First-line treatment of EGFR-mutated nonsmall cell lung cancer: critical review on study methodology

Martin Sebastian[1], Alexander Schmittel[2] and Martin Reck[3]

Affiliations: [1]Dept of Hematology/Oncology, University Hospital Frankfurt, Frankfurt, [2]Hematology, Oncology, Infectious Diseases, Aerzteforum Seestrasse, Medizinisches Versorgungszentrum, Berlin, and [3]LungClinic Grosshansdorf, Member of the German Centre for Lung Research (DZL), Grosshansdorf, Germany.

Correspondence: M. Sebastian, Dept of Hematology/Oncology, University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Frankfurt, Germany. E-mail: martin.sebastian@kgu.de

ABSTRACT Recent advances in understanding the mechanisms of nonsmall cell lung cancer (NSCLC) has led to the development of targeted treatments, including the reversible epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors gefitinib and erlotinib, and the irreversible ErbB family blocker afatinib. Several important activating EGFR mutations have now been identified, which correlate strongly with response to treatment with these agents. Multiple randomised controlled trials have confirmed the association between the presence of activating EGFR mutations and objective response to gefitinib, erlotinib and afatinib, thus demonstrating their superiority over platinum-based chemotherapy as first-line treatment for NSCLC patients with EGFR mutation-positive tumours, and resulting in approval of these agents for use in this setting. It can be tempting to compare outcome data across multiple clinical trials and agents; however, substantial differences in methodology between studies, including investigator *versus* independent assessment and differences in patient eligibility, makes such comparisons fraught with difficulty. This critical review provides an overview of the evolution of the methodology used in eight phase III trials investigating first-line targeted treatment of NSCLC, identifies key differences in methodology and reporting, and critically assesses how these differences should be taken into account when interpreting the findings from such trials.

@ERSpublications
First-line treatment of EGFR-mutated NSCLC: differences in study methodology limit data comparisons http://ow.ly/sBvso

Introduction

Despite recent advances in therapy for advanced lung adenocarcinoma, there continues to be an unmet medical need for effective treatment of stage IIIb/IV nonsmall cell lung cancer (NSCLC). In recent years, our understanding of the mechanisms of this disease has substantially increased in parallel with the development of the reversible epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) gefitinib and erlotinib, and afatinib, which binds irreversibly to EGFR as well as to the other members of the ErbB family.

Early trials with gefitinib and erlotinib revealed subsets of patients achieving prolonged responses to treatment not seen with standard chemotherapy [1, 2]. Females, nonsmokers, Japanese patients and patients with lung adenocarcinoma were found to have higher response rates than patients who were of European

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

Received: Nov 29 2013 | Accepted after revision: Dec 20 2013

Conflict of interest: Disclosures can be found alongside the online version of this article at err.ersjournals.com

Provenance: Publication of this peer-reviewed article was supported by Boehringer Ingelheim, Germany (article sponsor, *European Respiratory Review* issue 131).

origin, male, smokers or who had other NSCLC histology [3–5]. In light of this, further investigations identified several important activating EGFR mutations occurring in specific patient types that correlate strongly with response to gefitinib and erlotinib treatment [2, 6, 7].

The first randomised clinical trial to specifically compare EGFR TKI therapy with chemotherapy in patients with EGFR mutation-positive tumours was IPASS (Iressa Pan-Asia Study) [8]. In East Asian patients with stage IIIB/IV lung adenocarcinoma who never smoked tobacco (or only smoked lightly), initial treatment with an EGFR TKI was found to be superior to standard platinum-based chemotherapy [8]. Patients with EGFR mutation-positive tumours achieved significantly longer progression-free survival (PFS) with gefitinib *versus* those receiving chemotherapy (hazard ratio (HR) for progression or death 0.48 (95% CI 0.36–0.64); $p<0.001$) [8].

Subsequently, a number of trials, published within a relatively short period, have specifically addressed first-line treatment options in patients with stage IIIb/IV NSCLC and suspected or known EGFR mutations, confirming the association between the presence of activating EGFR mutations and objective response to gefitinib, erlotinib and afatinib. These trials are as follows: EURTAC (European Randomised Trial of Tarceva *versus* Chemotherapy) [9], OPTIMAL [10, 11], NEJ002 (North East Japan 002) [12], West Japan Thoracic Oncology Group (WJTOG) 3405 [13], IPASS [8, 14], LUX-Lung 3 [15], LUX-Lung 6 [16] and ENSURE [17]. As a result, erlotinib, gefitinib and, most recently, afatinib have received approval for first-line treatment of EGFR mutation-positive NSCLC [18–20]. Furthermore, recognition of the significance of acquired genetic mutations in therapeutic targets, including EGFR, has led to alterations in the NSCLC treatment paradigm, with upfront molecular testing for EGFR and other mutations now recommended [21].

Although results from these trials are frequently compared, it is important to recognise that direct comparisons do not take into account substantial differences in trial methodology, *e.g.* mutation testing, assessment of progression (independent *versus* investigator) and differences in patient inclusion criteria. For example, inclusion of local populations and differences in EGFR mutation status have the potential to impact on extrapolation of the findings and the generalisability of the conclusions, and may, therefore, have a bearing on regulatory processes and approval. Furthermore, differences in trial documentation can impact on the utility of trial data.

The objective of this review is to provide an overview of the evolution of methodology of phase III trials investigating first-line treatment of NSCLC over time, and to assess how differences in methodology should be taken into account when interpreting the findings from such trials. The results of the chemotherapy arms are not discussed extensively because all trials concluded that in patients with EGFR mutation-positive NSCLC, the chemotherapy comparator was inferior.

Methodology

Clinical trials were searched using www.ClinicalTrials.gov and www.citeline.com. The results of identified trials were obtained *via* PubMed, American Society of Clinical Oncology and European Society for Medical Oncology/European CanCer Organisation supplements, and World Conference on Lung Cancer abstracts. Where possible, data from fully published peer-reviewed literature were included. However, for more recent studies, LUX-Lung 6 and ENSURE [16, 17], data were only available in abstract, poster or presentation form. Phase III trials that investigated the first-line treatment of patients with stage IIIb/IV NSCLC were included. This was further restricted to trials that compared EGFR TKI monotherapy with standard platinum-based chemotherapy. Each of the trials had TKI and chemotherapy comparator arms. The comparators are relevant to understanding the trial methodology and generalisability of the study results and are, therefore, included in the methodological comparison. Phase III studies meeting these criteria, but where only a subpopulation of patients were EGFR mutation-positive, were included providing efficacy data reported for the EGFR mutation-positive subgroup were sufficient for comparison with other studies. Phase III studies meeting these criteria, but where only a subpopulation of patients had stage IIIb/IV NSCLC, were also included.

For a qualitative comparison of the studies, results were analysed using the CONSORT (Consolidated Standards of Reporting Trials) criteria (table S1) [22]. For conciseness, not all CONSORT criteria are discussed in full for all studies. For instance, the CONSORT criteria require that all trial protocols be made publically available in a registration database (*e.g.* www.ClinicalTrials.gov); however, since not all study protocols were available, no attempt was made to systematically retrieve study details from this source. Other CONSORT criteria, *e.g.* eligibility criteria, were reviewed as published but are not repeated here in detail. Only differences that were considered of relevance to the interpretation and comparison of study findings are discussed.

Quantitative and qualitative analyses of phase III first-line trials in EGFR mutation-positive NSCLC

Identification of trials

Nine trials were identified initially. Of these, eight had PFS as the primary end-point: NEJ002 [12, 23], WJTOG3405 [13], IPASS [8, 14], EURTAC [9], LUX-Lung 3 [15], OPTIMAL [10, 11], LUX-Lung 6 [16], and ENSURE [17] (table 1). The First-SIGNAL trial [27] (www.ClincalTrials.gov identifier NCT00455936) was also identified, but differed from the other trials in that the primary end-point was overall survival. The First-SIGNAL trial was conducted exclusively in South Korea and investigated first-line gefitinib *versus* gemcitabine-cisplatin in never-smokers with lung adenocarcinoma (stage IIIb/IV). Only 42 (14%) out of 313 patients were EGFR mutation-positive. As a result, the published results for the EGFR mutation-positive subgroup, especially for the secondary end-points of PFS and overall response rate, were limited, compromising comparison with other trials.

