recent studies, assessing the prognostic role of pathologic TNM in 9137 resected NSCLC. In multivariate analysis, pathological stage using the 7th version proposed by the IASLC was a significant prognostic factor for survival (p < 0.0001) (114). In large series or registry studies of more than 500 patients, pTNM (111;115-117) and, when analysed separately, pT or pN (112;118;119), were found to be independently associated with survival. Lastly, pTNM, pT, or pN was found as a significant predictor for survival in six randomised trials assessing the role of adjuvant chemotherapy after surgery and considering pathological stage in analyses of prognostic factors (120-125). In these studies, the 5th version of the International Union Against Cancer (UICC) and American Joint Committee on Cancer Staging system (AJCC) was most often used, although the 4th version was applied in three studies.

Recommendations

1) Complete resection is a prognostic factor for survival in resected NSCLC and can be used as an intermediate criterion for overall survival.

Level of evidence: cohort and retrospective studies; quality of evidence: 3; grade of recommendation: Fs

2) Pathological TNM is a prognostic factor for survival in resected NSCLC and can be used as an intermediate criterion for overall survival.

Level of evidence: cohort studies; quality of evidence: 3; grade of recommendation: Fs

Statement

The definition of complete resection proposed by the Complete Resection subcommittee of the IASLC is recommended.
Question 7: Can metabolic response, assessed by $^{18}$F-fluorodeoxyglucose gathering on positron emission tomography ($^{18}$F-FDG-PET or PET-CT), predict the efficacy of chemotherapy?

In a systematic review of the literature with a meta-analysis (126;127), primary tumour SUVmax measured on $^{18}$F-FDG-PET has been shown to have prognostic value in NSCLC. This observation needs to be confirmed by a multivariate analysis taking into account the known prognostic factors; nevertheless, it strongly suggests that PET may be more useful than only providing imaging.

PET has been proposed for the assessment of response to chemotherapy. Criteria were defined in 1999 by the European Organization for Research and Treatment of Cancer (EORTC) PET Study Group (128). A complete metabolic response is complete resolution of $[18F]$-FDG uptake within the tumour volume so that it is indistinguishable from surrounding normal tissue. A partial metabolic response consists of the reduction of a minimum of 15% in tumour SUV after one cycle of chemotherapy, and greater than 25% after more than one treatment cycle. A reduction in the extent of the tumour uptake is not a requirement for partial metabolic response. A stable metabolic disease is an increase in tumour SUV < 25% or a decrease < 15% and no visible increase in the extent of tumour uptake (20% in the longest dimension). A progressive metabolic disease consists of an increase in tumour SUV > 25% within the tumour region defined on the baseline scan, visible increase in the extent of tumour uptake (20% in the longest dimension), or appearance of new FDG uptake in metastatic lesions. For the experts, 25% was found to be a useful cut-off point, but there was a need for reproducibility analysis to determine the appropriate cut-off for statistical significance. Ten years later, that analysis has yet to be performed.

Data on lung cancer in this context are limited and heterogeneous. A few small studies have assessed the predictive effect of metabolic response for further survival. In a series of 57
patients with advanced NSCLC treated by chemotherapy, changes in SUV have been shown to have relatively similar adequacy to changes in tumour/muscle (t/m) ratio and to be better than changes in net-influx constants (129). In another series of 51 NSCLC patients (15 treated by induction chemotherapy and 31 by palliative chemotherapy), metabolic response assessed by the EORTC criteria was not associated with a significant difference in survival, while it became significant when patients were divided into two groups according to the median change in SUV (130). In a group of 30 patients with stage IIIAN2 NSCLC treated by induction chemotherapy, the best cut-off in decrease of SUVmax to predict 5-year survival was 60%; a decrease of 25% was not discriminatory (131). In a systematic review of the literature about the prediction of histopathological response and further survival in stage III disease treated by induction treatment, De-Geus-Oei et al. showed the great heterogeneity of studies in terms of cut-off of SUV used and definition of pathologic response (132). In the series reported above (131), Dooms et al. defined a group with good prognosis by combining pathological (pN0 or pN2 with < 10% cancer cells in mediastinal lymph nodes) and metabolic (> 60% decrease of SUV max in the primary tumour) criteria. Finally, complete (but not partial) metabolic response has been associated with better survival in a series of 31 stage III NSCLC patients (133).

The predictive effect of initial SUVmax of the primary tumour for morphological response to chemotherapy has been the topic of one paper (134) on advanced NSCLC (87 patients). SUVmax (dichotomised by the median) was the single significant predictor in univariate and multivariate analyses, but the patients with high SUVmax had shorter duration of response and there was no overall difference in survival between patients with high and low SUVmax.
The early assessment of the metabolic response after the first course of chemotherapy has been investigated in two small studies, suggesting some potential utility for predicting further survival (135;136).

Studies of the prediction of morphological response by the metabolic response have been described in two reports. In the first one (129), the metabolic response was shown to have high predictive value (96%), meaning that, if PET scan does not show a response, it is very unlikely for one to be documented by morphological tests. The positive predictive value is however not very good (71%). In the second study (72), performed on 89 patients, the correlation between metabolic response (either visual or based on SUVmax) and morphological response was poor and the SUVmax measured after chemotherapy did not predict further survival. We also have to take into account potential methodological problems due to the lack of standardisation.

