



Systematic qualitative review of randomised trials conducted in nonsmall cell lung cancer with a noninferiority or equivalence design

Marianne Paesmans¹, Bogdan Grigoriu², Sebahat Ocak³, Martine Roelandts⁴, Jean-Jacques Lafitte⁵, Stéphane Holbrechts⁶, Thierry Berghmans⁷, Anne-Pascale Meert⁷, Luigi Moretti⁴, Sophie Danyi⁸, Françoise Pasleau⁸, Lieveke Ameye¹, Alain Van Meerhaeghe⁹ and Jean-Paul Sculier⁷

Number 7 in the series “Challenges and controversies in thoracic oncology”
Edited by J-P. Sculier, B. Besse and P. Van Schil

Affiliations: ¹Data Centre, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium. ²Thoracic Oncology Dept, Regional Institute of Oncology Iasi, University of Medicine and Pharmacy, “Gr. T. Popa”, Iasi, Romania. ³Pneumology Dept, CHU Mont-Godinne, Université Catholique de Louvain, Yvoir, Belgium. ⁴Radiotherapy Dept, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium. ⁵Pneumology Dept, CHU Lille, University of Lille 2, Lille, France. ⁶Medical Oncology Dept, CHU Ambroise Paré, Mons, Belgium. ⁷Thoracic Oncology and Intensive Care Dept, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium. ⁸Library of Life Sciences, Université de Liège, Liège, Belgium. ⁹Pneumology Department, CHU Charleroi, Charleroi, Belgium.

Correspondence: Marianne Paesmans, Data Centre, Institut Jules Bordet, Université Libre de Bruxelles, Rue Héger-Bordet 1, B 1000 Brussels, Belgium. E-mail: marianne.paesmans@bordet.be

ABSTRACT The use of noninferiority randomised trials for patients with advanced non-small cell lung cancer has emerged during the past 10–15 years but has raised some issues related to their justification and methodology. The present systematic review aimed to assess trial characteristics and methodological aspects.

All randomised clinical trials with a hypothesis of noninferiority/equivalence, published in English, were identified. Several readers extracted *a priori* defined methodological information. A qualitative analysis was then performed.

We identified 20 randomised clinical trials (three phase II and 17 phase III), 11 of them being conducted in strong collaboration with industry. We highlighted some deficiencies in the reports like the lack of justification for both the noninferiority assumption and the definition of the noninferiority margin, as well as inconsistencies between the results and the authors’ conclusions. CONSORT guidelines were better followed for general items than for specific items ($p < 0.001$).

Improvement in the reporting of the methodology of noninferiority/equivalence trials is needed to avoid misleading interpretation and to allow readers to be fully aware of the assumptions underlying the trial designs. They should be restricted to limited specific situations with a strong justification why a noninferiority hypothesis is acceptable.



@ERSpublications

Reporting on noninferiority trials should be improved and results interpreted using the noninferiority margin <http://ow.ly/EB6dg>

This article has supplementary material available from erj.ersjournals.com

Received: May 20 2014 | Accepted after revision: Nov 17 2014 | First published online: Dec 23 2014

Conflict of interest: None declared.

Copyright ©ERS 2015

Introduction

Despite the intense research efforts in advanced non-small cell lung cancer (NSCLC) in the past 15 years, the prognosis of these patients remains dismal [1]. Moreover, the number of trials involving new agents reporting a positive outcome without meeting their primary end-point has increased over time and the magnitude of benefit in survival has tended to decrease over time, together with a sharp increase in trial sample size [2].

We face, then, with a situation where there is definitely a need to improve patient outcome (already in the first-line setting and even more in the second-line setting) while, in parallel, the conduct of noninferiority trials has developed during the last 10–15 years [3]. This means that investigators and pharmaceutical companies are willing to test experimental drugs (or therapeutic schemes), accepting that those drugs will have decreased efficacy. Indeed, noninferiority does not mean equivalence (statistically, strict equivalence could be demonstrated only if an infinite sample size could be reached) and noninferiority will in fact be translated, for implementation in a clinical trial, into a decreased efficacy lower than an amount to be specified. This amount is called the noninferiority margin. There are, of course, some arguments that may lead to the design of noninferiority trials [3]. Improvement can be hoped for in patients' quality of life (favouring then a treatment with a better quality of survival rather than a longer quantity of survival). Reducing toxicity might be another situation when a noninferiority design could be considered, in particular because this can improve patient quality of life or lead, should noninferiority be established, to a treatment opportunity for further patients who would otherwise have been assessed as unfit for treatment. Decreasing costs might be another parameter, as cost-effectiveness is taken into account by payers for deciding on reimbursement of new drugs. This expected benefit in a secondary outcome should be obvious in order to accept the potential reduced efficacy and should be clearly formulated at the time of formulating trial design. There is therefore a need to carefully justify the rationale for conducting a noninferiority trial. Beyond this justification, a second inherent difficulty is to define what one means by decreased efficacy and how the noninferiority margin will be determined. There are essentially two proposed methods to determine the noninferiority margin [3–6]: either to choose it by a clinical judgment that should take into account the expected benefits on secondary outcomes, with the drawback that there will be no proof that the experimental treatment performs better than the historical standard, or to choose it in order to maintain a proportion of the benefit that has been obtained by the reference treatment in the trials or in the meta-analyses that led one to consider this treatment as a new standard. In fact, the trial results would be more convincing in the case of success when a small margin is used but the required sample sizes may become very large.

Importantly, noninferiority trials require a specific methodology [7] and special guidelines (Consolidated Standards of Reporting Trials (CONSORT)) have been published on the items to report when publishing [8]. For the interested reader, some specific methodological aspects of noninferiority trials are developed in the online supplementary material.

There have been previous reviews assessing the purpose and the methodological aspects of noninferiority trials [3, 9, 10] in oncology (mainly breast cancer, NSCLC and colorectal cancer): all three identified critical deficiencies in design and reporting, namely related to the choice and justification of the noninferiority margin. However, no review, to our knowledge, focused on trials performed in patients with advanced NSCLC. Therefore, we thought it worth conducting a systematic review on that field with the aims to: 1) report on the general methodology of the trials and on the specific aspects linked to the noninferiority assumption(s); 2) make an overview of the noninferiority margins that have been used; 3) assess result interpretation according to the noninferiority hypothesis; 4) assess whether the CONSORT guidelines were followed; and 5) assess the reported reasons for carrying out noninferiority trials.