Qualitative analyses of the trials

Six studies were conducted in East Asia (NEJ002, WJTOG3405, IPASS, OPTIMAL, LUX-Lung 6 and ENSURE; 100% East Asian population), one was global (LUX-Lung 3) with a 72% East Asian population, and one was European (EURTAC) with a 99% Caucasian population. The earliest studies commenced in March 2006 (NEJ002, WJTOG3405 and IPASS) and were completed by June 2009, while OPTIMAL and the LUX-Lung 3 and 6 studies were conducted between 2008 and 2011. ENSURE was conducted between 2011 and 2013.

All studies were open-label randomised controlled trials. However, randomisation methodology was not reported consistently across trials (notably lacking in NEJ002 and IPASS) and stratification criteria varied widely. Most trials had a 1:1 treatment allocation ratio. Only the LUX-Lung 3 and 6 studies had a treatment allocation ratio of 2:1. The sample size of the studies varied from 154 patients in OPTIMAL to 364 patients in LUX-Lung 6. The number of protocol violations in terms of patients' eligibility was low; four of the studies reported no violations, while EURTAC reported two, LUX-Lung 3 reported one and OPTIMAL reported four protocol violations.

Some key differences in trial conception and design were noted, including differences in EGFR mutation status between trials, ranging from not mandating EGFR mutation-positive status at baseline to requirement for specific EGFR mutations. With the exception of IPASS, all trials focused on patients whose mutation status was confirmed by various detection methods (table 2). In IPASS, the overall study population was clinically enriched for patients with an EGFR mutation-positive status (table 3); however, only a subgroup of patients had known EGFR mutation status. Of all the included studies, WJTOG3405 and IPASS differed most from the others with regard to heterogeneous patient population (table 3). Again with the exception of IPASS, all studies aimed to show superiority of the EGFR targeting TKI over chemotherapy. In contrast, with its overall study population clinically enriched for EGFR mutations, IPASS was designed to show noninferiority between treatments for the overall population.

''Measurable disease'', an important baseline criterion for the evaluation of response to treatment (according to the gold-standard RECIST (Response Evaluation Criteria In Solid Tumours) criteria), was not used consistently across studies. EURTAC included patients with ''measureable or evaluable disease'' and WJTOG3405 included patients with ''measurable and nonmeasurable disease''. Both NEJ002 and WJTOG3405 also had limitations on patients' age, leading to the exclusion of a relevant cohort of elderly patients. There was variation in whether investigator or independent assessments were conducted. Three studies relied on investigator review only. One trial (IPASS) did not describe the method of assessment. The key features of each trial, according to CONSORT criteria, are summarised in table 3.

Quantitative analyses of the trials

An overview of the outcomes presented across trials is provided in table 4, and in the context of EGFR mutation status in table 5. All studies significantly show the efficacy of EGFR-targeting TKIs in the EGFR mutant population, although there are differences in methodology and numerical outcome. For example, overall, PFS was shortest in WJTOG3405 (8.4 months) and longest in LUX-Lung 6 (13.7 months by investigator assessment). However, these trials differed in their use of investigator *versus* (blinded) independent assessment, as well as other methodologies, as described previously. The overall response rate was highest in OPTIMAL (83%) and lowest in EURTAC (58%) and LUX-Lung 3 (56%, independent review). Again, however, differences in assessment methodology were noted.

Findings for overall survival are also presented where available (table 4); however, the potential impact of crossover and the lack of assessment and reporting of outcomes with treatment post-progression further limits the comparability of these data.

DOI: 10.1183/09059180.00008413

TABLE 1 Overview of randomised phase III trials investigating first-line treatment of patients with epidermal growth factor receptor mutation-positive nonsmall cell lung cancer

CONSORT checklist entry [24]	Study [ref.]							
	NEJ002 [12, 23]	WJTOG3405 [13, 25]	IPASS# [8, 14]	EURTAC [9]	LUX-Lung 3 [15]	OPTIMAL [10, 11]	LUX-Lung 6 [16, 26]	ENSURE [17]
Publication status	Fully published as primary manuscript; overall survival update published subsequently	Fully published as primary manuscript; updated overall survival published at ASCO 2012	Fully published as primary manuscript; overall survival update published subsequently	Fully published as primary manuscript	Fully published as primary manuscript	Fully published as primary manuscript; overall survival update published at ASCO 2012	Primary outcomes and quality of life data published at ASCO 2013	Primary outcome published at WCLC 2013
Generic name	Gefitinib	Gefitinib	Gefitinib	Erlotinib	Afatinib	Erlotinib	Afatinib	Erlotinib
Registration number¶	UMIN-CTR number C000000376	UMIN-CTR number 000000539	NCT00322452	NCT00446225	NCT00949650	NCT00874419	NCT01121393	NCT01342965
Inclusion of protocol in the primary manuscript	Not attached to primary paper	Not attached to primary paper	Abbreviated CSR freely available online	Not attached to primary paper	Redacted trial protocol part of the primary paper	Part of the primary paper	Currently only conference presentations available	Currently only conference presentation available
Trial funding source	Japan Society for Promotion of Science; Japanese Foundation for the Multidisciplinary Treatment of Cancer and Tokyo Cooperative Oncology Group (IIT)	No sole study sponsor for this trial (IIT)	AstraZeneca	Roche (IIT)	Boehringer Ingelheim	Roche (IIT)	Boehringer Ingelheim	Roche

CONSORT: Consolidated Standards of Reporting Trials; WJTOG: West Japan Thoracic Oncology Group; NEJ002: North East Japan 002; IPASS: Iressa Pan-Asia Study; EURTAC: European Randomised Trial of Tarceva *versus* Chemotherapy; ASCO: American Society of Clinical Oncology; CSR: clinical study report; WCLC: World Conference on Lung Cancer; IIT: investigator-initiated trial #: all patients; ¶: www.ClinicalTrials.gov unless otherwise stated.

TABLE 2 Qualitative analyses of trials according to EGFR mutation status and laboratory methods used to confirm EGFR mutation status

	Study [ref.]							
	NEJ002 [12, 23]	WJTOG3405 [13, 25]	IPASS# [8, 14]	EURTAC [9]	LUX-Lung 3 [15]	OPTIMAL [10]	LUX-Lung 6 [16]	ENSURE [17]
Included patient group	EGFR mutant and absence of T790M	Del19 and L858R	Nonsmoker or former light smoker (no EGFR status required)	Del19 or L858R	EGFR mutation-positive	Del19 or L858R	EGFR mutation-positive	Del19 or L858R
EGFR mutations that could be included	Del19 L858R Other (not described)	Del19 L858R	Del19 (19 different deletions) L858R T790M L861Q G719X S768I Three insertions in exon 20	Del19 L858R	Del19 (19 different deletions) L858R T790M L861Q G719X S768I Three insertions in exon 20	Del19 L858R	Del19 (19 different deletions) L858R T790M L861Q G719X S768I Three insertions in exon 20	Del19 or L858R
Testing methodology used	Peptide nucleic acid-locked nucleic acid PCR clamp method	For those tested centrally, detection by fragment analysis for Del19 and by cycleave for L858R; confirmed by direct sequencing	Detection by Therascreen EGFR29 in central laboratory	Sanger sequencing confirmed by PNAClamp for Del19 and by TaqMan assay for L858R	Standardised allele-specific quantitative real-time PCR kit (Therascreen EGFR 29) in central laboratory	PCR-based direct sequencing in central laboratory	TheraScreen EGFR RGQ PCR kit in central laboratory	Not reported
Specificity	100% [28]	Direct sequencing: 100% [29]	100% [29]	100% [28]	100% [29]	100% [29]	100% [29]	
Sensitivity	89% [28]	Direct sequencing: 40–89% [29]	40–67% [29]	89% for PNAClamp [28] 99% for TaqMan	40–67% [29]	40–89% [29]	40–67% [29]	

EGFR: epidermal growth factor receptor; WJTOG: West Japan Thoracic Oncology Group; NEJ002: North East Japan 002; IPASS: Iressa Pan-Asia Study; EURTAC: European Randomised Trial of Tarceva *versus* Chemotherapy #: all patients. Therascreen EGFR 29 and TheraScreen EGFR RGQ PCR kit are manufactured by Qiagen (Manchester, UK); TaqMan is manufactured by Life Technologies Europe BV (Bleiswijk, the Netherlands); PNAClamp is manufactured by Panagene Inc. (Daejeon, Korea).