**Recommendation**

*Metabolic response assessed by PET scan should not be used for the routine assessment of response to treatment in lung cancer patients in place of morphological criteria.*

Level of evidence: retrospective studies and case series; quality of evidence: 1; grade of recommendation: Aw

**Statements**

*The cut-off point of 25% to define response/progression, as proposed by the EORTC for metabolic response, needs validation assessment.*

*Metabolic response requires further investigation with strict methodology in terms of PET scan examination standardisation, morphological response assessment, and outcome definition.*

**Question 8: Which is the place of tissue biological markers for the evaluation of overall survival?**
Numerous biomarkers have been tested in lung cancer for prognostic purposes or for predicting response to chemotherapy or radiotherapy. Determination of the presence or absence of a tissue biomarker can be performed by various techniques (immunohistochemistry, PCR, FISH, high-throughput techniques, etc.) that assess cancer biology at different levels: protein expression, RNA, or DNA abnormalities. The question is whether differential expression of tissue biomarker during treatment, before and after chemotherapy for example, could reliably predict treatment efficacy and overall survival.

Many studies on tissue biomarkers in lung cancer have been published. They looked at the prognostic role for survival of these biological markers, usually obtained at initial diagnosis, mainly in NSCLC surgical series. To a lesser extent, their predictive role for treatment sensitivity was studied. However, we found only one published study assessing the relationship between the modifications of tissue biomarkers during therapy and overall survival (137). Fifty-four patients with stage III NSCLC were treated by induction chemotherapy followed by concomitant chemoradiotherapy, and surgery if the tumour became resectable. There was no difference in the apoptotic indices before and after neoadjuvant treatment and no statistically significant impact on survival was observed.

One potential explanation of the lack of studies is related to the difficulties in easily obtaining adequate tissue samples in lung cancer patients. This problem is yet more pronounced when a good response is observed and no surgery is planned.

**Recommendation**

*Tissue biological markers have not to be used for evaluation of treatment efficacy and are not adequate intermediate criteria for overall survival in lung cancer patients.*

Level of evidence: case series; quality of evidence: 1; grade of recommendation: Aw

**Statements**
The role of tissue biological markers in assessing overall survival in lung cancer needs further investigation with adequate methodology, that is, in terms of positivity definition, and standardisation in the evaluation methods and statistical methodology (multivariate analysis taking into account known prognostic factors).

Question 9: Which is the place for circulating markers (molecules or cells) to assess overall survival?

We aimed to determine if modification in the levels of circulating markers before and after therapeutic procedures in lung cancer patients can reliably predict overall survival. We selected the markers as described in the methodology section. We also specifically searched for biological markers known as potential prognostic factors in either NSCLC or SCLC: LDH, alkaline phosphatases, white cell and neutrophil counts, platelet count, haemoglobinemia, natremia, calcemia, bilirubinemia, cyfra 21-1, CEA, NSE, pro-GRP (gastrin-releasing peptide), and other cancer markers like CA19-9 and CA125. Publications were eligible if the correlation between the marker and overall survival was assessed.

Although the prognostic role for survival of the selected variables taken at initial diagnosis has been extensively studied, few studies, summarised in Table 7, have been published that assess the prognostic role on survival of their expression before and after treatment (138-151). Criteria used by the authors to define response based on the levels of circulating markers were variable, allowing only indirect comparisons between studies. It must be pointed out that the majority of these studies were retrospective.

Post-therapy CEA (carcinoembryonic antigen) normalisation or significant decrease seems to be related to better survival in early stage NSCLC treated by surgery (138;140;141;152), in advanced NSCLC with chemotherapy (139;142), and after salvage gefitinib in relapsing NSCLC (143). Cyfra 21-1 decrease has been reported to be significantly associated with improved survival in two studies (139;142) while no association was observed
in a third one (144). For other factors, it could be suggested from small studies that decreases in the following in terms of blood levels were predictive of better survival: pro-GRP in SCLC, CA 19-9 in relapsing NSCLC, CA125 in NSCLC, and NSE in NSCLC and SCLC (142;143;145-147). Anemia occurring during chemotherapy or chemoradiotherapy has not been demonstrated to have prognostic significance in two larger studies on NSCLC and SCLC (148;151). Lastly, the persistence of circulating tumour cells in blood was of prognostic significance in locoregional NSCLC treated by chemoradiotherapy and in relapsing NSCLC treated with gefitinib (149;150;153).

**Recommendations**

1) **Some circulating markers (CEA, Cyfra 21-1, pro-GRP and, to a lesser extent, NSE, CA-125, and CA19-9), used as single criterion to assess overall survival, could be adequate intermediate criteria for survival in lung cancer patients.**

   Level of evidence: retrospective studies; quality of evidence: 2; grade of recommendation: Fw

2) **The persistence of circulating tumour cells in NSCLC may have a prognostic impact on further survival.**

   Level of evidence: cohort studies; quality of evidence: 1; grade of recommendation: Fw

**Statement**

*The suggested role of circulating markers in assessing overall survival in lung cancer needs further prospective investigation with adequate methodology (including the cut-off determination), that is, in terms of statistical methodology (multivariate analysis taking into account known prognostic factors)*.
**Conclusions**