Methods

Trial identification

We defined the following eligibility criteria for a trial to be included in our review: 1) randomised clinical trial conducted in patients with advanced NSCLC; 2) trial testing a noninferiority hypothesis for anticancer drugs/regimens (or an equivalence hypothesis); and 3) trial published as a full paper in English (or French, as French is spoken by all the authors of this review). We excluded publications reporting protocols only, as

Previous articles in this series: **No. 1:** Powell HA, Baldwin DR. Multidisciplinary team management in thoracic oncology: more than just a concept? *Eur Respir J* 2014; 43: 1776–1786. **No. 2:** Shlomi D, Ben-Avi R, Balmor GR, *et al.* Screening for lung cancer: time for large-scale screening by chest computed tomography. *Eur Respir J* 2014; 44: 217–238. **No. 3:** De Ruysscher D, Nakagawa K, Asamura H. Surgical and nonsurgical approaches to small-size nonsmall cell lung cancer. *Eur Respir J* 2014; 44: 483–494. **No. 4:** Van Schil PE, Opitz I, Weder W, *et al.* Multimodal management of malignant pleural mesothelioma: where are we today? *Eur Respir J* 2014; 44: 754–764. **No. 5:** Kim L, Tsao MS. Tumour tissue sampling for lung cancer management in the era of personalised therapy: what is good enough for molecular testing? *Eur Respir J* 2014; 44: 1011–1022. **No. 6:** Blum T, Schönfeld N. The lung cancer patient, the pneumologist and palliative care: a developing alliance. *Eur Respir J* 2015; 45: 211–226.

well as duplicate publications, and publications reporting outcomes other than that tested for noninferiority or data pooled from several trials.

The literature search was performed by two experienced librarians (F. Pasleau and S. Danyi) in April 2014. In addition, we examined the bibliography of review papers selected by the same search. The search strategy is described in the online supplementary material.

Data extraction

The following data, defined before the literature search, were extracted from all publications. General characteristics: 1) year of publication, journal, first author, involvement of a pharmaceutical company in the conduct of the trial or in the interpretation of the results (that we scored as none, uncertain when a grant was provided or certain when some authors of the publication were company employees); 2) patients and treatment data (stage of disease, standard and experimental treatments, and treatment setting). Trial methodology: 1) trial phase, trial design and trial type (frequentist or Bayesian), primary outcome, secondary outcomes, description in the methods section of type of hypothesis (noninferiority/superiority) for each outcome, rationale for noninferiority, whether interim analyses were planned and, if yes, the method for adjusting the final analysis; 2) if a noninferiority hypothesis was used for the primary outcome, measure of treatment effect, margin of noninferiority, method for choosing the margin, whether the sample size was consistent with the design of noninferiority and, if yes, the alternative hypothesis for power calculation were extracted. Trial results for the primary outcome for two patient populations: 1) intent-to-treat (ITT) and per protocol (PP). Conclusion: of the authors and whether the conclusions were in accordance with the hypothesis of noninferiority.

In addition, we qualitatively assessed the trials using: 1) the CONSORT guidelines (the general guideline for reporting parallel-group randomised trials [11] and its extension for reporting noninferiority and equivalence randomised trials [8]); and 2) the risk of bias evaluation tool developed by the Cochrane Collaboration [12]. The CONSORT checklist consists of 37 items to be reported for randomised clinical trials and 12 extended items specific to studies with noninferiority hypotheses. We scored them as 1 (not reported), 2 (reported, but partially or inadequately) or 3 (reported).

The risk of bias tool from the Cochrane Collaboration is specifically dedicated to aspects of trials that are most linked to the risk of obtaining a biased estimate of the treatment effect. The scoring was performed as recommended by the Cochrane Collaboration, as low risk of bias, unclear risk of bias or high risk of bias in six domains (two related to the randomisation procedure, two related to the blinding of the study, one related to bias induced by incomplete outcome data and one related to bias induced by selective reporting).

Each publication was read, for data extraction, by three to 11 readers (median 5.5). All extractions were reviewed by one author (M. Paesmans) who performed a final and common evaluation by resolving the inconsistencies or, in case of subjectivity in the assessments, by calculating an average score. All final evaluations were submitted to the readers, reviewed and approved.

Analysis

We summarised our data extraction in descriptive tables providing individual results for the qualitative characteristics that we extracted from each publication as well as for each domain of the risk of bias tool.

For the items from the CONSORT guidelines, we used frequency tabulations and, for each of the items, calculated the mode, median and mean (although the numeric scores, in fact, reflect only an ordinal scale). We also calculated, for the general and extended items, an overall mean result. For the risk of bias tool, we calculated the number of items with a poor assessment, *i.e.* assessment of unclear risk of bias or high risk of bias (ranging from 0 to 6). We compared the medians of the distributions of the mean CONSORT scores (general and extended) with a Wilcoxon rank-signed test. We also assessed the impact of the following variables, both for the risk of bias tool (number of items with poor assessment) and the mean CONSORT scores, with Mann–Whitney tests: year of publication (dichotomised according to the median value), setting (first or second line), primary outcome (overall survival or not), reproducibility of sample size (yes or no) possible with the details provided in the publications, consistency of the authors conclusions with the results and the pre-planned design (yes or no), involvement of company members in the conduct of the trials and the analysis of the data, and high or low risk of bias according to the Cochrane tool.

SAS v9.3 (SAS Institute, Cary, NC, USA) was used for the statistical analysis.

Results

Trials identified

We identified 20 eligible trials (all published in English), which were published between 2002 and 2013 [13–32]: 11 were conducted in chemotherapy-naïve patients (*i.e.* first-line setting) [13–23] and nine in

populations of patients with a progression or relapse following previous chemotherapy [24–32]. By definition, all were conducted in patients with locally advanced or metastatic tumour. Only one trial was specific to patients with stage IV disease. There were three phase II trials and 17 phase III trials. Two trials were purely academic, and 11 were fully conducted together with the pharmaceutical industry with members of the company included as authors (table 1). Table 1 provides the other general characteristics of the trials including the description of the standard and experimental arms.

