Addendum

LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP (LRiG)

Erlotinib for the first-line treatment of EGFR-TK mutation positive non-small cell lung cancer

This report was commissioned by the NIHR HTA Programme as project number 11/08

Completed 18th January 2012

DOES NOT CONTAIN IN CONFIDENCE DATA

UNIVERSITY OF LIVERPOOL

LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP

A MEMBER OF THE RUSSELL GROUP

1 INTRODUCTION

This document provides further comment from the ERG regarding the indirect comparisons of erlotinib and gefitinib presented by the manufacturer in the MS. It is provided in response to a NICE request arising from the pre-meeting briefing discussions.

2 ERG COMMENTARY ON THE INDIRECT COMPARISONS PRESENTED IN THE MANUFACTURER SUBMISSION

2.1 Description and critique of scenarios

The manufacturer pooled the hazard ratios (HRs) for the gefitinib trials and then conducted a variety of indirect comparisons using different combinations of erlotinib trials as comparators (Table 1).

Table 1 Indirect comparisons presented in the manufacturer's submission

Scenario	Comparison	Indirect PFS HR erlotinib vs gefitinib
1	OPTIMAL compared with IPASS/First-SIGNAL/WJTOG3405/NEJGSG002	0.36 (0.22 to 0.59)
2	Fixed effects pooled estimate of EURTAC/OPTIMAL compared with IPASS/First-SIGNAL/WJTOG3405/NEJGSG002	0.58 (0.41 to 0.81)
3	Random effects pooled estimate of EURTAC/OPTIMAL compared with IPASS/First-SIGNAL/WJTOG3405/NEJGSG002	0.56 (0.24 to 1.28)
4	EURTAC compared with IPASS/First-SIGNAL/WJTOG3405/NEJGSG002	0.82 (0.54 to1.26)

Scenario 1

Scenario 1 compares an Asian population treated with erlotinib with an Asian population treated with gefitinib. If Scenario 1 is accepted as credible, then the OPTIMAL [1] trial must form the basis of the MS, with EURTAC[2] as its support. The ERG notes that in its deliberations regarding the manufacturer's application to extend the existing marketing authorisation, the European Medicines Agency was unable to discuss the results of the OPTIMAL[1] trial as no clinical study report could be made available.

Scenarios 2 and 3

The manufacturer considers that the EURTAC[2] and OPTIMAL[1] trials should not be pooled due to heterogeneity; in that case, the ERG is of the opinion that neither Scenario 2 nor Scenario 3 is credible. The ERG agrees that the magnitude of the heterogeneity presented by the manufacturer in the MS would preclude a pooled analysis of the two trials; however, further investigations by the ERG indicate that the heterogeneity noted by the manufacturer could be the result of using **median** PFS as an outcome measure rather than being due to any of the three factors suggested by the manufacturer in the MS. This issue is discussed in detail in Section 2.2.

Scenario 4

Scenario 4 compares the results of a trial with a European patient population (EURTAC[2]) with the results of trials with Asian patient populations. The ERG considers that this comparison is invalid if it is accepted that these two patient populations are critically different based on differences in ethnicity. The ERG notes the wide confidence intervals around the indirect HR for Scenario 4; these provide little confidence in the estimate.

2.2 ERG critique of the manufacturer's selection of outcome evidence for efficacy of erlotinib

The clinical trial evidence available for use in an indirect comparison of the efficacy of erlotinib compared to chemotherapy doublet regimens for EGFR M+ patients with advanced/metastatic NSCLC is limited to two trials (OPTIMAL[1] and EURTAC[2]). The manufacturer favours use of EURTAC[2] only in the indirect comparison on the grounds that it recruited patients only from non-Asian (predominantly Caucasian) populations and would therefore be more representative of UK patients. This argument rests on the assumption that Asian patients with NSCLC generally have better outcomes than Caucasian patients, and therefore their inclusion in the indirect comparison may distort the estimation of relative efficacy.