TABLE 3 Qualitative analyses of trials according to the CONSORT criteria

CONSORT checklist	Study [ref.]							
	NEJ002 [12, 23]	WJTOG3405 [13, 25]	IPASS# [8, 14]	EURTAC [9]	LUX-Lung 3 [15]	OPTIMAL [10, 11]	LUX-Lung 6 [16]	ENSURE [17]
2b: Specific objectives or hypotheses	Superiority of gefitinib over carboplatin/paclitaxel	Superiority of gefitinib over cisplatin/docetaxel	Non-inferiority of gefitinib over carboplatin/paclitaxel	Superiority of erlotinib over standard chemotherapy	Superiority of afatinib over cisplatin-pemetrexed	Superiority of erlotinib over standard chemotherapy	Superiority of afatinib over gemcitabine/ cisplatin	Superiority of erlotinib over gemcitabine/ cisplatin
3a: Allocation ratio	Two-arm parallel groups 1:1 randomised	Two-arm parallel groups 1:1 randomised	Two-arm parallel groups 1:1 randomised	Two-arm parallel groups 1:1 randomised	Two-arm parallel groups 2:1 randomised	Two-arm parallel groups 1:1 randomised	Two-arm parallel groups 2:1 randomised	Two-arm parallel groups 1:1 randomised
3b: Important changes after trial commencement	Not reported	Change of inclusion criteria; site for EGFR testing; sample size	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
4a: Eligibility criteria	Stage IIIb/IV	Initially recurrence after surgery (n=71); amended to stage IIIb/IV (n=101)	Stage IIIb/IV adenocarcinoma plus BAC	Stage IIIb/IV	Stage IIIb/IV adenocarcinoma	Stage IIIb/IV	Stage IIIb/IV adeno-carcinoma	Stage IIIb/IV
	Measurable disease (RECIST)	Measurable or non-measurable disease (RECIST)	Measurable disease (RECIST)	Measurable or evaluable disease (Not specified)	Measurable disease (RECIST v1.1)	Measurable disease (RECIST v1.0)	Measurable disease (RECIST v1.1)	Measurable disease (not specified)
	EGFR mutant and absence of T790M	Del 19 and L8585R	Nonsmokers or former light smokers (no EGFR status required)	Del19 or L858R	EGFR mutation-positive	Del19 or L858R	EGFR mutation-positive	Del19 or L858R
	ECOG 0-1	WHO 0–1	WHO 0–2	ECOG 0–2	ECOG 0–1	ECOG 0–2	ECOG 0–1	ECOG 0-2
	Age ⩽75 years	Age up to 75 years						
4b: Settings and locations where the data were collected	43 sites in Japan	36 sites in Japan	87 sites in East Asia	42 sites in Spain, Italy and France	133 sites in 25 countries globally	22 sites in China	36 sites in Asia	28 sites in Asia
5: Interventions	250 mg gefitinib or paclitaxel 200 mg·m^{-2} + carboplatin AUC6 three times per week	250 mg gefitinib or cisplatin 80 mg·m^{-2} + docetaxel 60 mg·m^{-2} three times per week	250 mg gefitinib or paclitaxel 200 mg·m^{-2} + carboplatin AUC5-6 three times per week	150 mg erlotinib three times a week or either: cisplatin 75 mg·m^{-2} + docetaxel 75 mg·m^{-2}; or cisplatin + gemcitabine 1250 mg·m^{-2}; or, for patients ineligible for cisplatin, carboplatin AUC6 + docetaxel or + gemcitabine (1000 mg·m^{-2} and carboplatin AUC5)	40 mg afatinib or cisplatin 75 mg·m^{-2} + 500 mg·m^{-2} pemetrexed three times a week	150 mg erlotinib or carboplatin AUC5 + gemcitabine 1000 mg·m^{-2} three times a week	40 mg afatinib or cisplatin 75 mg·m^{-2} + 1000 mg·m^{-2} gemcitabine three times per week	150 mg erlotinib or cisplatin 75 mg·m^{-2} + gemcitabine 1250 mg·m^{-2} three times per week
6a: Pre-specified primary and secondary end-points including assessment	At least three cycles Primary: PFS by two monthly CT	3–6 cycles Primary: PFS by 2 monthly CT/MRI	Up to 6 cycles Primary: PFS every 6 weeks	Up to 4 cycles Primary: PFS every 6 weeks by CT	Up to 6 cycles Primary: PFS every 6 weeks by CT/MRI	Up to 4 cycles Primary: PFS every 6 weeks by CT/MRI/bone scan	Up to 6 cycles Primary: PFS every 6 weeks by CT/MRI	Up to 4 cycles Primary: PFS
	Investigator review RECIST 1.0	Investigator review RECIST	Method and assessment not described RECIST	Investigator review Confirmation by central review board. RECIST 1.0	Independent central review RECIST 1.1	Investigator review with input from radiologist RECIST 1.0	Independent central review RECIST 1.1	Investigator review with Independent Review Committee assessment for sensitivity analysis [17]
	Secondary: OS, ORR, QoL, safety	Secondary OS, ORR, DCR, mutation type specific survival, safety	Secondary: OS, ORR, QoL, correlation of efficacy to baseline status of EGFR, safety	Secondary: OS, ORR, EGFR mutation analysis in serum	Secondary: OS, ORR, QoL, PK, safety	Secondary: OS, ORR, TTP, doR, QoL, safety	Secondary: OS, ORR, QoL, safety	Secondary: OS, ORR, safety
7a: Sample size	Sample size: 230 based on a PFS of 9.7 *versus* 6.7 months to achieve a power of 80% and a two-sided significance level of 5	Sample size: 146 to achieve a HR of 0.5 with 90% power to show superiority α 0.05 two-sided; HR amended to 0.48	With 944 progression events, the study would have 80% power to demonstrate non-inferiority, with a two-sided 5% probability of an erroneous demonstration of non-inferiority	Sample size 135 events (174 patients) to show PFS of 10 months *versus* 6 months with 80% power to show superiority α 0.05 two-sided	Sample size: 330 to show a HR of 0.64, equating to an increase in median PFS from an expected 7 months for chemotherapy to 11 months for afatinib to provide 90% power with a two-sided 5% significance level	Sample size: 152 patients based on PFS 11 months *versus* 6 months for a HR of 0.54 with a power of 80% and an α of 0.025	Sample size: at least 217 events reported by independent review needed to detect a HR of 0.64 (or median increase in PFS from 7 to 11 months) at two-sided 5% significance level with 90% power	Sample size: 217 patients randomised
7b: Interim analyses	One interim analysis after enrolment of 200 patients resulting in premature closure	Initially planned but not done; prematurely stopped after a DMC recommendation	One interim analysis was planned The purpose of this analysis was to detect inferiority of gefitinib compared with carboplatin/paclitaxel in terms of PFS, DMC recommended to continue with the trial	One planned interim analysis at 88 events The DMC recommended halting enrolment	No interim analysis	No interim analysis	No interim analysis	One planned interim analysis was conducted after 73% of PFS events (cut-off July 20, 2012) An additional exploratory updated analysis (cut-off November 19, 2012), included all planned PFS events