Overall survival, or healing for curable disease, is the leading criterion of efficacy of a treatment in lung cancer patients. Intermediate criteria are potential or adequate surrogates for survival, despite some limitations: time-to-progression, progression-free survival, objective response, local chest control after radiotherapy, mediastinal and TNM downstaging in locally advanced NSCLC after induction therapy before surgery, complete resection and pathological TNM in resected NSCLC, and a few circulating markers (CEA, Cyfra 21-1, proGRP and, to a lesser extent, NSE, CA-125, and CA19-9). At the present time, some criteria cannot be used for predicting overall survival in lung cancer (and therefore certainly not as surrogates for overall survival): conventional criteria for radiotherapy response assessment of primary lung tumour, quality of life evolution under treatment, metabolic response assessment by FDG-PET scan, and tissue biological markers. The recommendations in this study highlight the need for well-designed prospective studies in many instances where intermediate criteria are of possible interest.
Potential limitations and problems

1. Literature biases

Firstly, we searched for prospective studies and meta-analyses looking, as the primary objective, at the association between intermediate criteria and overall survival. This approach was recently discussed by Buyse et al. to demonstrate a possible association between a surrogate and a clinical endpoint, requiring data from multiple randomised trials or meta-analyses (154). In most instances considered in these guidelines, this literature was lacking. Secondly, we searched for secondary endpoints of prospective studies and retrospective series. Although we used multiple key words for each topic and an extensive literature search system (Ovid), it is possible that some studies and series in which some of the selected criteria were actually reported as secondary endpoints were not retrieved. However, the impact of this literature bias on the conclusions of these guidelines seems limited as the level of evidence of this type of study is relatively weak (low or very low grade of evidence according to the GRADE system). Furthermore, the inclusion of non-randomised studies should induce biases in the interpretation of the literature results for assessing association between criteria like determination of objective response and survival. Another potential source of bias is related to the considered languages. We only took into account reports written in English, French, German, and Dutch languages that were comprehensible by all of the authors. It must be underlined that most of the results came from clinical trials in which patients were strictly selected and do not fully represent the population treated in routine practice.

2. Validation studies

Although we found some statistical associations between intermediate criteria and survival in meta-analyses and small-sized randomised studies, validation studies are required before definite conclusions can be drawn in most instances. These studies need to be prospectively designed with the primary objective being the determination of the relationship
between the considered surrogate and the true endpoint. They have to consider the clinical value of the association when the effect of the intermediate criterion in predicting survival accounts for only a limited part of the variability of the survival ratio, as we observed with TTP (40;41). The potential surrogate may have a substantial noise component due to issues in defining the intermediate criterion or other factors that may affect the clinical outcome. Therefore, there are some issues concerning the application of a surrogate in situations other than those in which the surrogacy was asserted. These require knowledge about the relevant biological variables whereby the treatment or the considered populations may affect the outcome of primary interest.

3. Grades of recommendations

In order to grade the guidelines, the ELCWP has chosen the GRADE system. This internationally accepted grading system for recommendations has previously been used for lung cancer by the American College of Chest Physicians (ACCP) (155). GRADE has certain advantages. It allows grading of recommendations based on the level of scientific evidence, which is further modulated by the quality of the literature (Table 1). It also allows integration into the grade of the recommendations of some subjective variables like the balance between desirable and undesirable effects, and the values and preferences or the costs (resource allocation) by guideline writers, which may be of importance in routine practice but cannot always be derived from the literature.

4. Impact of supplementary treatment on overall survival

We selected studies assessing a possible relationship between a surrogate and overall survival. Preceding or subsequent therapeutics after the investigated treatment can impact on overall survival and potentially on the relationship with the intermediate criteria. Unfortunately, in light of the currently available data, it was not possible to avoid this potential bias in our literature interpretation.
**Perspectives**

We found that validation of intermediate criteria as surrogate endpoints in lung cancer has not really been achieved; furthermore, such a validation may be dependent on the treatment category considered and the population of patients included in the trial. The development of new therapies leads us to consider the use of intermediate criteria to assess their effectiveness in order to achieve possible reductions in sample size, study duration, or cost. A further set of intermediate outcomes based on data from gene expression or proteomic platforms will be proposed in the near future. In this context, the identification of a set of intermediate outcomes may be important, taking into account the fact that biological pathways could influence the beneficial or harmful effects of the treatment, influencing the final outcome.

To convincingly validate these early markers as surrogates, a strict methodology has to be developed. Some recommendations have already been proposed by Buyse et al. (8;154) on the use of meta-analyses for validation of clinical surrogates and biomarkers.
Table 1. Grades of Recommendation Assessment, Development and Evaluation (GRADE) system (adapted from (12)).

<table>
<thead>
<tr>
<th>First step: Quality of evidence and definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. High quality</td>
</tr>
<tr>
<td>3. Moderate quality</td>
</tr>
<tr>
<td>2. Low quality</td>
</tr>
<tr>
<td>1. Very low quality</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second step: Criteria for assigning grade of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of evidence</strong></td>
</tr>
<tr>
<td>Randomised trial = high</td>
</tr>
<tr>
<td>Observational study = low</td>
</tr>
<tr>
<td>Any other evidence = very low</td>
</tr>
</tbody>
</table>

**Decrease grade if:**
- Serious (−1) or very serious (−2) limitation to study quality
- Important inconsistency (−1)
- Some (−1) or major (−2) uncertainty about directness
- Imprecise or sparse data (−1)
- High probability of reporting bias (−1)

**Increase grade if:**
- Strong evidence of association—significant relative risk of > 2 (< 0.5) based on consistent evidence from two or more observational studies, with no plausible confounders (+1)
- Very strong evidence of association—significant relative risk of > 5 (< 0.2) based on direct evidence with no major threats to validity (+2)
- Evidence of a dose response gradient (+1)
- All plausible confounders would have reduced the effect (+1)