This argument has a serious flaw in that the trials identified to reflect the performance of gefitinib in comparison with erlotinib within the indirect comparison network all recruited patients exclusively from East Asian populations, so that if the objective is to avoid confounding evidence from very different geographic sources then the logical approach would be to exclude EURTAC[2] from the analysis, resulting in a consistently non-Caucasian evidence net.

An alternative pragmatic approach is to consider to what extent there is *prima facie* evidence of heterogeneity between trial results, and aim to include as much data in the evidence analysis as is not clearly contraindicated, since the patient numbers recruited to individual trials are quite low and any unnecessary exclusions would inevitably increase the uncertainty in any conclusions drawn from the analysis findings. The implicit assumption underlying such an approach is that differential outcomes in different ethnic populations are not wholly ascribable to ethnicity *per se*, but are the result of differential prevalence of key prognostic factors (in this case activating mutations) in the populations. This hypothesis has been previously proposed by clinical experts.

The manufacturer conducted a meta-analysis based on the comparison of median PFS between the two trials, and concluded that there was evidence of heterogeneity; the manufacturer then speculated on three possible causes of this apparent difference in median PFS (under-performance in the OPTIMAL[1] comparator arm, ethnic differences, and better compliance in OPTIMAL[1]).

The PFS Kaplan-Meier results from the OPTIMAL[1] and EURTAC[2] trials are shown in Figure 1. The comparator arms show very close correspondence across the two trials, and the erlotinib trials follow very similar trends albeit slightly separated. Crucially, across successive time periods the gradients of the cumulative hazard are very similar. It appears likely that the heterogeneity result obtained by the manufacturer is simply a consequence of using the **median** PFS as the outcome measure, since the median is notoriously volatile and a poor indicator of relative performance in the whole population. Without access to detailed patient data from both trials it is not possible to carry out full formal heterogeneity tests. **However, the ERG takes the view that the balance of evidence favours including the two erlotinib trials in any indirect comparison.**