DOI: 10.1183/09059180.00008413

TABLE 3 Continued

CONSORT checklist	Study [ref.]							
	NEJ002 [12, 23]	WJTOG3405 [13, 25]	IPASS# [8, 14]	EURTAC [9]	LUX-Lung 3 [15]	OPTIMAL [10, 11]	LUX-Lung 6 [16]	ENSURE [17]
8-10: Methods of randomisation	Randomisation not described Strata: sex, stage, site	Central Fax randomisation Strata: site, adjuvant chemotherapy, interval between surgery and recurrence, stage, sex	Randomisation not described Dynamic balancing: WHO performance status, smoking status, sex, site	Computer-generated central randomisation by CRO Strata: ECOG performance status and mutation type	Computer-generated central randomisation by IVRS Strata: ethnicity, mutation type	Central randomisation *via* telephone or email Strata: histology, smoking status, mutation type	Computer-generated central randomisation by IVRS Strata: mutation type	Not reported Strata: ECOG performance status, mutation type, sex, country
12a: Statistical methods for primary and secondary outcomes	Kaplan–Meier using log-rank test, HR using Cox proportional hazard model ORR and safety were compared between the two groups with Fisher's exact test and the Wilcoxon test, respectively Each analysis was performed with the use of a two-sided, 5% significance level and a 95% CI	Kaplan–Meier using log-rank test, HR using Cox proportional hazard model The Chi-squared test was used to compare proportions Differences were considered significant at a two-sided p-value of ⩽0.05	Cox proportional hazards model in the ITT, ORR and QoL were assessed with the use of a logistic regression model with the same covariates as those considered for PFS to calculate ORs and 95% CIs Adverse events were compared with the use of Fisher's exact test; adjustment for multiple comparisons was performed with the use of the method of Westfall and Young	Kaplan–Meier curves using the log-rank test HR (95% CI) by Cox proportional hazards analysis Prespecified adjustment factors included ECOG performance status and type of mutation (exon 19 deletion *versus* L858R) Response rates were compared between the two groups using the Chi-squared test	Stratified log-rank test, using the same stratification factors used in randomisation Cox proportional hazard models were used to compare PFS between arms, and Kaplan–Meier estimates were calculated PFS analysis in patients with common EGFR mutations (L858R and exon 19 deletions) was prespecified Logistic regression models were used to compare arms	Survival was estimated with Kaplan-Meier methodology A two-sided log-rank test was used to compare survival between the two treatment groups Exploratory and pre-planned subgroup analyses of PFS were performed with the Cox proportional hazards model and included the stratification factors from randomisation	Stratified log-rank and Cox proportional hazard for PFS comparisons (ITT for all randomised patients) Prespecified subgroup analyses included sex, age, mutation type, performance status and smoking status	Not reported
12b: Methods for additional analyses including subanalyses	One interim analysis was planned to analyse the primary end-point (significance level, p=0.003) The Lan–DeMets method was used to adjust for multiple comparisons The O'Brien–Fleming type α-spending function was also used	HRs in the overall population and in patient subsets were calculated using the Cox proportional hazards model The Chi-squared test was used to compare proportions	Tests to determine interactions of treatment with covariates were used to identify predictive factors by assessing whether there was a significant difference in the treatment effect for PFS (HR for progression or death) between subgroups	A Lan–DeMets α-spending function with a Pocock stopping boundary was used to maintain the significance level at 5% with a 0.037 significance level at interim and 0.025 for the final analysis based on 135 events	NA	NA	NA	NA
13a: Participant flow for primary outcome	Screened: not reported Enrolled: 230 Excluded: 2	Screened: 337 Enrolled 118 + 71 (detected at commercial clinical laboratory, not central laboratory) Excluded: 12	Screened 1329 Enrolled: 1217 Excluded: 0	Screened: 1227 Enrolled: 174 Excluded: 42 for change in target lesion	Screened: 1269 Enrolled: 345 Excluded: 0	Screened; 549 Enrolled: 165 Excluded: 11	Screened: 910 Enrolled 364 Excluded 0	NA
13b: Participant flow for losses and exclusions	No protocol violations, six patients were excluded from the PFS analysis	No protocol violations, five randomised patients were excluded from efficacy analyses	No protocol violations	Two protocol violations: two patients less than stage IIIb in erlotinib arm Not excluded from analyses 1 patient received treatment before randomisation (protocol violation)	One protocol violation: ECOG 2	Four protocol violations (patients allocated to chemotherapy received an EGFR TKI) were excluded from analyses	No protocol violations	NA
14a: Recruitment dates	March 2006 to May 2009 Median follow-up >17 months	March 2006 to June 2009 Median (range) follow-up was 81 (74–1253) days	March 2006 to October 2007 Median follow-up for OS 17 months	February 2007 to January 2011 Median follow-up 14.4 months for chemotherapy and 18.9 months for erlotinib	August 2009 to February 2011 Median follow-up 16.4 months	August 2008 to July 2009 Median follow-up 15.6 months	April 2010 to November 2011 Median follow-up not reported	April 2010 to November 2011 Median follow-up 10.3 months for chemotherapy and 11.7 months for erlotinib
16: Numbers analysed	224 for PFS 228 in ITT	172	261 EGFR mutation positive	173 (131 for change in target lesion)	345	154	364	217

CONSORT: Consolidated Standards of Reporting Trials; NEJ002: North East Japan 002; WJTOG: West Japan Thoracic Oncology Group; IPASS: Iressa Pan-Asia Study; EURTAC: European Randomised Trial of Tarceva *versus* Chemotherapy. RECIST: Response Evaluation Criteria In Solid Tumours; EGFR: epidermal growth factor receptor; ECOG: Eastern Cooperative Oncology Group; AUC*n*: target area under the free carboplatin plasma concentration *versus* time curve of $n\times$ (glomerular filtration rate + 25) mg·m^{-2}; PFS: progression-free survival; CT: computed tomography; OS: overall survival; ORR: overall response rate; QoL: quality of life; HR: hazard ratio; ITT: intention-to-treat; WHO: World Health Organization; MRI: magnetic resonance imaging; DCR: disease control rate; DMC: data monitoring committee; BAC: bronchoalveolar carcinoma; PK: pharmacokinetics; CRO: contract research organisation; IVRS: contract research organisation; NA: not available; TTP: time to progression; doR: duration of response; TKI: tyrosine kinase inhibitor. #: all patients.

TABLE 4 Quantitative analyses of included clinical trials: tyrosine kinase inhibitors (TKI) *versus* chemotherapy

Study [ref.]	Patients treated with TKI n	PFS	ORR %	Overall survival	Incidence of grade 3-5 adverse events[#] >1% of patients	Rate of discontinuation due to adverse events %[¶]
NEJ002 [12, 23]	114	10.8 months *versus* 5.4 months; HR 0.32 (95% CI 0.24–0.44), p<0.001	74 *versus* 31; p<0.001	27.7 months *versus* 26.6 months; HR 0.89 (95% CI 0.63–1.24), p=0.483	AST/ALT elevation 25%, rash 5.3%, appetite loss 5.3%, fatigue 2.6%, pneumonitis 2.6%	Not reported
WJTOG3405 [13, 25]	51 for PFS (stage IIIb/IV subgroup) 86 for overall survival	8.4 months *versus* 5.3 months; HR 0.33 (95% CI 0.21–0.54), p<0.0001 (stage IIIb/IV subgroup)	62 *versus* 32[++]; p<0.0001	36 months *versus* 39 months; HR 1.19 (95% CI 0.771.83), p=0.443	Whole population (including those with recurrent disease): ALT elevations 27.6%, AST elevations 16.1%, fatigue 2.3%, rash 2.3%, diarrhoea 1.1%, paronychia 1.1%, nausea 1.1%, sensory disturbance 1.0%	16
IPASS[+] [8, 14]	132	9.5 months *versus* 6.3 months; HR 0.48 (95% CI 0.36–0.64), p<0.001	71 *versus* 47; p<0.001	21.6 months *versus* 21.9 months; HR 1.00 (95% CI 0.76–1.33), p=0.990)	Whole population: diarrhoea 3.8%, neutropenia 3.7%, rash or acne 3.1%, anaemia 2.2%, anorexia 1.5%, leukopenia 1.5%	Whole population, 7
EURTAC[§] [9, 19]	86	9.7 months *versus* 5.2 months; HR 0.37 (95% CI 0.25–0.54), p<0.0001	58 *versus* 15; p-value not reported	19.3 months *versus* 19.5 months; HR 1.04 (95% CI 0.65–1.68), p=0.87	Rash 13%, fatigue 6%, diarrhoea 5%, AST/ALT 2%, anaemia 1%, neuropathy 1%, arthralgia 1%, pneumonitis 1%	13
EURTAC[/] [9, 19]	86	10.4 months *versus* 5.4 months; HR 0.47 (95% CI 0.28–0.78), p=0.0030				
LUX-Lung 3[##]	230	11.1 months *versus* 6.9 months; HR 0.58 (95% CI 0.43–0.78); p=0.001	56 *versus* 23; p=0.001	28.1 months *versus* 28.2 months; HR 0.91 (95% CI 0.66–1.25), p=0.55 (yet immature)	Rash/acne 16.2%, diarrhoea 14.4%, paronychia 11.4%, stomatitis/mucositis 8.7%, decreased appetite 3.1%, vomiting 3.1%, fatigue 1.3%	8
LUX-Lung 3[§] [15]	230	11.1 months *versus* 6.7 months; HR 0.49 (95% CI 0.37–0.65); p=0.001	69 *versus* 44; p=0.001			
OPTIMAL [10, 11]	82	13.1 months *versus* 4.6 months; HR 0.16 (95% CI 0.10–0.26), p<0·0001	83 *versus* 36; p<0.0001	22.7 months *versus* 28.9 months; HR 1.04 (95% CI 0.69–1.58), p=0.69 (yet immature)	ALT 4%, rash 2%	1
LUX-Lung 6[¶¶] [16]	242	11.0 months *versus* 5.6 months; HR 0.28 (95% CI 0.20–0.39), p<0.0001	67 *versus* 23; p<0.0001	Not reported; immature	Rash/acne 14.6%, diarrhoea 5.4%, stomatitis/mucositis 5.4%, ALT increase 1.7%, decreased appetite 1.3%	6
LUX-Lung 6[§] [16]	242	13.7 months *versus* 5.6 months; HR 0.26 (95% CI 0.19–0.36), p<0.0001	74 *versus* 31; p-value not reported			
ENSURE[¶¶] [17]	110	11.0 months *versus* 5.6 months; HR 0.42 (95% CI 0.27–0.66), p<0.0001	63 *versus* 34; p=0.0001	Not reported; immature	Rash 6.4%, diarrhoea 1.8%[§§]	3
ENSURE[§] [17]	110	11.0 months *versus* 5.5 months; HR 0.34 (95% CI 0.22–0.51), p<0.0001)				

NEJ002: North East Japan 002; WJTOG: West Japan Thoracic Oncology Group; IPASS: Iressa Pan-Asia Study; EURTAC: European Randomised Trial of Tarceva *versus* Chemotherapy; PFS: progression-free survival; ORR: overall response rate; HR: hazard ratio; ALT: alanine transaminase; AST: aspartate aminotransferase. [#]: TKI treatment only; [¶]: TKI treatment only, independent of relation to study drug; [+]: mutation-positive subgroup; [§]: investigator assessed; [/]: investigator assessment based on 45 patients, independent review based on 31 patients treated with erlotinib; [##]: independent review (primary end-point); [¶¶]: independent assessed; [++]: measurable disease but stage-independent (TKI and chemotherapy combined n=117); [§§]: only adverse events of special interest were reported.