<table>
<thead>
<tr>
<th>Third step: Strength of recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fs</strong> Strong recommendation for using an intervention</td>
</tr>
<tr>
<td><strong>Fw</strong> Weak recommendation for using an intervention</td>
</tr>
<tr>
<td><strong>Aw</strong> Weak recommendation against using an intervention</td>
</tr>
<tr>
<td><strong>As</strong> Strong recommendation against using an intervention</td>
</tr>
</tbody>
</table>

Determinants of strength: balance between desirable and undesirable effects; quality of evidence; values and preferences; costs (resource allocation)
Table 2. Meta-analyses or pooled analysis assessing the role of time-to-progression and progression-free survival as intermediate endpoint for overall survival in lung cancer patients.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Endpoint</th>
<th>Population</th>
<th>Considered studies</th>
<th>Period</th>
<th>N study</th>
<th>N patients</th>
<th>Relationship between the intermediate criteria and overall survival</th>
</tr>
</thead>
</table>
| Johnson (40) | TTP      | Advanced NSCLC | RCT of 1st line CT | 1966-2005 | 191     | 44125      | $r^2 = 0.19$  
|             |          |              |                    |        |         |            | $p = 0.0003$  
| Hotta (41)   | Median TTP ratio | Advanced NSCLC | Phase III RCT of 1st line CT | 1987-2002 | 54      | 23457      | $r^2 = 0.33$  
|             |          |              |                    |        |         |            | $p < 0.01$  
| Buyse (42)   | PFS      | Advanced NSCLC | RCT of 1st line CT* | 2003-2006 | 7       | 2838       | $r^2 = 0.61$  |

TTP = time-to-progression; PFS = progression-free survival; NSCLC = non-small cell lung cancer; RCT = randomised controlled trial; CT = chemotherapy

* RCT comparing docetaxel to vincalcaloids regimens
**Table 3.** WHO, SWOG and RECIST criteria for response assessment (14;46;47)

<table>
<thead>
<tr>
<th></th>
<th>WHO</th>
<th>SWOG</th>
<th>RECIST*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>complete disappearance of all measurable and assessable tumoural lesions not less than 4 weeks apart</td>
<td>complete disappearance of all measurable and assessable tumoural lesions, including normalisation of markers and other abnormal laboratory values for at least 3-6 weeks</td>
<td>complete disappearance of all tumoural lesions for at least 4 weeks</td>
</tr>
<tr>
<td>response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial</td>
<td>decrease ≥ 50% of the product of the 2 greatest perpendicular diameters for at least 4 weeks</td>
<td>decrease ≥ 50% of the product of the 2 greatest perpendicular diameters for at least 3-6 weeks</td>
<td>decrease ≥ 30% of the sum of the greatest diameters</td>
</tr>
<tr>
<td>response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stable</td>
<td>no response nor progression</td>
<td>decrease &lt; 50% or increase &lt; 50% (&lt;10 cm²) of the product of the 2 greatest diameters for at least 3-6 weeks</td>
<td>no response nor progression</td>
</tr>
<tr>
<td>disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progression</td>
<td>increase &gt; 25% of the product of the 2 greatest perpendicular diameters or appearance of a new lesion</td>
<td>increase &gt; 50% or &gt; 10 cm² of the product of the 2 greatest diameters or appearance of a new lesion or clear aggravation of any assessable lesion</td>
<td>increase &gt; 20% of the sum of the greatest diameters or appearance of a new lesion</td>
</tr>
</tbody>
</table>

*The final evaluation according to RECIST criteria takes into account the response on target and non-target lesions*
Table 4. Meta-analyses assessing response as an intermediate marker for overall survival in lung cancer

<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Effect on survival</th>
<th>U/M Response criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NSCLC treated by chemotherapy (1st line and salvage CT combined)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sekine (52)</td>
<td>6768</td>
<td>OR: $r^2 = 0.36$ ($p = 0.00003$)</td>
<td>M -</td>
</tr>
<tr>
<td>Tsujino (54)</td>
<td>6171</td>
<td>OR: $p &lt; 0.0001$ (↑MST 0.26m/↑1% OR)</td>
<td>U WHO, RECIST</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DCR: $p = 0.003$ (↑MST 0.17m/↑1% DCR)</td>
<td></td>
</tr>
<tr>
<td><strong>Advanced NSCLC receiving 1st line chemotherapy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson (40)</td>
<td>44125</td>
<td>OR: $r^2 = 0.16$ ($p &lt; 0.0001$)</td>
<td>M WHO</td>
</tr>
<tr>
<td>Hotta (50)</td>
<td>43551</td>
<td>OR: $r = 0.08$; $p &lt; 0.0001$</td>
<td>M -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DCR: $r = 0.08$; $p &lt; 0.0001$</td>
<td></td>
</tr>
<tr>
<td>Shanafelt (53)</td>
<td>2794</td>
<td>OR: $r^2 = 0.22$ (↑MST 1w/↑3.3% OR)</td>
<td>U -</td>
</tr>
<tr>
<td><strong>NSCLC receiving salvage chemotherapy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hotta (50)</td>
<td>4318</td>
<td>OR: $p = 0.69$</td>
<td>M -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DCR: $p = 0.013$</td>
<td></td>
</tr>
<tr>
<td>Kurata (51)</td>
<td>4292</td>
<td>OR: $p &lt; 0.001$ (↑MST 0.07m/↑1% OR)</td>
<td>M -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD: $p = 0.04$ (↑MST 0.038m/↑1% SD)</td>
<td></td>
</tr>
<tr>
<td><strong>Extensive disease SCLC receiving 1st line chemotherapy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hotta (49)</td>
<td>8779</td>
<td>OR: $r^2 = 0.33$</td>
<td>U WHO, RECIST, other</td>
</tr>
</tbody>
</table>