Trial methodology

General methodology

The general methodological characteristics of the trials are reported in table 2. All had primary objectives with a noninferiority assumption (equivalence in one) and none of them had a Bayesian design (not

TABLE 1 General characteristics of the trials

First author [ref.]	Year	Journal	Company [#]	Disease stage	Standard arm	Experimental arm(s)
First-line setting						
NEGORO [13]	2003	<i>Br J Cancer</i>	3	IIIB–IV	Cisplatin+vindesine	Irinotecan
ROSELL [14]	2002	<i>Ann Oncol</i>	2	III–IV	Cisplatin+paclitaxel	Carboplatin+paclitaxel
FOSSELLA [15]	2003	<i>J Clin Oncol</i>	3	IIIB–IV	Cisplatin +vinorelbine	Cisplatin+docetaxel Carboplatin+docetaxel
OHE [16]	2007	<i>Ann Oncol</i>	3	IIIB–IV	Cisplatin+irinotecan	Carboplatin+paclitaxel Cisplatin+gemcitabine Cisplatin+vinorelbine
NOVELLO [17]	2007	<i>Ann Oncol</i>	1	IV	Cisplatin +gemcitabine (2 cycles) followed by 3 further cycles in nonprogressive patients	Cisplatin-gemcitabine (2 cycles) followed by 3 cycles of gemcitabine in nonprogressive patients
PARK [18]	2007	<i>J Clin Oncol</i>	1	IIIB–IV	4 cycles of cisplatin-based CT in patients nonprogressive after 2 cycles	2 cycles of cisplatin based CT in patients nonprogressive after 2 cycles
SCAGLIOTTI [19]	2008	<i>J Clin Oncol</i>	2	IIIB–IV	Cisplatin +gemcitabine	Cisplatin+pemetrexed
HEYMACH [20]	2008	<i>J Clin Oncol</i>	2	IIIB–IV	Carboplatin +paclitaxel	Vandetanib Vandetanib+paclitaxel +carboplatin (superiority assumption for this arm)
MOK [21]	2009	<i>N Engl J Med</i>	2	IIIB–IV	Carboplatin +paclitaxel	Gefitinib
OKAMOTO [22]	2010	<i>J Clin Oncol</i>	3	IIIB–IV	Carboplatin +paclitaxel	Carboplatin+S-1 [¶]
LEE [23]	2013	<i>Clin Lung Cancer</i>	3	IIIB–IV	Cisplatin+paclitaxel	Paclitaxel-loaded polymeric micelle +cisplatin
Second-line setting						
SCHUETTE [24]	2005	<i>J Clin Oncol</i>	3	IIIB–IV	Docetaxel (3-weekly regimen)	Docetaxel (weekly regimen)
HANNA [25]	2004	<i>J Clin Oncol</i>	2	IIIB–IV	Docetaxel	Pemetrexed
RAMLAU [26]	2006	<i>J Clin Oncol</i>	2	IIIB–IV	Docetaxel	Topotecan
KRZAKOWSKI [27]	2010	<i>J Clin Oncol</i>	2	IIIB–IV	Docetaxel	Vinflunine
MARUYAMA [28]	2008	<i>J Clin Oncol</i>	2	IIIB–IV	Docetaxel	Gefitinib
KIM [29]	2008	<i>Lancet</i>	2	IIIB–IV	Docetaxel	Gefitinib
LI [30]	2012	<i>Pulm Pharm Ther</i>	3	IIIB–IV	Docetaxel	Pemetrexed
NATALE [31]	2011	<i>J Clin Oncol</i>	2	IIIB–IV	Erlotinib	Vandetanib
SHI [32]	2013	<i>Lancet Oncol</i>	2	IIIB–IV	Gefitinib	Icotinib

The studies are ordered by the date of inclusion of the first patient in the study. CT: chemotherapy. [#]: involvement of a pharmaceutical company was scored as 1 (no), 2 (yes, grant provided and company members as authors) or 3 (uncertain, grant provided but real role of company members not reported); [¶]: an oral fluoropyrimidine agent.

reported in the table). In none of the trials was it explicitly specified in the methods section whether the secondary objectives were analysed with a superiority assumption or not, except that by NATALE *et al.* [31], where all outcomes were first tested for superiority (noninferiority was to be tested only in the case of failure to detect improvement). Among the three phase II trials, two used response and one progression-free survival as the primary outcome. For the 17 remaining trials, 12 had overall survival (with a methodology focusing on 1-year survival in three), four analysed progression-free survival and one made use of response as the primary end-point. Planned interim analyses were not so infrequent (six (30%) trials) and their results were adequately taken into account at the time of the final analyses in five cases.

Specific aspects linked to the noninferiority hypothesis

Trial characteristics specific to the assumption of noninferiority are described in table 3. By rationale for noninferiority, we mean an explicit statement about an expected benefit on a secondary outcome and we found such a statement in only six trials. This expected benefit was in all cases reduced toxicity (plus, in one trial, ease of administration). It should be noted that two trials planned to test first for superiority (classical methodology and classical test based on the observed results) and then, in case of failure to detect superiority, for noninferiority. Three trials were designed to test first for the noninferiority hypothesis but calculated their sample size *a priori* with the alternative hypothesis of superiority of the experimental treatment, which decreases the power to detect true equality for both treatments (more details can be found in the online supplementary material). The method used to define the noninferiority margin was described in only six articles. In five studies, it was the effect-retention method (50% retention for four trials and proportion not specified in one). Use of the conventional method, on the basis of clinical reasoning and taking into account toxicity, was reported once, as was the use of guidelines from European Medicines Agency, without providing a reference to these guidelines [23]. In one study, two different margins were used and the effect-retention method was mentioned for the secondary definition of the margin. Except for the two trials that tested first for superiority and the trial in which sample size was not justified, all but one article reported an *a priori* sample size estimation that was consistent with the noninferiority margin. However, sufficient details allowing reproduction of the calculation of sample size were often missing: in six cases, the alternative hypothesis for determining statistical power was not reported and the targeted statistical power was not specified in one trial. When specified, the statistical power ranged from 75% to 90%. As most of the trials expressed the noninferiority margin in relative terms, we calculated the absolute difference accepted by the authors for concluding on noninferiority, using a summary parameter from the results (expected if reported or observed) in the control arm (table 3). When survival was the primary end-point, the median accepted absolute loss was 1.35 months (range 1.1–3.5 months).