In terms of safety, increases in alanine transaminase/aspartate aminotransferase were most common in the gefitinib trials (NEJ002 and WJTOG3405), while fatigue had the highest incidence in the EURTAC study, and rash and diarrhoea were most commonly reported in the afatinib studies (LUX-Lung 3 and LUX-Lung 6). However, rates of adverse event-related discontinuation were presented differently between studies (some studies only reported treatment-related adverse events/EGFR mutation-positive patients while others presented overall data). No information was available regarding the quality of adverse event reporting in terms of source data monitoring.

Comparison of the evaluation of health-related quality of life/patient-reported outcomes

Three different questionnaires were used in the five studies addressing health-related quality of life (HRQoL): Care Notebook; European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ); and the Functional Assessment of Cancer Therapy – Lung (FACT-L) questionnaire (including lung cancer-specific modules of the latter two). Key details of the quality of life (QoL) analyses are summarised by study in table 6. In NEJ002 [30], QoL was assessed for 20 weeks after initiation of first-line therapy using the Care Notebook [34–36], a self-administered, cancer-specific questionnaire that comprises 24 domains structured in multidimensional scales, assessed using one word or a short phrase graded on an 11-point linear analogue scale (scored 0–10). Patients complete the questionnaire before therapy and then weekly during first-line treatment. Deterioration is noted when

DOI: 10.1183/09059180.00008413

TABLE 5 Quantitative analyses regarding epidermal growth factor receptor mutation status of included clinical trials

Trial	Common mutations			Del19			Exon 21			Uncommon mutation		
	Patients treated with TKI n	Median PFS months	HR (95% CI), p-value	Patients treated with TKI n	PFS months	HR (95% CI), p-value	Patients treated with TKI n	PFS months	HR (95% CI), p-value	Patients treated with TKI n	PFS months	HR (95% CI), p-value
NEJ002 [12, 23]	107	NR	NR	58	11.5	NR	49	10.8	NR	7	NR	NR
WJTOG3405 [13, 25]	51 (stage IIIb/IV subgroup)	8.4	0.33 (0.21–0.54), p<0.0001	NR	NR	NR	NR	NR	NR	NA	NA	NA
IPASS [8, 14]	130	NR	NR	66	11.0	0.38 (0.26–0.56), p=NR	64	9.2	0.55 (0.35–0.87), p=NR	8	NR	NR
EURTAC [9]	86	9.7	0.37 (0.25–0.54), p<0.0001	57	11.0	0.3 (0.18–0.50), p<0.0001	29	8.4	0.55 (0.29–1.02), p=0.0539	NA	NA	NA
LUX-Lung 3 [15]	204	13.6	0.47 (0.34–0.65), p=0.001	113	13.7	0.28 (0.18–0.44), p=0.01	91	10.8	0.73 (0.46–1.17), p=0.01	26	2.8	
OPTIMAL [10]	82	13.1	0.16 (0.10–0.26), p<0.0001	43	15.3	0.13 (0.07–0.25), p=NR	39	12.5	0.26 (0.14–0.49), p=NR	NA	NA	NA
LUX-Lung 6 [16]	216	13.7	0.25 (0.18–0.35)	124	13.7	0.20 (0.13–0.33), p=NR	92	9.6	0.32 (0.19–0.52), p=NR	26		0.55 (0.22–1.43)
ENSURE [17]	110	11.0	0.34 (0.22–0.51), p<0.0001	NR	11.1	0.20 (0.11–0.37), p=NR	NR	8.3	0.57 (0.31–1.05), p=NR	NA	NA	NA

All data are from investigator assessment, except LUX-Lung 3 which is from independent review. NEJ002: North East Japan 002; WJTOG: West Japan Thoracic Oncology Group; IPASS: Iressa Pan-Asia Study; EURTAC: European Randomised Trial of Tarceva *versus* Chemotherapy; TKI: tyrosine kinase inhibitor; PFS: progression-free survival; HR: hazard ratio; CI: confidence interval; NR; not reported.

TABLE 6 Quantitative analyses of included clinical trials: health-related quality of life data[#]

	NEJ002 [12, 23, 30]	IPASS [8, 14, 31]	OPTIMAL [10, 32]	LUX-Lung 3 [15, 33]	LUX-Lung 6 [16, 26]
Questionnaire **Assessment until disease progression**	Care notebook Baseline, weekly	FACT-L (incl. LCS/TOI) Baseline, weeks 1 and 3 Every 3 weeks until week 18, then every 6 weeks	FACT-L (incl. LCS/TOI) Baseline, every 6 weeks	EORTC-QLQ C30 and LC13 Baseline, every 3 weeks	EORTC-QLQ C30 and LC13 Baseline, every 3 weeks
Compliance with completing questionnaires[+]	Gefitinib: 63% chemotherapy: 69% (at least two time-points)	Gefitinib: 95% chemotherapy: 90% (time-point NR)	Erlotinib: 96%/91% cycle 2/cycle 6 chemotherapy: 100%/50% cycle 2/cycle 6	Afatinib: 97%/98% cycle 2/cycle 6 chemotherapy: 97%/83% cycle 2/cycle 6	Afatinib: 96%/85% cycle 2/cycle 6 chemotherapy: 98%/90% cycle 2/cycle 6
Significant and clinically relevant[#] symptom improvement[¶,§]	Loss of appetite (p=0.014) Constipation (p<0.0001) Pain and shortness of breath (p<0.0001)	Maintaining at least 21 days FACT-L (70% *versus* 45%), TOI (70% *versus* 38%), LCS (76% *versus* 54%)	FACT-L, TOI, LCS	Dyspnoea (64% *versus* 50%) Pain (59% *versus* 48%; only significant for individual pain items)	Cough (76% *versus* 55%) Dyspnoea (71% *versus* 48%) Pain (64% *versus* 47%) Global health status (63% *versus* 33%) Physical (54% *versus* 29%) Role (50% *versus* 35%) Social (55% *versus* 35%)
Significant and clinically relevant differences in time to worsening/ deterioration	Pain and shortness of breath (0.2 *versus* 2.1 months) Daily functioning (0.4 *versus* 3.0 months)	FACT-L (15.6 *versus* 3.0 months) TOI (16.6 *versus* 2.9 months) LCS (11.3 *versus* 2.9 months)	NA	Cough (NE *versus* 8.0 months) Dyspnoea (10.3 *versus* 2.9 months)	Cough (NE *versus* 10.3 months) Dyspnoea (7.7 *versus* 1.7 months) Pain (6.4 *versus* 3.4 months)
Significant and clinically relevant changes in longitudinal analyses	NA	NA	NA	Cough Dyspnoea	Cough Dyspnoea Pain

NEJ002: North East Japan 002; IPASS: First Line IRESSA *versus* Carboplatin/Paclitaxel in Asia; FACT-L: Functional Assessment of Cancer Therapy – Lung; LCS: lung cancer subscale; TOI: trial outcome index; EORTC-QLQ C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire C30; LC13: lung cancer-specific module; NR: not reported; NA: not available; NE: not evaluable. [#]: different definitions of ''clinically meaningful'' were used in the different evaluations; [¶]: to date, no quality of life data have been published from the ENSURE trial; [+]: baseline, cycle 6; [§]: data are presented as % patients, tyrosine kinase inhibitor *versus* chemotherapy.

worsening from baseline by one of 11 points (9.1%) occurs at any time-point [37, 38]. To the best of our knowledge, the Care Notebook is not used outside Japan.