N = number of patients; U/M: results obtained in univariate/multivariate analysis; WHO = world health organisation; CT = chemotherapy; OR = overall response; DCR = disease control rate; MST = median survival time; SD = stable disease; m = month; w = week
Table 5. Retrospective analyses of prospective studies assessing response as an intermediate marker for overall survival in lung cancer

<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Population</th>
<th>T</th>
<th>Effect on survival</th>
<th>p</th>
<th>U/M</th>
<th>Response criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pooled analyses assessing the predictive value of response on survival as primary objective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lara (55)</td>
<td>984</td>
<td>NSCLC IV</td>
<td>CT</td>
<td>OR: HR 0.61 DCR: HR 0.45</td>
<td>&lt;0.001</td>
<td>U</td>
<td>RECIST</td>
</tr>
<tr>
<td>Yamamoto (56)</td>
<td>806</td>
<td>NSCLC III/IV</td>
<td>CT</td>
<td>DCR Phase II 14.5 vs 10.4m</td>
<td>0.04</td>
<td>U</td>
<td>RECIST</td>
</tr>
<tr>
<td>Paesmans (57)</td>
<td>1052</td>
<td>Advanced NSCLC</td>
<td>CT</td>
<td>OR: RR 2.20</td>
<td>&lt;0.001</td>
<td>M</td>
<td>WHO</td>
</tr>
<tr>
<td>Lebeau (58)</td>
<td>1280</td>
<td>SCLC LD/ED</td>
<td>CT</td>
<td>CR: MST 16m PR: 12m No response 10m</td>
<td>&lt;0.001</td>
<td>M</td>
<td>WHO</td>
</tr>
<tr>
<td><strong>Retrospective analyses of individual prospective randomised trials having as secondary objective the assessment of response as intermediate criteria for survival</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depierre (60)</td>
<td>179</td>
<td>NSCLC IB-III A</td>
<td>CT → surgery</td>
<td>OR: RR0.37 CR: RR 0.42</td>
<td>0.0001</td>
<td>U</td>
<td>WHO</td>
</tr>
<tr>
<td>Nagai (61)</td>
<td>31</td>
<td>NSCLC IIAN2</td>
<td>CT → surgery</td>
<td>PR: MST 30m SD: MST 18m PD: MST 12m</td>
<td>0.05</td>
<td>U</td>
<td>-</td>
</tr>
<tr>
<td>Sculier (59)</td>
<td>462</td>
<td>NSCLC III</td>
<td>CT-RT</td>
<td>PR: MST 57w SD: MST 32w PD: MST 16w</td>
<td>&lt;0.001</td>
<td>M</td>
<td>WHO</td>
</tr>
<tr>
<td>Groen (62)</td>
<td>160</td>
<td>NSCLC III</td>
<td>RT or CT-RT</td>
<td>OR: MST 16m No response: MST 10m</td>
<td>&lt;0.001</td>
<td>U</td>
<td>WHO</td>
</tr>
<tr>
<td>Crino (63)</td>
<td>66</td>
<td>NSCLC III</td>
<td>RT or CT-RT</td>
<td>OR: RR 3.86</td>
<td>&lt;0.001</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>Huber (64)</td>
<td>303</td>
<td>NSCLC III</td>
<td>CT → RT or CT-RT</td>
<td>OR to induction CT correlated with survival</td>
<td>0.01</td>
<td>M</td>
<td>WHO</td>
</tr>
<tr>
<td>Joss (74)</td>
<td>266</td>
<td>SCLC LD/ED</td>
<td>CT or CT-RT</td>
<td>CR &gt; PR /no change &gt; Progression</td>
<td>&lt;0.001</td>
<td>U</td>
<td>WHO</td>
</tr>
</tbody>
</table>

N = number of patients; U/M: results obtained in univariate/multivariate analysis; T = therapy; WHO = world health organisation; SWOG = Southwest Oncology Group; CT =
chemotherapy; CT-RT = chemoradiotherapy; OR = overall response; DCR = disease control rate; CR = complete response; PR = partial response; m = months; MST = median survival time; w = weeks; RR = relative risk; LD = limited disease; ED = extensive disease
### Table 6. Comparison between WHO and RECIST 1.0 for response assessment in lung cancer

<table>
<thead>
<tr>
<th>Authors</th>
<th>P/R</th>
<th>N</th>
<th>Population</th>
<th>RR-RECIST</th>
<th>RR-WHO</th>
<th>PD-RECIST</th>
<th>PD-WHO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Therasse (47)</td>
<td>R</td>
<td>1197</td>
<td>ND</td>
<td>32%</td>
<td>32%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Werner-Wasik (77)</td>
<td>P</td>
<td>22</td>
<td>locally advanced NSCLC</td>
<td>87%</td>
<td>87%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cortes (78)</td>
<td>R</td>
<td>164</td>
<td>Metastatic NSCLC</td>
<td>52%</td>
<td>52%</td>
<td>26%</td>
<td>26%</td>
</tr>
<tr>
<td>Watanabe (79)</td>
<td>R</td>
<td>120</td>
<td>Metastatic NSCLC</td>
<td>19%</td>
<td>20%</td>
<td>13%</td>
<td>17.5%</td>
</tr>
<tr>
<td>Park (75)</td>
<td>R</td>
<td>28</td>
<td>NSCLC</td>
<td>39%</td>
<td>43%</td>
<td>25%</td>
<td>29%</td>
</tr>
<tr>
<td>Konishi (76)</td>
<td>R</td>
<td>32</td>
<td></td>
<td>38%</td>
<td>47%</td>
<td>6%</td>
<td>6%</td>
</tr>
</tbody>
</table>

P/R = prospective/retrospective study; N = number of patients, RR = response rate; PD = progressive disease; ND = no data
Table 7. Studies assessing the relationship between modification of circulating markers levels after treatment and survival in lung cancer.