Presentation of trial results

The results were presented using an ITT approach in 17 trials and using a PP approach in four (table 4). Both results were presented in two trials. In one trial, the results on the primary outcome were not presented. A benefit in a secondary outcome was documented in 12 trials. In total, for the 20 trials, 23 comparisons for noninferiority were presented (one trial had two experimental arms and one trial had three experimental arms). We compared the qualitative conclusions presented by the authors with the results, and we considered that the qualitative conclusions were consistent with the results and the trial design in 11 out of these 23 comparisons. The most frequent reason for inconsistency was claiming equivalence on the basis of absence of statistically significant difference, even if noninferiority was not demonstrated according to the *a priori* margin. For instance, in the study by MARUYAMA *et al.* [28], the primary end-point was overall survival and the noninferiority margin was defined as a hazard ratio ≤ 1.25 . The 95% confidence interval for the hazard ratio estimate was 0.89–1.40 (ITT analysis only), with an upper boundary > 1.25 . Although the authors correctly concluded that noninferiority could not be demonstrated, they also added “however, there was no statistically significant difference in overall survival”. This last statement suggests equivalence between both treatments, which is not correct. Another example of misleading conclusion can be found in the study by HANNA *et al.* [25]. The primary definition of noninferiority was a hazard ratio for overall survival ≤ 1.11 (primary analysis, conventional method) or ≤ 1.21 (secondary analysis, effect-retention method). The confidence interval obtained from the trial data was 0.80–1.20 (ITT analysis only). With the definition to be used for the primary analysis, the conclusion of noninferiority cannot be claimed. However, in the abstract, there was no mention of the noninferiority assumptions, no report of the hazard ratio estimate and a conclusion stating “treatment with pemetrexed resulted in clinically equivalent efficacy outcomes, but with significantly fewer side effects [...] and should be considered a standard treatment option”. This conclusion is not consistent with the trial design and is subject to false interpretation by the reader.

TABLE 2 General methodological characteristics of the trials

First author [ref.]	Phase	Primary outcome	Secondary outcomes	IA	Adjustment for IA
First-line setting					
NEGORO [13]	III	OS	Response Duration of response TTP Toxicity QoL	No	NA
ROSELL [14]	III	Response	Toxicity QoL OS	No	NA
FOSSELLA [15]	III	OS	Response Duration of response TTP Toxicity QoL	Yes	Reported
OHE [16]	III	1-year OS	Response Duration of response TTF TTP Toxicity QoL	No	NA
NOVELLO [17]	III	OS	Response rate 1-year OS Toxicity QoL	No	NA
PARK [18]	III	OS	Response TTP Toxicity QoL	No	NA
SCAGLIOTTI [19]	III	OS	Response Duration of response TTF TTP PFS Toxicity	No	NA
HEYMACH [20]	II	PFS	Response Disease control Toxicity OS	Yes	Not reported
MOK [21]	III	PFS	Response Reduction in symptoms Toxicity OS QoL	No	NA
OKAMOTO [22]	III	OS	Response PFS Toxicity QoL	Yes	Reported
LEE [23]	IIB	Response	PFS Toxicity OS	No	NA
Second-line setting					
SCHUETTE [24]	III	1-year OS	Response TTP Toxicity QoL	No	NA
HANNA [25]	III	OS	Response Duration of CB TTF TTP PFS QoL	No	NA

Continued

TABLE 2 Continued

First author [ref.]	Phase	Primary outcome	Secondary outcomes	IA	Adjustment for IA
RAMLAU [26]	III	1-year OS	Response Duration of response TTR TTP Toxicity OS QoL	No	NA
KRZAKOWSKI [27]	III	PFS	Drug exposure Response CB Toxicity OS QoL	No	NA
MARUYAMA [28]	III	OS	PFS Toxicity QoL	Yes	Reported
KIM [29]	III	OS	Response PFS Toxicity QoL	Yes	Reported
LI [30]	II [#]	Response	Disease control Performance index scores Toxicity	No	NA
NATALE [31]	III	PFS	Response Time to symptom deterioration	Yes	Reported
SHI [32]	III	PFS	Response TTP Toxicity OS QoL	No	NA

IA: interim analysis; OS: overall survival; TTP: time to progression; QoL: quality of life; NA: not applicable; TTF: time to treatment failure; PFS: progression-free survival; CB: clinical benefit; TTR: time to response. #: stated as exploratory.

Qualitative assessment

The assessment of the CONSORT items is presented in table 5 (frequency tabulations). The general mean score was significantly higher (2.3) than the mean score for extended items (2.0) (p<0.001, Wilcoxon’s rank-signed test). Among the general items that were poorly described (median <2) were the items related to the randomisation procedure (generation of the allocated arm, allocated treatment concealment and implementation of the randomisation procedure), the item about the availability of the trial in a registry and that about the public availability of the protocol. In the extended items, the lack of clarity about the noninferiority hypothesis in the title, the lack of specification of the rationale for noninferiority and the absence of an explicit statement about patient selection (are they similar to those included in superiority trials?) were the main drawbacks. It should be stressed that the CONSORT checklist does not allow documentation of the lack of transparency of the *a priori* sample size estimation as the only requirement is to report a sample size estimation calculated using the noninferiority assumption and on the noninferiority margin. We did not identify, for any of the covariates that we pre-specified, any significant association with the general mean score (all p>0.05, Mann–Whitney tests). The same applies to the extended items. For the risk of bias tool, we only identified a higher risk when the primary outcome was different from overall survival, higher risk being explained by the lack of blinding in all the trials but one. The individual results for the risk of bias tool assessment are shown in table 6.

Discussion

The use of noninferiority designs is a recent development and, up to now, has remained quite modest in patients with advanced NSCLC. Indeed, a recent review by SACHER *et al.* [2] identified 118 randomised phase III clinical trials (targeting the same populations) published during 2001–2010 while, in the present review, we selected 15 such studies during the same period (13%, 95% CI 8–20%). The authors, although

TABLE 3 Trial characteristics specific to the assumption of noninferiority (NI)