In the LUX-Lung 3 and 6 studies, patient-reported outcomes (PROs) were comprehensively assessed at randomisation and then every 21 days until disease progression [26, 33, 39] using the self-administered, cancer-specific EORTC QLQ-C30 [40, 41], comprising 30 questions of both multi- and single-item measures, and the lung cancer-specific module QLQ-LC13 [42, 43], comprising 13 questions and designed for use in patients with lung cancer undergoing chemotherapy or radiotherapy. Each item utilises a four-point linear analogue scale with a seven-point scale for overall health and QoL. A linear transformation is then applied to standardise the raw score to a range from 0 to 100 (high scores represent a high/healthy level of functioning or high/severe level of symptomatology) [41, 44]. A 10-point change in an item or domain was accepted as a clinically meaningful change [38], with a $\geqslant$10-point decrease from baseline at any time during the study used to define symptom improvement. Time to deterioration in PROs was defined as months from randomisation to the first instance of symptom worsening (10 points from baseline) [38, 45], and changes in PROs scores over time were assessed using mixed-effects growth curve models [46].

Both IPASS [31] and OPTIMAL [32] assessed HRQoL using the total score of the FACT-L questionnaire and the Trial Outcome Index (TOI; sum of the physical well-being, functional well-being and lung cancer subscale (LCS) scores of the FACT-L), and lung cancer symptom improvement was assessed using the LCS domain of the FACT-L. Questionnaires were completed at baseline, week 1 and week 3, then every 3 weeks until week 18, and then every 6 weeks until tumour progression, and at treatment discontinuation. Each item uses a five-point linear analogue scale, with clinically relevant improvement/worsening in HRQoL and symptoms predefined as an increase or decrease from baseline of $\geqslant$6 points for FACT-L and TOI, and $\geqslant$2 points for LCS, maintained for $\geqslant$21 days [37].

All studies showed clinically relevant symptom improvement; however, for a direct comparison, it is critical that different definitions of what was considered to be clinically meaningful are applied. EORTC has the highest threshold for what was considered to be clinically meaningful (10%); Care Notebook has the most granular questionnaire with the most requirement of data collection. This may be why the compliance rate for patients recording these data was so low in this study. The definition of ''clinically meaningful'' used in EORTC questionnaires was prospectively developed by Osoba *et al.* [38]. This is a subjective significance questionnaire where a change in score of 5–10 points is perceived by patients as having little difference. A difference in score of 10–20 points is perceived as moderately different, and a difference of >20 points is

DOI: 10.1183/09059180.00008413

perceived as very different. By combining systematic reviews, expert opinions and meta-analysis, COCKS *et al.* [44] demonstrated similar results. An improvement by 0–4 points was considered trivial, an improvement by 4–10 points was considered little difference, an improvement by 10–15 points was considered moderately different, and an improvement by >15 points was considered very different. The approach for FACT-L was a retrospective estimation. CELLA *et al.* [37] defined criterion-related validity as the relationship of test scores to meaningful anchors such as performance status rating, weight loss and presence of primary disease symptoms, and used this information to provide meaning to scores based on group-level differences from one trial. Clinically relevant changes were estimated as 2–3 points for the LCS and 5–7 points for the TOI [37]. Furthermore, longitudinal analysis, which shows that the effects are long lasting and not only snapshots, was only available for afatinib. Interestingly, the positive impact of afatinib over chemotherapy in the LUX-Lung studies could also be shown for the time-period where patients were on drug holiday from chemotherapy but still on afatinib.

Limitations of comparisons across trials

There is no question that TKIs targeting EGFR are superior to platinum-based chemotherapy in the first-line setting of NSCLC patients with EGFR mutation-positive tumours. However, as a result of the substantial differences in methodology between studies and reporting, caution should be exercised when comparing outcomes between trials.

The high incidence of patients with EGFR mutation-positive NSCLC in Asia is reflected by the focus on Asian populations in studies conducted in this setting. While a large proportion of patients in the LUX-Lung 3 study came from Asia, LUX-Lung 3 was the only study designed to be global, with subanalyses using race (Asian *versus* non-Asian) as a stratification showing no significant difference between the ethnicities, hence further increasing the global relevance of the data.

Some trial design features apply to all studies. All were randomised controlled trials and all had a low rate of protocol violations; however, when looking more closely at our quantitative analysis, two studies warrant specific discussion as they differ most from the others: WJTOG3405 and IPASS. In the WJTOG3405 study important changes were implemented during study conduct with the inclusion of a heterogeneous patient population (both post-operative recurrent as well as stage IIIb/IV patients). This makes the findings from this study somewhat difficult to put into perspective. IPASS was conducted to show noninferiority of TKIs *versus* chemotherapy in clinically enriched patients. The subgroup analysis of patients with EGFR mutation-positive tumours was pre-planned and superiority for the whole population could be concluded from the same analysis without statistical penalty. With these exploratory analyses, IPASS was certainly a milestone for the understanding of the activity of gefitinib in patients with EGFR mutation-positive tumours, but unlike the other studies discussed here, IPASS was not designed to show superiority. The similar design and robust methodology used in LUX-Lung 3 and LUX-Lung 6 has led to a high reproducibility of the efficacy results for afatinib. This has not been observed in the erlotinib study (ENSURE, OPTIMAL or EURTAC). Currently, we cannot provide reasons for this.

Limitations of source data

Some of the included studies provided only limited information required by CONSORT. In some cases, study protocols were not included as supplementary information to the primary publications (despite being recommended by CONSORT and requested by many journals). The exceptions to this were OPTIMAL and LUX-Lung 3, both of which included study protocols as part of the primary publication.

CONSORT also requires the disclosure of randomisation methods, as these are critical for judging of the quality of a trial [24]. However, to our knowledge, this was not carried out for NEJ002 or IPASS. The term ''random'' is frequently used to describe treatment allocation methods that do not fit with its precise definition. These include nonrandom methods to determine allocation, such as alternation and the use of hospital numbers or birth date, which have the potential to lead to biases in the study design and, ultimately, study outcomes. Furthermore, while various different methods of sequence generation, such as the nonrandom process of minimisation, are acceptable, this cannot be determined from descriptors centred on randomness alone. As such, more detailed descriptions of randomisation and sequence generation methods are needed [24].

In some publications, results of secondary end-points and data on stratification factors were not reported. In EURTAC, data from external review were, to our knowledge, not published in a peer-reviewed journal. In the primary OPTIMAL publication, the secondary end-points duration of response and time to progression were not reported. Furthermore, the influence of stratification on PFS was not shown. In the OPTIMAL trial, only patients who ''had received at least one dose of study drug'' rather than the intention to treat population were included in the efficacy analyses. Furthermore, in response to the request by the European

Medical Agency, no clinical study report has been made available for this trial [48]. The ongoing ENSURE study [17] was conducted in the same setting as the OPTIMAL study, and may provide robust data for the efficacy and safety of TKIs in EGFR mutation-positive NSCLC.

The independent review of PFS in the EURTAC study was conducted retrospectively. In this study, not all of the scans were available for independent review. Based on the number of available scans and relevant clinical information, 30 patients were considered to have had an event by independent review in the chemotherapy arm *versus* 31 patients in the erlotinib arm (data cut-off August 2, 2010). In the investigators' assessment, the number of patients who experienced an event in the chemotherapy and erlotinib treatment arms were 47 and 45, respectively [47].

It should also be noted that for studies where not only EGFR-mutated stage IIIB/IV patients were enrolled (*e.g.* IPASS and WJTOG3405), safety data were reported for the whole study population (and not the subpopulation of interest), potentially leading to artificially low occurrence rates. A recently published analysis showed that some methodological aspects of adverse event collection and analysis are poorly reported in trials. Given the importance of adverse events in evaluating new treatments, authors should be encouraged to adhere to the 2004 CONSORT guidelines regarding adverse event reporting [48].

Limitations of trial methodology

Eligibility criteria

There were substantial differences in eligibility criteria across the studies, some of which have the potential to introduce bias and compromise direct comparison of trial outcomes. Three trials included patients with an Eastern Coopertaive Oncology Group (ECOG) performance status of 0–2. However, the majority of patients included in these trials had a performance status of 0 or 1, meaning that few patients with a performance status of 2 were included. Of note, the afatinib trials excluded ECOG performance status 2 patients, and two studies had restrictions regarding the inclusion of elderly patients, excluding a relevant subgroup of patients in the stage IIIb/IV NSCLC setting, [15, 49]. Most studies were not restricted to patients with adenocarcinoma. As the histology of the tumour may influence prognosis, this difference may need to be taken into account before making any direct comparisons across studies. WJTOG3405 recruited patients with "non-measurable" disease and EURTAC included patients with "evaluable" disease, which compromises the assessment of the efficacy of these studies by standard RECIST criteria [50, 51].