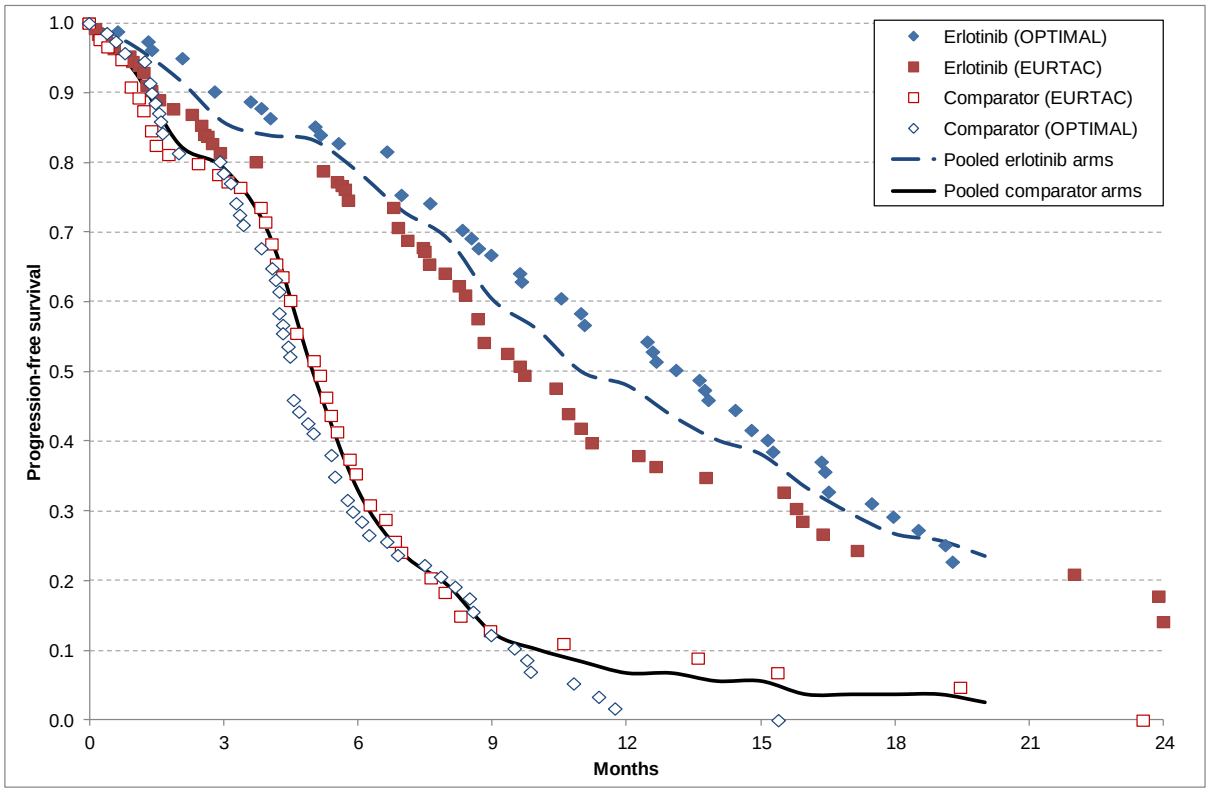

Figure 1 Comparison of PFS outcomes between OPTIMAL and EURTAC clinical trials

2.3 *ERG commentary on the consistency of the outcome evidence for efficacy of gefitinib*

Four gefitinib clinical trials were identified for inclusion in the indirect comparison evidence set. Two trials recruited exclusively from Japanese populations (NEGJSG[3] and WJTOG[4]), one from multiple East Asian locations (IPASS[5, 6]) and a Korean trial including only a very small sub-group of EGFR M+ patients (First-SIGNAL[7]).

In Figure 2 the PFS Kaplan-Meier results for the four comparator arms are compared and show a good correspondence, following the same general trajectory. The corresponding results for the gefitinib

arms (Figure 3) suggest that after 12 months the results for IPASS[5, 6] patients diverge markedly from those of the two Japanese trials, indicating that larger numbers of patients were being identified at regular follow-up assessments than in the other trials.