<table>
<thead>
<tr>
<th>Authors</th>
<th>P/R</th>
<th>Marker</th>
<th>N</th>
<th>Population</th>
<th>Therapy</th>
<th>Effect on survival</th>
<th>p</th>
<th>U/M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Okada (138)</td>
<td>R</td>
<td>CEA</td>
<td>1000</td>
<td>NSCLC cI</td>
<td>Surgery</td>
<td>↓ for high postsurgical level</td>
<td>&lt; 0.0001</td>
<td>M</td>
</tr>
<tr>
<td>Nisman (139)</td>
<td>P</td>
<td>CEA</td>
<td>45</td>
<td>Advanced NSCLC</td>
<td>CT</td>
<td>↑ if ↓ ≥ 35%</td>
<td>0.04</td>
<td>U</td>
</tr>
<tr>
<td>Sawabata (140)</td>
<td>R</td>
<td>CEA</td>
<td>19</td>
<td>ADC I</td>
<td>Surgery</td>
<td>↓ for high postsurgical level</td>
<td>0.006</td>
<td>U</td>
</tr>
<tr>
<td>Sawabata (141)</td>
<td>R</td>
<td>CEA</td>
<td>242</td>
<td>NSCLC IA</td>
<td>Surgery</td>
<td>↓ for high postsurgical level</td>
<td>&lt; 0.0001</td>
<td>M</td>
</tr>
<tr>
<td>Sawabata (152)</td>
<td>R</td>
<td>CEA</td>
<td>297</td>
<td>NSCLC cI</td>
<td>Surgery</td>
<td>↓ for high postsurgical level</td>
<td>0.002</td>
<td>M</td>
</tr>
<tr>
<td>Bréchot (142)</td>
<td>P</td>
<td>CEA</td>
<td>116</td>
<td>NSCLC I-IV</td>
<td>Various</td>
<td>↑ if ↓ ≥ 50%</td>
<td>0.006</td>
<td>U</td>
</tr>
<tr>
<td>Chiu (143)</td>
<td>R</td>
<td>CEA</td>
<td>89</td>
<td>Relapsing NSCLC</td>
<td>Gefitinib</td>
<td>↑ if ↓ ≥ 50% at 4 weeks</td>
<td>0.03</td>
<td>U</td>
</tr>
<tr>
<td>Spiridonidis (156)</td>
<td>R</td>
<td>CEA, CA19-9, CA125</td>
<td>36</td>
<td>NSCLC I-IIIB</td>
<td>CT or CT-RT</td>
<td>↑ if ↓ if decrease</td>
<td>0.0002</td>
<td>U</td>
</tr>
<tr>
<td>Nisman (139)</td>
<td>P</td>
<td>Cyfra 21-1</td>
<td>45</td>
<td>Advanced NSCLC</td>
<td>CT</td>
<td>↑ if ↓ ≥ 35%</td>
<td>0.01</td>
<td>M</td>
</tr>
<tr>
<td>Hamzaoui (144)</td>
<td>?</td>
<td>Cyfra 21-1</td>
<td>63</td>
<td>NSCLC/SCLC III-IV</td>
<td>CT</td>
<td>↑ if ↓ ≥ 70%</td>
<td>0.44</td>
<td>U</td>
</tr>
<tr>
<td>Bréchot (142)</td>
<td>P</td>
<td>Cyfra 21-1</td>
<td>116</td>
<td>NSCLC I-IV</td>
<td>Various</td>
<td>↑ if ↓ ≥ 50%</td>
<td>0.0001</td>
<td>M</td>
</tr>
<tr>
<td>Sunaga (145)</td>
<td>R</td>
<td>proGRP</td>
<td>29</td>
<td>SCLC I-IV</td>
<td>CT</td>
<td>↑ if ↓ ≥ 50%</td>
<td>0.001</td>
<td>M</td>
</tr>
<tr>
<td>Okusaka (146)</td>
<td>R</td>
<td>proGRP</td>
<td>18</td>
<td>SCLC in PR</td>
<td>CT</td>
<td>↑ if return to normal</td>
<td>&lt; 0.05</td>
<td>U</td>
</tr>
<tr>
<td>Bréchot (142)</td>
<td>P</td>
<td>NSE</td>
<td>116</td>
<td>NSCLC I-IV</td>
<td>Various</td>
<td>↑ if ↓ ≥ 50%</td>
<td>0.018</td>
<td>U</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Sample Size</th>
<th>Disease Type</th>
<th>Treatment</th>
<th>Effect</th>
<th>p-value (Univariate)</th>
<th>p-value (Multivariate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fizazi (147)</td>
<td>R</td>
<td>NSE</td>
<td>135</td>
<td>SCLC LD/ED</td>
<td>CT-RT CT</td>
<td>↑ if return to normal</td>
<td>0.02</td>
</tr>
<tr>
<td>Bréchot (142)</td>
<td>P</td>
<td>CA125</td>
<td>116</td>
<td>NSCLC I-IV</td>
<td>Various</td>
<td>↑ if ≥ 50%</td>
<td>0.0006</td>
</tr>
<tr>
<td>Chiu (143)</td>
<td>R</td>
<td>CA125</td>
<td>89</td>
<td>Relapsing NSCLC</td>
<td>Gefitinib</td>
<td>no effect</td>
<td>0.12</td>
</tr>
<tr>
<td>Chiu (143)</td>
<td>R</td>
<td>CA19-9</td>
<td>89</td>
<td>Relapsing NSCLC</td>
<td>Gefitinib</td>
<td>↑ if ≥ 25% at 8 weeks</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Gauthier (148)</td>
<td>R</td>
<td>Anemia</td>
<td>242</td>
<td>NSCLC IB-II</td>
<td>Surgery + CT</td>
<td>No effect</td>
<td>0.06</td>
</tr>
<tr>
<td>Laurie (151)</td>
<td>R</td>
<td>Anemia</td>
<td>652</td>
<td>SCLC LD</td>
<td>CT-RT</td>
<td>No effect</td>
<td>0.33</td>
</tr>
<tr>
<td>Ishikawa (150)</td>
<td>P?</td>
<td>CTC</td>
<td>74</td>
<td>Relapsing NSCLC</td>
<td>Gefitinib</td>
<td>↑ if ratio T0/T2w or T0/T4w &lt; 1.2</td>
<td>0.002</td>
</tr>
<tr>
<td>Chen (149)</td>
<td>P</td>
<td>CTC</td>
<td>67</td>
<td>NSCLC I-IIIB</td>
<td>CT-RT</td>
<td>Persistence of CTC</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Yamashita (153)</td>
<td>P</td>
<td>CTC</td>
<td>103</td>
<td>NSCLC I-IIIA</td>
<td>Surgery</td>
<td>Persistence of CTC</td>
<td>NS</td>
</tr>
</tbody>
</table>