Four studies were limited to patients with common mutations (Del19/L858R), meaning that these study populations were more homogeneous *versus* studies that included patients with both common and uncommon mutations [52]. For direct comparison of study results, this difference has a significant impact. The benefit of TKI treatment in patients with common mutations is well-established. This was also shown in the LUX-Lung 3 study, where the median PFS for patients with common mutations was 13.6 months compared with 11.1 months for all patients including those with uncommon mutations [15]. Due to the low number of patients with uncommon mutations, it still remains unclear as to what is the best treatment for these individuals. The testing methods for EGFR mutation detection were of different sensitivity; PCR methods have demonstrated lower invalid rates and higher sensitivity than Sanger in the detection of EGFR mutations [53, 54]. Highly sensitive testing methods, such as the peptide nucleic acid-locked nucleic acid PCR clamp methods used in NEJ002, have the potential to identify patients with low numbers of EGFR mutation-positive tumour cells. This could be different compared with methods where patients with a high percentage of EGFR-mutated cells may be selected and who might be expected to respond better to TKI.

Choice of comparator treatment

The comparator and number of cycles used in NEJ002, EURTAC, OPTIMAL and ENSURE varied as only three or four cycles of chemotherapy were allowed. In EURTAC, a defined variety of different chemotherapy regimens was allowed. This is an important consideration, given the impact that comparator treatment has on the comparability of efficacy data across trials. The relevance of the comparator arm is illustrated by the differences in HRs for PFS in the LUX-Lung 3 and LUX-Lung 6 studies. In these studies, the choice of comparators was driven by the differences in regulatory approvals for chemotherapies across the countries in which the studies were conducted. In LUX-Lung 6, the use of cisplatin/gemcitabine resulted in a lower PFS (HR 0.26) than with cisplatin/pemetrexed in LUX-Lung 3 (HR 0.58) (table 4). The experimental arms were essentially identical but chemotherapies differed substantially regarding PFS. However, when assessing LUX-Lung 6 and ENSURE, both comparing the respective investigational compound with cisplatin/gemcitabine, and reporting the same median PFS by independent review (11.0 months and 5.5–5.6 months for EGFR TKI and chemotherapy, respectively), a difference in HRs was observed (0.25 in LUX-Lung 6 compared with 0.42 in ENSURE for patients with common mutations). However, as neither trial is fully published yet we cannot speculate about the underlying reason.

DOI: 10.1183/09059180.00008413

Assessment and reporting of trial outcomes

The recognition of inherent differences in assessment and reporting of trial outcomes is critical for comparison of data across studies. Not all studies had prospective independent review by blinded oncologists/radiologists, which is regarded as the most conservative approach to assessing response to therapy, and is recommended in RECIST guidelines [50, 51]. Studies in this setting also lack assessment and reporting of survival outcomes with post-progression crossover treatment. As a result, the optimal sequence of EGFR TKI/afatinib and chemotherapy in patients with EGFR mutations has yet to be clarified.

Due to the difficulty in assessing overall survival benefits in clinical trials, HRQoL is an important method of measuring response to treatment in the first-line setting. From the consistent picture, it can be concluded that in general there is a benefit with both reversible and irreversible EGFR TKIs, especially when this is indicated by robust results from studies with validated HRQoL questionnaires and the inclusion of longitudinal analysis. The robustness of EORTC questionnaires ensures high generalisability of results. In the studies analysed here, HRQoL was improved with EGFR TKIs and afatinib compared with chemotherapy, and all trials showed clinically relevant differences in time to deterioration. The high return of completed questionnaires in IPASS, OPTIMAL, LUX-Lung 3 and LUX-Lung 6 strengthens the reliability of these data; however, in NEJ002, the QoL assessment gives limited information as patients had a low compliance in completing these data. Furthermore, differences in the approach to determining clinically meaningful improvements and a lack of longitudinal analyses in most trials means that, as with the other outcomes discussed in this review, caution must be exercised when comparing HRQoL results.

Conclusions

Taken together, these clinical trials provide substantial evidence that erlotinib, gefitinib and afatinib are the standard of care for patients with EGFR mutation-positive NSCLC, and should be considered as first-line treatment options. The results of QoL analyses, as a sum of side-effects and symptom improvement, support this view. However, cross-trial comparisons generally have strong scientific limitations. This is particularly obvious when comparing differences in trial design, comparator choice, inclusion criteria and reporting standards. Without highlighting these differences, the outcomes of these studies may be misinterpreted as comparable. For a reader not familiar with the intricacies of these studies, it is tempting to relate directly to the eye-catching single values of median PFS or response rate. This review shows that such comparisons are not valid. Furthermore, the optimal sequencing of EGFR TKIs, afatinib and chemotherapy in patients with EGFR mutations requires more investigation. The head-to-head comparisons of afatinib with gefitinib in the first-line setting (LUX-Lung 7; www.ClinicalTrials.gov identifier NCT01466660, fully recruited) and dacomitinib with gefitinib (ARCHER 1050; www.ClinicalTrials.gov identifier NCT01774721, ongoing) will shed more light on how these agents compare. For oncologists and patients, it is of high importance that clinical trial results are robust and generalisable. Therefore, it is highly desirable that future studies in NSCLC make use of the appropriate tools: independent tumour assessment; most appropriate randomisation methods; clearly defined patient populations; and well-recognised QoL questionnaires, to name a few.

Acknowledgements

Medical writing assistance, supported financially by Boehringer Ingelheim Pharma GmbH & Co. KG, was provided by C. Smart (Ogilvy Healthworld, London, UK) during the preparation of this manuscript. The authors were fully responsible for all content and editorial decisions, were involved at all stages of manuscript development, and have approved the final version.

References

1 Miller VA, Kris MG, Shah N, *et al.* Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non-small-cell lung cancer. *J Clin Oncol* 2004; 22: 1103–1139.
2 Giaccone G, Gallegos Ruiz M, Le Chevalier T, *et al.* Erlotinib for frontline treatment of advanced non-small cell lung cancer: a phase II study. *Clin Cancer Res* 2006; 12: 6049–6055.
3 Cohen MH, Williams GA, Sridhara R, *et al.* FDA drug approval summary: gefitinib (ZD1839) (Iressa) tablets. *Oncologist* 2003; 8: 303–306.
4 Fukuoka M, Yano S, Giaccone G, *et al.* Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial). *J Clin Oncol* 2003; 21: 2237–2246.
5 Kris MG, Natale RB, Herbst RS, *et al.* Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *JAMA* 2003; 290: 2149–2158.
6 Lynch TJ, Bell DW, Sordella R, *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004; 350: 2129–2139.
7 Paez JG, Janne PA, Lee JC, *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004; 304: 1497–1500.
8 Mok TS, Wu YL, Thongprasert S, *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009; 361: 947–957.