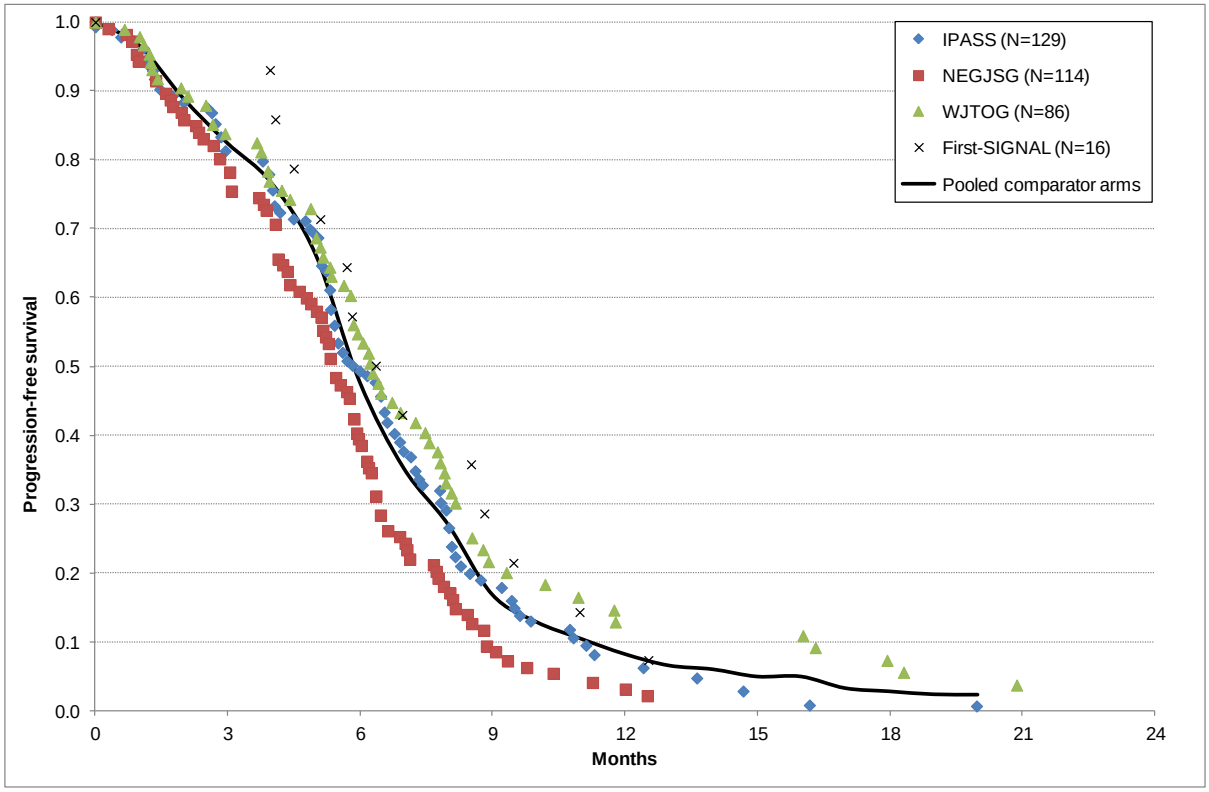

Figure 2 Comparison of PFS between comparator arms in four gefitinib clinical trials

The significance of this difference becomes apparent when the pooled data for the two treatments and two comparator arms are compared in terms of cumulative hazard profiles (Figure 4). It is evident that the outcomes for the chemotherapy arms in both sets of trials are well matched. The pooled gefitinib (intervention) arms appear to show a consistently higher cumulative hazard than that found in the pooled erlotinib arms (lines with yellow infill). However, if IPASS[5, 6] gefitinib results are separated from the other gefitinib trials, the IPASS[5, 6] trend worsens markedly from 12 months onward, and the results for the other trials more closely track the pooled erlotinib results.

This analysis suggests that there may be an important difference between IPASS[5, 6] and all other trials in relation to long-term risk of progression, which at present is unexplained. **The ERG is of the opinion that there is good reason to consider that a sensitivity analysis that excludes IPASS[5, 6] data should be undertaken as part of the indirect comparison exercise. It appears likely that this would show a smaller difference between erlotinib and gefitinib, and consequently increase the size of any estimated ICER.**