First author [ref.]	Rationale	Measure	Margin of NI	Choice of margin	Consistency of sample size estimation	Power estimation	Power %	Planned number per arm	E/O in control arm	Absolute decrease accepted
First-line setting										
NEGORO [13]	No	Hazard ratio	≤1.33	NR	Yes	Superiority of experimental arm	81	81 events	Median OS 35 weeks (E)	7 weeks
ROSELL [14]	No	Rates difference (E–C)	>–10%	NR	Yes	Equivalence	80 [#]	260	Response rate 30% (E)	10%
FOSSELLA [15]	No	Hazard ratio	≥0.85	ERM (50%)	No [¶]	NR	85	360	Median OS 8 months (E)	1.2 months
OHE [16]	No	Rates difference (E–C)	>–10%	NR	Yes	Superiority of experimental arm	80	135	1-year survival 43% (E)	10%
NOVELLO [17]	No	Hazard ratio	≤1.33	CM	Yes	Equivalence	80 [#]	110 events	1-year survival 33% (E)	1.9 months
PARK [18]	Yes	Rates difference (E–C)	>–15%	NR	Yes	NR	80	109	Median OS 6.2 months (E)	15%
SCAGLIOTTI [19]	No	Hazard ratio	<1.176	NR	Yes	Equivalence	80	595 events	Median OS 10.3 months (O)	1.5 months
HEYMACH [20]	No	Hazard ratio	NR	NR	Yes	NR	75 ⁺	NR	Median PFS 11.5 weeks (E)	NC
MOK [21]	No	Hazard ratio	≤1.20	NR	Yes	Equivalence	80	472 events	NR (NPH)	NA
OKAMOTO [22]	Yes	Hazard ratio	≤1.33	NR	Yes	Equivalence	85	280	Median OS 14 months (E)	3.5 months
LEE [23]	No	Rates difference (E–C)	>–15%	Other	Yes	Equivalence	80	143	Response rate 37% (E)	15%
Second-line setting										
SCHUETTE [24]	Yes	Rates difference (E–C)	>–2%, <2% [§]	NR	Unclear	NR	80	102	Unclear	2%
HANNA [25]	Yes	Hazard ratio	≤1.11	NR	Yes	Hazard ratio 0.83	81%	NR	Median OS: 7.9 months (O)	0.7 months
			≤1.21	ERM (50%)		NA	NA			
RAMLAU [26]	No	Rates difference (E–C)	>–10%	NR	Yes	Equivalence	81	380	1-year survival 37%	10%
KRZAKOWSKI [27]	No	Hazard ratio	≥0.755	NR	Yes	NR	80	275	Median PFS 2.3 months (O)	0.6 months
MARUYAMA [28]	No	Hazard ratio	≤1.25	NR	Yes	Superiority of experimental arm	90	242	Median OS 12 months (E)	2.4 months
KIM [29]	Yes	Hazard ratio	≤1.154	ERM (50%)	Yes	NR	NR	585 events	Median OS: 8 months (O)	1.1 months
LI [30]	Yes	Rates difference (E–C)	>–10%	NR	NR	NR	NR	NR	Response rate 4.9% (O)	10%
NATALE [31]	No	Hazard ratio	NR	ERM (50%)	No [¶]	NR	NR	555 events	Median PFS 2.2 months	0.6 months

Continued

TABLE 3 Continued

First author [ref.]	Rationale	Measure	Margin of NI	Choice of margin	Consistency of sample size estimation	Power estimation	Power %	Planned number per arm	E/O in control arm	Absolute decrease accepted
Shi [32]	No	Hazard ratio	<1.14	ERM	Yes	NR	80	200	Median PFS 4.2 months	0.5 months

Rationale: was the rationale for NI specified? We assessed whether it was explicitly specified in the methods section (yes or no). Consistency of sample size estimation: with the NI margin; Power: targeted statistical power; Power estimation: choice of the alternative hypothesis for power estimation. For further details of the margin of NI and method for choosing the margin, and power estimation, see the online supplementary material. E: expectation; O: observation; NR: not reported; OS: overall survival; E-C: experimental arm minus control; ERM: effect-retention method; CM: conventional method; PFS: progression-free survival; NC: not calculated (margin of NI not specified); NPH: nonproportional hazards. NA: not applicable.[#]: α -level 10% (two-sided); [¶]: trial designed for superiority, noninferiority to be tested only in case of failure to detect superiority; ^{*}: α -level 20% (two-sided); [§]: equivalence trial.

not addressing the issue of noninferiority assumption, reported a significant shift in the design and interpretation of trials with an increasing use of progression-free survival as primary outcome, despite the subjectivity of its assessment and despite the lack of demonstrated surrogacy for overall survival. They also raised concerns because of the increasing rate of inadequate result interpretation: indeed, they observed, more frequently in recent years than before, trials considered positive despite not reaching statistical significance on their primary outcome, on the basis of arguments related to secondary outcomes. A benefit observed in a secondary outcome without reaching the pre-defined difference on the primary outcome cannot be considered sufficient evidence to claim that the experimental arm is to be recommended.

To conclude about the value of an experimental treatment on the basis of benefits on other outcomes without a positive, statistically significant result on the primary outcome is, in fact, the aim of noninferiority trials. This could be viewed as a better approach than wrongly interpreting the results of superiority trials. Indeed, in a noninferiority trial, the definition of noninferiority is decided upfront with the choice of the noninferiority margin and the conclusion should be reached only if the noninferiority is statistically proven according to the design of the trial and the *a priori* assumptions. However, we found comparable methodological flaws in the present review, with about half of the conclusions (11 out of 23) not being consistent with trial results or design. As the absence of a statistically significant difference in a superiority trial cannot be interpreted as equivalence between the standard-care and experimental arm [11], failure to demonstrate noninferiority cannot be shifted to an interpretation of “similar efficacy” even if the point estimates for the primary end-point appear to be close. Indeed, when the experimental treatment is not “noninferior” to the standard one, the confidence interval for the treatment effect includes the value of the treatment effect corresponding to the noninferiority margin and the trial results are, therefore, statistically compatible with the inferiority of the experimental arm (table 7).

The issue of defining noninferiority is complex and, although debated in the literature [3], there is no consensus about the best method to be used. In the trials analysed here, the choice of the noninferiority margin is neither described nor discussed and the noninferiority margin itself is most often defined in relative terms, preventing the reader from interpreting the translation of this definition into absolute differences. We calculated these absolute differences and we believe that, in some cases, the price to pay to get a secondary possible benefit is potentially as high as losing 2.4 months from 12 months [28] or 3.5 months from 14 months [22] of median survival. In one trial, investigators even accepted an absolute loss of 10% in response starting from an absolute rate of 5% [30]: with such an assumption, no treatment could be definitely noninferior and also less toxic, and there is no need to conduct a clinical trial to achieve this result.

There is also a clear lack of proper documentation of why the investigators did not have the “ambition” to identify experimental treatments superior to the control arm but choose to attempt to demonstrate noninferiority. The expected benefits are rarely explicitly mentioned and, if so, the hypotheses on these benefits are never quantified. The planned switch from superiority to noninferiority that we observed in two trials in our review is of particular concern because there should be no other justification for noninferiority than the failure to detect superiority of the experimental arm. The use of the superiority of the experimental arm as alternative hypothesis for power calculation (at least in three trials out of 20), which is in contradiction with the noninferiority assumption as it consists of using a true situation different from the hypothesis, is also worrying. Some authors argued that there might be some pressure on