9 Rosell R, Carcereny E, Gervais R, *et al.* Erlotinib *versus* standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* 2012; 13: 239–246.
10 Zhou C, Wu YL, Chen G, *et al.* Erlotinib *versus* chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* 2011; 12: 735–742.
11 Zhou C, Wu Y, Liu X. Overall survival (OS) results from OPTIMAL (CTONG0802), a phase III trial of erlotinib (E) *versus* carboplatin plus gemcitabine (GC) as first-line treatment for Chinese patients with EGFR mutation-positive advanced non-small cell lung cancer (NSCLC). *J Clin Oncol* 2012; 30: Suppl., 485s.
12 Maemondo M, Inoue A, Kobayashi K, *et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* 2010; 362: 2380–2388.
13 Mitsudomi T, Morita S, Yatabe Y, *et al.* Gefitinib *versus* cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. *Lancet Oncol* 2010; 11: 121–128.
14 Fukuoka M, Wu YL, Thongprasert S, *et al.* Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib *versus* carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). *J Clin Oncol* 2011; 29: 2866–2874.
15 Sequist LV, Yang JC, Yamamoto N, *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013; 31: 3327–3334.
16 Wu YL, Zhou CC, Hu CP, *et al.* LUX-Lung 6: a randomized, open-label, phase III study of afatinib (A) *vs* gemcitabine/cisplatin (GC) as first-line treatment for Asian patients (pts) with EGFR mutation-positive (EGFR M+) advanced adenocarcinoma of the lung. *J Clin Oncol* 2013; 31: Suppl., 490s.
17 Wu YL, Liam CK, Zhou C, *et al.* First-line erlotinib *versus* cisplatin/gemcitabine (GP) in patients with advanced EGFR mutation-positive non-small-cell lung cancer (NSCLC): interim analyses from the phase 3, open-label, ENSURE study. *J Thoracic Oncol* 2013; 8: Suppl. 2, S603.
18 European Medicines Agency. Gefitinib European public assessment report. www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Summary_for_the_public/human/001016/WC500036359.pdf Date last accessed: January 20, 2014. Date last updated: May 2009.
19 US Food and Drug Administration. Erlotinib prescribing information. www.accessdata.fda.gov/drugsatfda_docs/label/2013/021743s018lbl.pdf Date last updated: May 2013. Date last accessed: January 20, 2014.
20 US Food and Drug Administration. Afatinib prescribing information. www.accessdata.fda.gov/drugsatfda_docs/label/2013/201292s000lbl.pdf Date last updated: July 2013. Date last accessed: January 20, 2014.
21 Lindeman NI, Cagle PT, Beasley MB, *et al.* Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *J Mol Diagn* 2013; 15: 415–453.
22 Schulz KF, Altman DG, Moher D, *et al.* CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c332.
23 Inoue A, Kobayashi K, Maemondo M, *et al.* Updated overall survival results from a randomized phase III trial comparing gefitinib with carboplatin-paclitaxel for chemo-naive non-small cell lung cancer with sensitive EGFR gene mutations (NEJ002). *Ann Oncol* 2013; 24: 54–59.
24 Moher D, Hopewell S, Schulz KF, *et al.* CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010; 63: e1–e37.
25 Mitsudomi T, Morita S, Yatabe Y, *et al.* Updated overall survival results of WJTOG 3405, a randomized phase III trial comparing gefitinib (G) with cisplatin plus docetaxel (CD) as the first-line treatment for patients with non-small cell lung cancer harboring mutations of the epidermal growth factor receptor (EGFR). *J Clin Oncol* 2012; 30: Suppl., 7521.
26 Geater S, Zhou C, Hu CP, *et al.* LUX-lung 6: patient reported outcomes (PROs) from a randomized open-label, phase III study in 1st-line advanced NSCLC patients (pts) harboring epidermal growth factor receptor (EGFR) mutations. American Society of Clinical Oncology. 2013. Abstract 8061.
27 Han JY, Park K, Kim SW, *et al.* First-SIGNAL: first-line single-agent iressa *versus* gemcitabine and cisplatin trial in never-smokers with adenocarcinoma of the lung. *J Clin Oncol* 2012; 30: 1122–1128.
28 Sutani A, Nagai Y, Udagawa K, *et al.* Gefitinib for non-small-cell lung cancer patients with epidermal growth factor receptor gene mutations screened by peptide nucleic acid-locked nucleic acid PCR clamp. *Br J Cancer* 2006; 95: 1483–1489.
29 Angulo B, Conde E, Suarez-Gauthier A, *et al.* A comparison of EGFR mutation testing methods in lung carcinoma: direct sequencing, real-time PCR and immunohistochemistry. *PLoS One* 2012; 7: e43842.
30 Oizumi S, Kobayashi K, Inoue A, *et al.* Quality of life with gefitinib in patients with EGFR-mutated non-small cell lung cancer: quality of life analysis of North East Japan Study Group 002 Trial. *Oncologist* 2012; 17: 863–870.
31 Wu YL, Fukuoka M, Mok TS, *et al.* Tumor response and health-related quality of life in clinically selected patients from Asia with advanced non-small-cell lung cancer treated with first-line gefitinib: *post hoc* analyses from the IPASS study. *Lung Cancer* 2013; 81: 280–287.
32 Chen G, Feng J, Zhou C, *et al.* Quality of life (QoL) analyses from OPTIMAL (CTONG-0802), a phase III, randomised, open-label study of first-line erlotinib *versus* chemotherapy in patients with advanced EGFR mutation-positive non-small-cell lung cancer (NSCLC). *Ann Oncol* 2013; 24: 1615–1622.
33 Yang JC, Hirsh V, Schuler M, *et al.* Symptom control and quality of life in LUX-Lung 3: a phase III study of afatinib or cisplatin/pemetrexed in patients with advanced lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013; 31: 3342–3350.
34 Care Notebook Center. Care Notebook. http://homepage3.nifty.com/care-notebook/en/index.html Date last accessed: January 20, 2014.
35 Andoh M, Kobayashi K, Kudoh S, *et al.* [Using ''Care Note'' to measure the level of satisfaction patients feel with their care, in palliative cancer care, as a measure of their quality of life]. *Nihon Ika Daigaku Zasshi* 1997; 64: 538–545.
36 Kobayashi K, Green J, Shimonagayoshi M, *et al.* Validation of the care notebook for measuring physical, mental and life well-being of patients with cancer. *Qual Life Res* 2005; 14: 1035–1043.

DOI: 10.1183/09059180.00008413

37 Cella D, Eton DT, Fairclough DL, *et al.* What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. *J Clin Epidemiol* 2002; 55: 285–295.

38 Osoba D, Rodrigues G, Myles J, *et al.* Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998; 16: 139–144.

39 Langer CJ. Epidermal growth factor receptor inhibition in mutation-positive non-small-cell lung cancer: is afatinib better or simply newer? *J Clin Oncol* 2013; 31: 3303–3306.

40 Aaronson NK, Ahmedzai S, Bergman B, *et al.* The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993; 85: 365–376.

41 Fayers P, Aaronson N, Bjordal K, *et al.*, eds. EORTC QLQ-C30 Scoring Manual. 3rd Edn. Belgium, European Organisation for Research and Treatment of Cancer, 2001.

42 Bergman B, Aaronson NK, Ahmedzai S, *et al.* The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. *Eur J Cancer* 1994; 30A: 635–642.

43 Earle CC, Weeks JC. The science of quality-of-life measurement in lung cancer. *In:* Lipscomb J, Gotay CC, Snyder C, eds. Outcomes Assessment in Cancer: Measures, Methods, and Applications. Cambridge, Cambridge University Press, 2005; pp. 160–177.

44 Cocks K, King MT, Velikova G, *et al.* Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *J Clin Oncol* 2011; 29: 89–96.

45 Bezjak A, Tu D, Seymour L, *et al.* Symptom improvement in lung cancer patients treated with erlotinib: quality of life analysis of the National Cancer Institute of Canada Clinical Trials Group Study BR.21. *J Clin Oncol* 2006; 24: 3831–3837.

46 Brown H, Prescott R, eds. Applied Mixed Models in Medicine. 2nd Edn. Chichester, Wiley, 2006.

47 European Medicines Agency. Assessment report. Tarceva. www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Assessment_Report_-_Variation/human/000618/WC500117593.pdf Date last updated: July 21, 2011. Date last accessed: January 20, 2014.

48 Peron J, Maillet D, Gan HK, *et al.* Adherence to CONSORT adverse event reporting guidelines in randomized clinical trials evaluating systemic cancer therapy: a systematic review. *J Clin Oncol* 2013; 31: 3957–3963.

49 Minegishi Y, Maemondo M, Okinaga S, *et al.* First-line gefitinib therapy for elder advanced non-small cell lung cancer patients with epidermal growth factor receptor mutations: multicenter phase II trial (NEJ 003 study). *J Clin Oncol* 2010; 28: Suppl., 7561.

50 Eisenhauer EA, Therasse P, Bogaerts J, *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45: 228–247.

51 Therasse P, Arbuck SG, Eisenhauer EA, *et al.* New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000; 92: 205–216.

52 Beau-Faller M, Prim N, Ruppert AM, *et al.* Rare EGFR exon 18 and exon 20 mutations in non-small-cell lung cancer on 10 117 patients: a multicentre observational study by the French ERMETIC-IFCT network. *Ann Oncol* 2014; 25: 126–131.

53 Lopez-Rios F, Angulo B, Gomez B, *et al.* Comparison of molecular testing methods for the detection of EGFR mutations in formalin-fixed paraffin-embedded tissue specimens of non-small cell lung cancer. *J Clin Pathol* 2013; 66: 381–385.

54 Beau-Faller M, Blons H, Domerg C, *et al.* A multicenter blinded study evaluating EGFR and KRAS mutation testing methods in the clinical non-small cell lung cancer setting-IFCT/ERMETIC2 project part 1: comparison of testing methods in 20 French molecular genetic national cancer institute platforms. *J Mol Diagn* 2014; 16: 45–55.