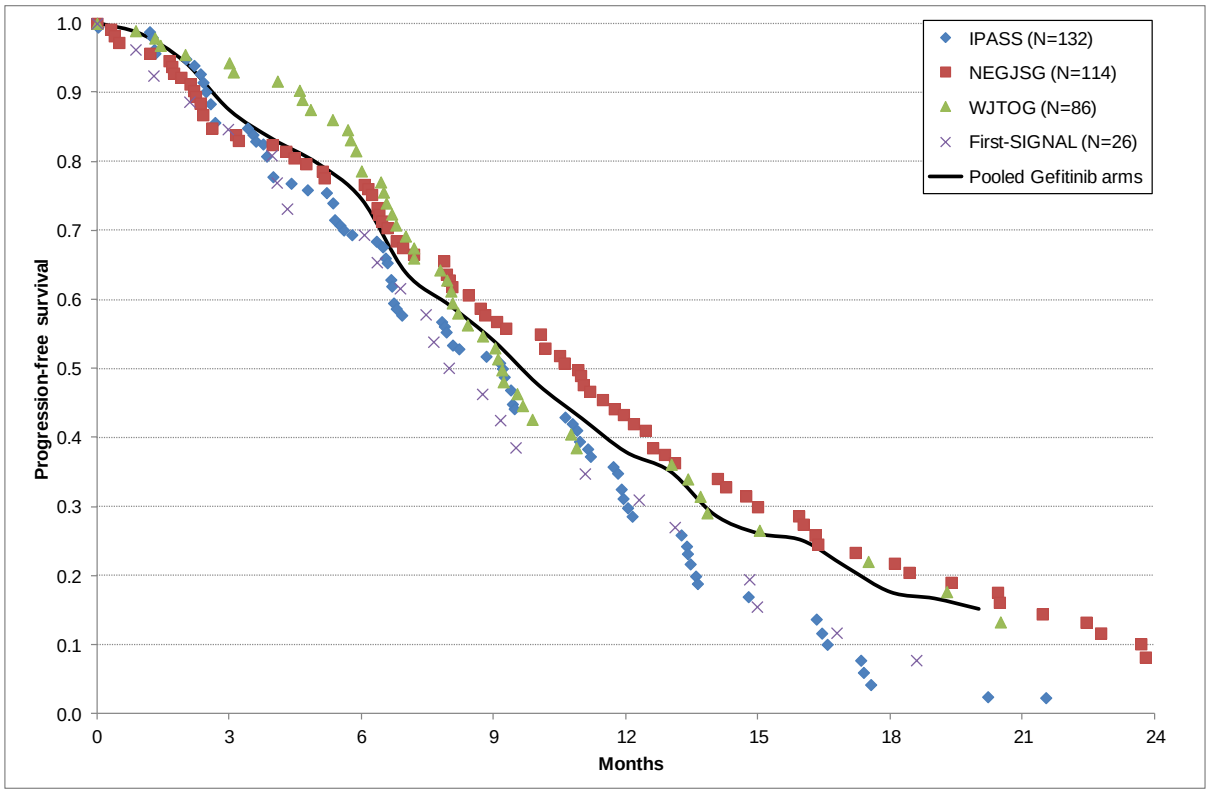

Figure 3 Comparison of PFS between gefitinib arms in four gefitinib clinical trials