TABLE 4 Results and conclusion

First author [ref.]	ITT	PP	Benefit on another outcome	Conclusion	Conclusion consistent with results?
First-line setting					
NEGORO [13]	Hazard ratio 0.83, 95% CI 0.64 to 1.09	NR	No	E not inferior to C	Yes
ROSELL [14]	Difference in rates (E–C) 3%, 90% CI NR	Difference in rates (E–C) 3%, 90% CI –10 to 3.4%	No	E not inferior to C for response but C superior to E for OS	Yes
FOSSELLA [15]	C versus E1: hazard ratio 1.18; 97.2% CI 0.99 to 1.42	NR	Yes	Superiority of E1 (abstract) NI of E1 (discussion)	No for E1 (superiority cannot be claimed)
	C versus E2: hazard ratio 1.05, 97.2% CI 0.88 to 1.25	NR		Failure of NI for E2	Yes for E2
OHE [16]	NR	Difference in rates (E–C) –8.2%, 95% CI –19.6 to 3.3%	No	The four regimens have similar efficacy	No
	NR	Difference in rates (E–C) 0.4%, 95% CI –10.0 to 11.7%			
	NR	Difference in rates (E–C) –10.9%, 95% CI –22.3 to 0.5%			
NOVELLO [17]	Hazard ratio 1.21, 90% CI 0.97 to 1.51	NR	Yes	Failure of NI for E	Yes
PARK [18]	Difference in rates (E–C) 3.4%, 95% CI –8 to 14%	NR	Yes	E not inferior to C but TTP better with C	Yes
SCAGLIOTTI [19]	Hazard ratio 0.94, 95% CI 0.84 to 1.05	NR	Yes	E provides similar efficacy with better tolerability	No, only NI can be claimed
HEYMACH [20]	Hazard ratio 1.26, 95% CI 0.83 to 1.91	NR	NA	Early closure: n=40 (C), n=73 (E); failure to detect NI	Yes
MOK [21]	Hazard ratio 0.74, 95% CI 0.65 to 0.85	NR	Yes	E superior to C	Yes (closed testing procedure)
OKAMOTO [22]	Hazard ratio 0.928, 99.2% CI 0.67 to 1.28 at IA	NR	Yes	E not inferior to C	Yes
LEE [23]	Difference in rates (E–C) 1.7%, 95% CI –10 to 13.3%	NR	No	E well tolerated and not inferior	Yes (higher neutropenia with E, despite wrong statement about 95% CI)
Second-line setting					
SCHUETTE [24]	NR	NR	Yes	E provides similar efficacy and better tolerability and is a feasible alternative	No, NR as an equivalence trial
HANNA [25]	Hazard ratio 0.99, 95% CI 0.80 to 1.20	NR	Yes	E equivalent to C, should be a new standard	No, neither equivalence nor NI can be claimed
RAMLAU [26]	Difference in rates (E–C) –3.6%, 95% CI –9.6 to 2.5%	NR	No	E provides activity and may be an option	No, no reference in the abstract to NI (indeed demonstrated)
KRZAKOWSKI [27]	Hazard ratio 1.004, 95% CI 0.84 to 1.12	NR	No	NI demonstrated, E is an option	No, as higher toxicities with E
MARUYAMA [28]	Hazard ratio 1.12, 95% CI 0.89 to 1.40	NR	Yes	NI not demonstrated; however, no differences	No, trial not designed to show a difference between arms
KIM [29]	Hazard ratio 1.02, 96% CI 0.901 to 1.143	Hazard ratio 1.02, 96% CI 0.905 to 1.150	Yes	NI demonstrated, E is a valid option	Yes
LI [30]	NR	Difference in rates (E–C) 4.5%, 95% CI –2.5 to NR	Yes	Equivalent efficacy outcomes and better safety profile for E	No

Continued

TABLE 4 Continued

First author [ref.]	ITT	PP	Benefit on another outcome	Conclusion	Conclusion consistent with results?
NATALE [31]	Hazard ratio 0.98, 95.22% CI 0.87 to 1.10	NR	No	No efficacy advantage and higher toxicity	No, as NI is demonstrated by design
SHI [32]	Hazard ratio 0.84, 95% CI 0.67 to 1.05	NR	Yes	Icotinib could be a new treatment option	Yes, although reference to NI missing

ITT: intention to treat; PP: per protocol; NR: not reported;. E: experimental arm; C: control arm; OS: overall survival; NI: noninferiority; TTP: time to progression; IA: interim analysis.

investigators to produce significant results (which lead to better publications) or influence from industry (often funding trials) willing to conclude positively on the trials results [2]. A pharmaceutical company was involved in all the trials included in our review, including at least funding but also often authorship for company employees. The exact role of company employees is seldom described but authorship should mean a significant contribution to design, data analysis or data interpretation.

Progression-free survival or, even more, response as primary objective is a nonmarginal choice (eight out of 20 trials) and, in our opinion, should be avoided in the context of noninferiority trials. Its use in superiority trials is debated namely due to its subjectivity, to the influence of the intervals between assessments and to the lack of impact on survival in the case of small benefits [33]. The impact of a noninferior progression-free survival on overall survival should be further studied using retrospective studies as it has been shown that, in the population of patients with advanced NSCLC, large benefits are needed in progression-free survival to warrant a (small) benefit in survival [33]. This justifies the need to study the impact on survival of a noninferior outcome in progression-free survival. Furthermore, the practical difficulty of blinding chemotherapy regimens for patients and investigators induces a higher risk of bias when using progression-free survival as outcome.

There were three phase II trials among those eligible for our review. Accepting noninferiority in early development of a drug/regimen is even more intriguing and challenging than for phase III trials, and should imply a much more careful justification of the rationale for this choice (anecdotally, this was not given for two of these three trials).

We showed that the quality of reporting of noninferiority trials is worse for the items linked to noninferiority than for the general aspects, as measured by the CONSORT scores. Our review also identified some deficiencies in the reporting of the determination of the sample size; reports were often not detailed enough to be reproducible. The important recommendation to carry out both ITT and PP analyses is, in practice, seldom followed. Indeed, contrary to superiority trials, in noninferiority trials, the PP analysis is important to make sure that there was no dilution in the effect of the experimental treatment in the ITT population (leading to a overoptimistic conclusion of noninferiority) [7].

In conclusion, we believe that noninferiority trials for advanced NSCLC might be misleading and their results have little value for clinical practice. The situation portrayed by ELIE *et al.* [7], who see a place for noninferiority trials when standard treatments have very high efficacy that cannot be improved any more meaningfully, is far from being common today in the population of patients with advanced NSCLC. Thus, the choice of a noninferiority design for future trials should be restricted to carefully selected situations. Such a situation was reached in the trial by PARK *et al.* [18] in which the question was about the number of cycles required, with an obvious hope of reducing toxicity. If this choice is made, as already stated by TANAKA *et al.* [3], improvement in methodology should be achieved by a full description of the choice of the noninferiority margin (the effect-retention method should be favoured as it is the only one warranting that the experimental treatment is better than the previous standard), clear hypotheses about expected benefits in secondary outcomes, conduct of both ITT and PP analyses, and strict interpretation of results according to the design and the pre-specified hypotheses.