NSCLC = non small cell lung cancer; SCLC = small cell lung cancer; P/R = prospective/retrospective study; N = number of patients, U/M: results obtained in univariate/multivariate analysis; CT = chemotherapy; ADC = adenocarcinoma; proGRP = pro-gastrin-releasing peptide; PR = partial remission; NSE = neuron specific enolase; LD = limited disease; ED = extensive disease; CT-RT = chemoradiotherapy; CTC = circulating tumour cells; NS = no statistically significant
Table 8. Summary of the recommendations based on level of evidence, quality of evidence and grade of recommendation (GRADE system).

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Level of evidence</th>
<th>Quality of evidence</th>
<th>Grade of recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1) Time-to-progression is an intermediate marker for overall survival in advanced NSCLC treated with first-line chemotherapy</td>
<td>MA</td>
<td>3</td>
<td>Fs</td>
</tr>
<tr>
<td>1 2) Progression-free survival is a potential intermediate marker for survival in the setting of docetaxel or vinca-alcaloids first-line based-regimens</td>
<td>MA</td>
<td>2</td>
<td>Fw</td>
</tr>
<tr>
<td>2 1) Objective response is an intermediate criterion for overall survival in advanced NSCLC treated with first-line or salvage chemotherapy</td>
<td>MA/RT</td>
<td>4</td>
<td>Fs</td>
</tr>
<tr>
<td>2 2) Objective response is an intermediate criterion for overall survival in extensive disease SCLC treated with first-line chemotherapy</td>
<td>MA/ReS</td>
<td>2</td>
<td>Fw</td>
</tr>
<tr>
<td>2 3) Objective response is a potential intermediate criterion for overall survival in operable NSCLC treated with induction chemotherapy and in locoregionally advanced NSCLC treated by chemotherapy and radiotherapy</td>
<td>ReS</td>
<td>2</td>
<td>Fw</td>
</tr>
<tr>
<td>3 1) Although commonly used, conventional criteria for response assessment of primary lung tumour treated by (chemo)radiotherapy can not be used as intermediate criteria for survival</td>
<td>ReS</td>
<td>1</td>
<td>Aw</td>
</tr>
<tr>
<td>3 2) Local control, for which the definition has to be clarified, may be a possible intermediate criterion for overall survival</td>
<td>ReS</td>
<td>1</td>
<td>Fw</td>
</tr>
<tr>
<td>4) Response at the QOL level cannot be recommended as an intermediate criterion for overall survival due to a lack of robust data</td>
<td>ReS</td>
<td>1</td>
<td>Aw</td>
</tr>
<tr>
<td>5) Although not strictly demonstrated, there are numerous arguments from subgroup analyses of randomised trials and retrospective studies that mediastinal downstaging and, in a less extent, TN downstaging are associated with better survival in locally advanced NSCLC treated by induction chemotherapy or chemoradiotherapy before surgery</td>
<td>ReS</td>
<td>2</td>
<td>Fw</td>
</tr>
<tr>
<td>6 1) Complete resection is a prognostic factor for survival in resected NSCLC and can be used as an intermediate criterion for overall survival</td>
<td>Cohort/R eS</td>
<td>3</td>
<td>Fs</td>
</tr>
<tr>
<td>6 2) Pathological TNM is a prognostic factor for survival in resected NSCLC and can be used as an intermediate criterion for overall survival</td>
<td>Cohort</td>
<td>3</td>
<td>Fs</td>
</tr>
<tr>
<td>7) Metabolic response assessed by PET scan should not been used for the routine assessment of response to treatment in lung cancer patients in place of morphological criteria</td>
<td>ReS/CS</td>
<td>1</td>
<td>Aw</td>
</tr>
</tbody>
</table>
8) Tissue biological markers have not to be used for evaluation of treatment efficacy and are not adequate intermediate criteria for overall survival in lung cancer patients