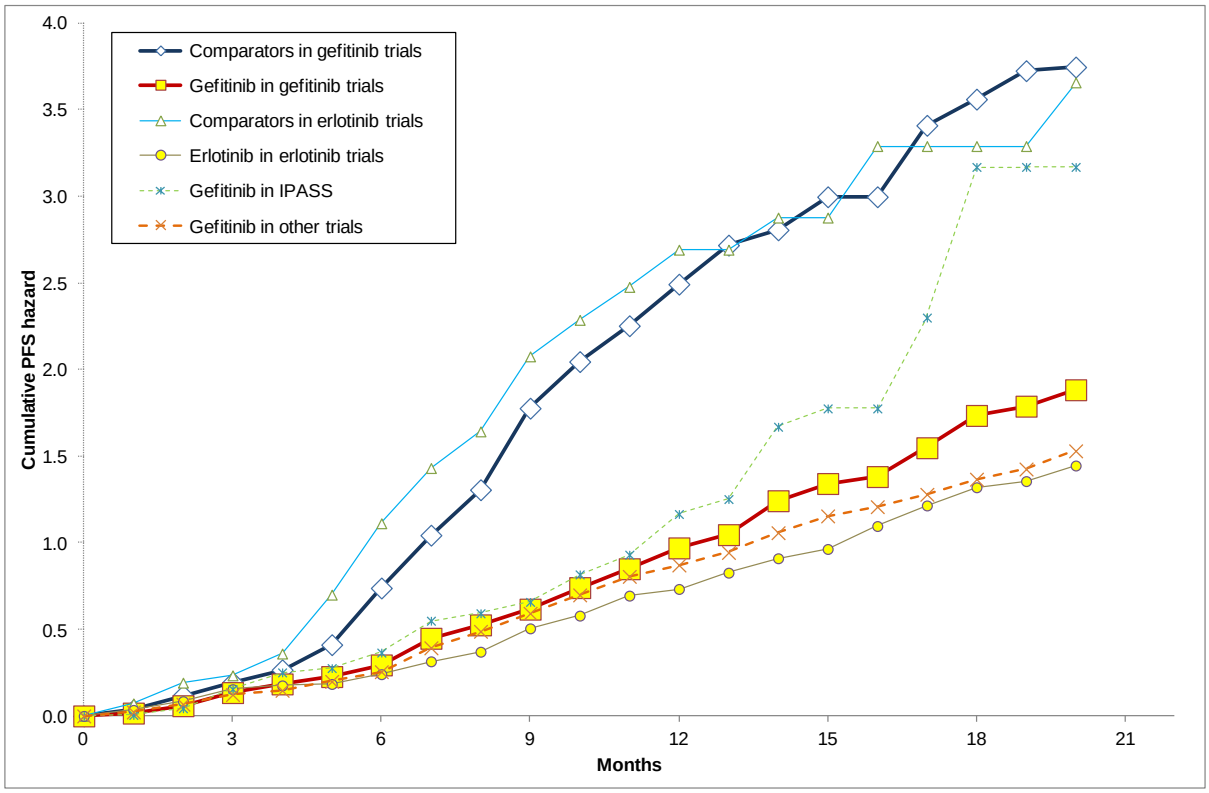

Figure 4 Pooled trends in PFS cumulative hazard between trial arms in erlotinib and gefitinib clinical trials (size of lines/symbols indicative of patient numbers)

2.4 ERG commentary on the manufacturer's estimation of hazard ratios

The manufacturer's model is heavily reliant on the estimated HR between erlotinib and gefitinib derived from indirect comparison, obtained by combining HRs within each trial (i.e. of intervention vs comparator). These calculations rely on the assumption that proportional hazards apply between arms within each trial. Examination of Figure 4 indicates that this assumption is not valid in this case. During the first 4 months of treatment (corresponding approximately to the period of standard chemotherapy), there is very little difference in hazard trends between intervention and comparator arms. However, in the following 2-3 months the trend lines in all trial arms increase, with the comparator arms diverging rapidly away from the erlotinib/gefitinib arms. The assumption of proportional hazards is equivalent to fitting average straight lines through the chart origin across all time periods, and then comparing the slopes of these lines. A more appropriate and accurate method involves treating these two time periods as separate phases of patient experience (equivalent to active therapy followed by observation/maintenance therapy) and deriving separate HRs for each phase (using a landmark analysis for the second phase). It would be necessary to modify the model logic to accommodate this approach, and it is not obvious how this may impact on any modelled ICER estimate. Based on these findings, **the ERG considers that revised relative efficacy measures should be estimated, and tested in any revised model.**

2.5 ERG conclusions

The ERG is of the opinion that the results of the IC presented in the MS by the manufacturer are of questionable reliability. In addition, the ERG notes that the preferred HR used in the manufacturer's indirect comparison will be that which is most likely to favour the manufacturer's case, and considers that the ERG's alternative analyses outlined above are more likely to increase than reduce the size of the estimated ICER for erlotinib compared to gefitinib.

3 REFERENCES

1. Zhou C, Wu Y-L, Chen G, Feng J, Liu X-Q, Wang C, *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(8):735-42.
2. Roche. EURTAC - Clinical study report (ML20650)2011.
3. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, *et al.* Gefitinib or chemotherapy for non–small-cell lung cancer with mutated EGFR.N Engl J Med. 2010; 362(25):2380-8.
4. Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Tsurutani J, *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. The Lancet Oncol. 2010; 11(2):121-8.
5. Fukuoka M, Wu Y-L, Thongprasert S, Sunpaweravong P, Leong S-S, Sriuranpong V, *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-sell lung cancer in Asia (IPASS). J Clin Oncol. 2011; 29(21):2866-74.
6. Mok TS, Wu YL, Thongsprasert S, Yang CH, Chu DT, Saijo N, *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med. 2009; 361(10):947-5
7. Lee J, Park K, Kim S, Lee D, H K. A randomized phase III study of gefitinib (IRESSA) versus standard chemotherapy (gemcitabine plus cisplatin) as a first-line treatment for never smokers with advanced or metastatic adenocarcinoma of the lung. 13th Biennial World Conference on Lung Cancer of the International Association of the Study of Lung Cancer (IASLC); San Francisco 2009.