Our key messages are the following. When an experimental treatment is shown to be noninferior to the standard one, the interpretation of this result should be careful and take into account the noninferiority margin, which represents the *a priori* acceptable loss in efficacy: noninferiority does not mean that there is no difference between the experimental approach and the standard one. The observed benefit(s) on other

TABLE 5 Assessment of CONSORT items

Item	General items			Extended items		
	Not reported	Partially reported	Reported	Not reported	Partially reported	Reported
1 Title and abstract						
1a Title	45	0	55	95	0	5
1b Abstract	5	35	60	45	55	0
2 Background and objectives						
2a Background	0	10	90	65	25	10
2b Objectives	10	45	45	5	85	10
3 Trial design						
3a Description	0	0	100			
3b Changes during the trial						
4 Participants						
4a Eligibility criteria	0	0	100	95	5	0
4b Settings and locations	25	75	0			
5 Interventions	0	0	100	5	0	95
6 Outcomes						
6a Definitions	10	45	45	20	70	10
6b Changes during trial						
7 Sample size						
7a Estimation	5	45	50	11 [#]	32 [#]	58 [#]
7b Interim analyses	33 [¶]	33 [¶]	33 [¶]	0 [¶]	33 [¶]	67 [¶]
8 Sequence generation						
8a Method	55	5	40			
8b Type and restrictions	55	10	35			
9 Allocation concealment	65	15	20			
10 Implementation	70	20	10			
11 Blinding						
11a Who?				Not reported as only one trial was blinded		
11b Similarity of interventions						
12 Statistical methods						
12a Methods	5	25	70	10	5	85
12b Subgroup analyses						
13 Participants flow						
13a Flow chart	5	15	80			
13b Reasons for losses and exclusion	10	25	65			
14 Recruitment						
14a Accrual period and follow-up	0	20	80			
14b Reason for stopping	5	20	75			
15 Baseline data	0	5	95			
16 Numbers analysed	5	10	85			
17 Outcomes and estimation						
17a Estimated effect and confidence interval	0	50	50	25	30	45
17b Absolute and relative effect sizes (binary outcomes)	15	85	0			
18 Ancillary analyses	6 ⁺	31 ⁺	63 ⁺			
19 Harms	0	0	100			
20 Limitations	10	50	40			
21 Generalisability	0	80	20			
22 Interpretation	10	30	60	15	55	30
23 Registration	70	0	30			
24 Protocol	95	0	5			
25 Funding	30	35	35			

Data are presented as %. CONSORT: Consolidated Standards of Reporting Trials. #: not applicable to one article; ¶: not applicable to 14 articles; +: not applicable to four articles.

end-points, especially those justifying the noninferiority design (if any) should be integrated in the conclusion too. However, the expected *a priori* benefits are seldom explicitly stated, which is a bad practice that needs to be improved as well as the reporting on methodological aspects of noninferiority trials. The impact of those trials on clinical practice should be low.

TABLE 6 Results of the risk of bias assessment

First author [ref.]	Random sequence generation	Allocation concealment	Blinding of participants	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Number of HR items
First-line setting							
NEGORO [13]	LR	LR	LR	LR	LR	LR	0
ROSELL [14]	LR	LR	HR	HR	?	LR	3
FOSSELLA [15]	?	LR	LR	LR	LR	LR	1
OHE [16]	?	LR	LR	LR	HR	LR	1
NOVELLO [17]	?	?	LR	LR	LR	LR	2
PARK [18]	LR	?	LR	LR	LR	LR	1
SCAGLIOTTI [19]	LR	?	LR	LR	LR	LR	1
HEYMACH [20]	?	?	LR	?	LR	LR	3
MOK [21]	LR	?	?	HR	LR	LR	3
OKAMOTO [22]	LR	LR	LR	LR	LR	LR	0
LEE [23]	?	?	LR	?	LR	LR	3
Second-line setting							
SCHUETTE [24]	?	?	LR	LR	LR	?	3
HANNA [25]	?	?	LR	LR	LR	LR	2
RAMLAU [26]	?	?	LR	LR	LR	LR	2
KRZAKOWSKI [27]	?	?	LR	?	LR	LR	3
MARUYAMA [28]	?	?	LR	LR	LR	LR	2
KIM [29]	LR	LR	LR	LR	LR	LR	0
LI [30]	?	?	HR	HR	HR	LR	5
NATALE [31]	?	?	LR	?	LR	LR	3
SHI [32]	LR	LR	LR	LR	LR	LR	0

LR: low risk of bias; ?: unclear risk of bias; HR: high risk of bias.

TABLE 7 Interpretation of treatment effect and of the 95% confidence interval for the true treatment effect

Scenario	Treatment effect estimate	Lower bound of 95% CI	Upper bound of 95% CI	Comment
1	<0	<0	<0 and < Δ	Noninferiority can be claimed
2	<0	<0	>0 and < Δ	Noninferiority can be claimed
3	0	<0	>0 and < Δ	Noninferiority can be claimed
4	>0 and < Δ	>0 and < Δ	>0 and < Δ	Quite infrequent, noninferiority can be claimed (by design)
5	<0	<0	> Δ	Failure to show noninferiority
6	>0	<0	> Δ	Failure to show noninferiority
7	>0 and < Δ	>0	> Δ	Failure to show noninferiority
8	> Δ	> Δ	> Δ	E inferior to C

Treatment effect is expressed as control arm (C) minus the experimental arm (E) with C-E<0 favouring the experimental arm and C-E>0 favouring the control arm. Δ is the noninferiority margin and is assumed to be >0.

References

- 1 Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012; 62: 10–29.
- 2 Sacher AG, Le LW, Leighl NB. Shifting patterns in the interpretation of phase III clinical trial outcomes in advanced non-small-cell lung cancer: the bar is dropping. *J Clin Oncol* 2014; 32: 1407–1411.
- 3 Tanaka S, Kinjo Y, Kataoka Y, et al. Statistical issues and recommendations for noninferiority trials in oncology: a systematic review. *Clin Cancer Res* 2012; 18: 1837–1847.
- 4 Rothmann M, Li N, Chen G, et al. Design and analysis of non-inferiority mortality trials in oncology. *Stat Med* 2003; 22: 239–264.
- 5 US Department of Health and Human Services, Food and drug administration, Center for Drug Evaluation and Research, et al. Guidance for industry. Non-inferiority clinical trials. www.FDA.Gov/downloads/drugs/guidancecomplianceregulatoryinformation/ucm202140.pdf Date last updated: March 2010.
- 6 European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline of the choice of non-inferiority margin. www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf Date last updated: July 17, 2005.