| 9 1) It is suggested that some circulating markers (CEA, Cyfra 21-1, proGRP and to a less extent NSE, CA-125, CA19-9), used as single criterion to assess overall survival, could be adequate intermediate criteria for survival in lung cancer patients |
|---|---|---|
| CS | 1 | Aw |

| 9 2) The persistence of circulating tumour cells in NSCLC may have a prognostic impact on further survival |
|---|---|---|
| ReS | 2 | Fw |
| Cohort | 1 | Fw |

MA = meta-analysis; RT = randomised trial; ReS = retrospective studies; Cohort = cohort study; CS = case series
Annex 1. List of MeSH headings and free-text keywords chosen to search Ovid Medline database

<table>
<thead>
<tr>
<th>1 : Pathological conditions (P criterion)</th>
<th>Searched Mesh terms, keywords and phrases</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>2 : Outcome (O criterion)</th>
<th>Searched Mesh terms, keywords and phrases (to be combined with the “OR” operator)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>3. Possible intermediate criteria (I criteria) (alphabetically listed)</th>
<th>Searched Mesh terms, keywords and phrases (to be combined with the “OR” operator)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer Staging : downstaging</td>
<td>downstagin$.tw OR tum$. regress$.tw</td>
</tr>
</tbody>
</table>
| Cancer Staging : pTNM versus cTNM | pTNM.tw AND cTNM.tw  
**Additional limit**: Comparative Study/ |
| Disease Progression | Disease -Free Survival/ OR Disease Progression/ OR Remission Induction/ OR disease* control$.tw OR disease exacerbation*.tw OR disease free survival*.tw OR disease progression*.tw OR disease recur$.tw. OR event free survival*.tw OR progression assessment*.tw OR progression free survival*.tw OR relaps$.tw OR remission*.tw OR time* to progression.tw  
**Additional limit**: Limit to meta-analysis and randomized controlled trial (Publication Types) |
**Additional limit**: NOT exp Lung Neoplasms/sc |
<table>
<thead>
<tr>
<th>Topic</th>
<th>Search Terms</th>
<th>Additional Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oncoproteins and Tumor Suppressor Proteins</td>
<td>[exp Oncogene Proteins/ OR exp Tumor Suppressor Proteins/ OR erb? receptor?.tw OR fibroblast growth factor?.tw OR EGF receptor?.tw OR tyrosine kinase?.tw OR kinase inhibitor?.tw OR tumor suppressor?.tw OR protein? P53.tw OR wt? protein?.tw] AND [Tumor Markers, Biological/ OR plasma?.tw OR blood sample?.tw ]</td>
<td>Additional limit : Comparative Study/</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>Attitude to Health/ OR Health Status Indicators/ OR Karnofsky Performance Status/ OR Patient participation/ OR Quality of Life OR Sickness Impact Profile/ OR health apprais$.tw OR health status ind$.tw OR Karnofsky.ty OR patient rated.tw OR patient reported.tw OR performance status.ty OR quality of life.ty OR self report.tw</td>
<td>Additional limit : Limit to meta-analysis and randomized controlled trial (Publication Types)</td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>Radiotherapy/ OR Radiotherapy, Adjuvant/ OR Radiotherapy, Computer-Assisted/ OR Radiotherapy, High Energy/</td>
<td></td>
</tr>
<tr>
<td>Resection</td>
<td>[lobectomy.tw OR pneumonectiony.tw OR resection.tw OR resectable.tw OR thoracotomy.tw OR tumor resection.tw OR sternectomy.tw OR unresectable.tw] Additional limit : Comparative Study/</td>
<td></td>
</tr>
<tr>
<td>Surrogates</td>
<td>surrogate*.tw</td>
<td></td>
</tr>
<tr>
<td>Time factors</td>
<td>[Time Factors/ OR “during and after” .tw OR before and after”.tw]</td>
<td>Additional limit : NOT exp Lung Neoplasms/sc [Secondary] + Limit to meta-analysis and randomized controlled trial (Publication Types)</td>
</tr>
<tr>
<td>Treatment Outcome</td>
<td>exp “Outcome Assessment (Health Care)”/; OR outcome assessment*.tw; OR outcome measure*.tw; OR recist.tw; OR response effectiveness.tw; OR response evaluation*.tw; OR response* to treatment.tw; OR therapeutic result*.tw; OR treatment efficac$.tw; OR treatment failure*.tw; OR</td>
<td></td>
</tr>
</tbody>
</table>
treatment outcome*.tw

**Additional limit**: Limit to meta-analysis and randomized controlled trial (Publication Types)


In the search strategies, MeSH terms are recognizable by the slash symbol (/) indicating that all the possible subheading combinations were taken into consideration. The “exp.” prefix indicates exploded MeSH terms. The other keywords are identified by “.tw” extension. The asterisk and question mark (* or ? standing for one single character) and dollar symbols ($) standing for a chain of characters) were used to account for every possible lexical or grammatical keyword variants. Completed search strategies systematically combined the "P" and "O" criteria with one of the "I" criteria.
References