- 7 Elie C, De Rycke Y, Jais JP, *et al.* Aspects méthodologiques et statistiques des essais d'équivalence et de non-infériorité [Methodological and statistical aspects of equivalence and non inferiority trials]. *Rev Epidemiol Sante Publique* 2008; 56: 267–277.
- 8 Piaggio G, Elbourne DR, Pocock SJ, *et al.* Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012; 308: 2594–2604.
- 9 Riechelmann RP, Alex A, Cruz L, *et al.* Non-inferiority cancer clinical trials: scope and purposes underlying their design. *Ann Oncol* 2013; 24: 1942–1947.
- 10 Saad ED, Buyse M. Non-inferiority trials in breast and non-small cell lung cancer: choice of non-inferiority margins and other statistical aspects. *Acta Oncol* 2012; 51: 890–896.
- 11 Moher D, Hopewell S, Schulz KF, *et al.* CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869.
- 12 Higgins JP, Altman DG, Gøtzsche PC, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343: d5928.
- 13 Negoro S, Masuda N, Takada Y, *et al.* Randomised phase III trial of irinotecan combined with cisplatin for advanced non-small-cell lung cancer. *Br J Cancer* 2003; 88: 335–341.
- 14 Rosell R, Gatzemeier U, Betticher DC, *et al.* Phase III randomised trial comparing paclitaxel/carboplatin with paclitaxel/cisplatin in patients with advanced non-small-cell lung cancer: a cooperative multinational trial. *Ann Oncol* 2002; 13: 1539–1549.
- 15 Fossella F, Pereira JR, von Pawel J, *et al.* Randomized, multinational, phase III study of docetaxel plus platinum combinations versus vinorelbine plus cisplatin for advanced non-small-cell lung cancer: the TAX 326 study group. *J Clin Oncol* 2003; 21: 3016–3024.
- 16 Ohe Y, Ohashi Y, Kubota K, *et al.* Randomized phase III study of cisplatin plus irinotecan *versus* carboplatin plus paclitaxel, cisplatin plus gemcitabine, and cisplatin plus vinorelbine for advanced non-small-cell lung cancer: Four-Arm Cooperative Study in Japan. *Ann Oncol* 2007; 18: 317–323.
- 17 Novello S, Bruzzi P, Barone C, *et al.* Phase III study in stage IV non-small-cell lung cancer patients treated with two courses of cisplatin/gemcitabine followed by a randomization to three additional courses of the same combination or gemcitabine alone. *Ann Oncol* 2007; 18: 903–908.
- 18 Park JO, Kim SW, Ahn JS, *et al.* Phase III trial of two versus four additional cycles in patients who are nonprogressive after two cycles of platinum-based chemotherapy in non small-cell lung cancer. *J Clin Oncol* 2007; 25: 5233–5239.
- 19 Scagliotti GV, Parikh P, von Pawel J, *et al.* Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naïve patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol* 2008; 26: 3543–3551.
- 20 Heymach JV, Paz-Ares L, De Braud F, *et al.* Randomized phase II study of vandetanib alone or with paclitaxel and carboplatin as first-line treatment for advanced non-small-cell lung cancer. *J Clin Oncol* 2008; 26: 5407–5415.
- 21 Mok TS, Wu YL, Thongprasert S, *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009; 361: 947–957.
- 22 Okamoto I, Yoshioka H, Morita S, *et al.* Phase III trial comparing oral S-1 plus carboplatin with paclitaxel plus carboplatin in chemotherapy-naïve patients with advanced non-small-cell lung cancer: results of a west Japan oncology group study. *J Clin Oncol* 2010; 28: 5240–5246.
- 23 Lee SY, Park HS, Lee KY, *et al.* Paclitaxel-loaded polymeric micelle (230 mg/m²) and cisplatin (60 mg/m²) *vs.* paclitaxel (175 mg/m²) and cisplatin (60 mg/m²) in advanced non-small-cell lung cancer: a multicenter randomized phase IIB trial. *Clin Lung Cancer* 2013; 14: 275–282.
- 24 Schuette W, Nagel S, Blankenburg T, *et al.* Phase III study of second-line chemotherapy for advanced non-small-cell lung cancer with weekly compared with 3-weekly docetaxel. *J Clin Oncol* 2005; 23: 8389–8395.
- 25 Hanna N, Shepherd FA, Fossella FV, *et al.* Randomized phase III trial of pemetrexed versus docetaxel in patients with non-small-cell lung cancer previously treated with chemotherapy. *J Clin Oncol* 2004; 22: 1589–1597.
- 26 Ramlau R, Gervais R, Krzakowski M, *et al.* Phase III study comparing oral topotecan to intravenous docetaxel in patients with pretreated advanced non-small-cell lung cancer. *J Clin Oncol* 2006; 24: 2800–2807.
- 27 Krzakowski M, Ramlau R, Jassem J, *et al.* Phase III trial comparing vinflunine with docetaxel in second-line advanced non-small-cell lung cancer previously treated with platinum-containing chemotherapy. *J Clin Oncol* 2010; 28: 2167–2173.
- 28 Maruyama R, Nishiwaki Y, Tamura T, *et al.* Phase III study, V-15-32, of gefitinib versus docetaxel in previously treated Japanese patients with non-small-cell lung cancer. *J Clin Oncol* 2008; 26: 4244–4252.
- 29 Kim ES, Hirsh V, Mok T, *et al.* Gefitinib *versus* docetaxel in previously treated non-small-cell lung cancer (INTEREST): a randomised phase III trial. *Lancet* 2008; 372: 1809–1818.
- 30 Li R, Sun L, Wang J, *et al.* Pemetrexed *versus* docetaxel in second line non-small-cell lung cancer: results and subsets analyses of a multi-center, randomized, exploratory trial in Chinese patients. *Pulm Pharmacol Ther* 2012; 25: 364–370.
- 31 Natale RB, Thongprasert S, Greco FA, *et al.* Phase III trial of vandetanib compared with erlotinib in patients with previously treated advanced non-small-cell lung cancer. *J Clin Oncol* 2011; 29: 1059–1066.
- 32 Shi Y, Zhang L, Liu X, *et al.* Icotinib *versus* gefitinib in previously treated advanced non-small-cell lung cancer (ICOGN): a randomised, double-blind phase 3 non-inferiority trial. *Lancet Oncol* 2013; 14: 953–961.
- 33 Laporte S, Squifflet P, Baroux N, *et al.* Prediction of survival benefits from progression-free survival benefits in advanced non-small-cell lung cancer: evidence from a meta-analysis of 2334 patients from 5 randomised trials. *BMJ Open* 2013; 3: e001802.