Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
RR1.	1.	All	All	Any test used to detect EGFR status should be fully validated. Laboratories providing a diagnostic service should participate in external quality assurance through a scheme such as UK NEQAS. This will ensure that resources utilised for the treatment of NSCLC are used appropriately and that patients are optimally treated according to the accurate assessment of all tumour characteristics.	Comment noted. No response required
RR2.	**1.**	11-18	Executive Summary	The report highlights the considerable challenges in *comparative* assessment of EGFR mutation tests as companion diagnostics for TKI treatments for NSCLC. There is no doubt of the considerable benefit patients gain from stratified treatment of NSCLC by EGFR mutation testing and the availability of highly effective treatments with anti-TKIs for EGFR mutation positive patients. However, when attempting a comparative cost-effectiveness assessment of different technologies to test for EGFR mutations the report highlights the considerable challenges that are the result of limited availability of data, in particular *direct head-to-head comparisons* of tests, and the attempt of the authors to base an analysis on data from *indirect* comparisons: The diagnostic test is only one element in a complex and variable patient journey. Costs and treatment outcomes will be strongly influenced not only by the mutation test performance but by other factors, for example the degree of progression of the tumour, patient age, treatment regime, biopsy outcomes and pre-analytic sample	Comment noted. The stakeholder notes issues which were discussed/highlighted in the report.

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
				preparation etc. It is therefore not clear to us to what extent the presented results can be attributed to differences in these variables or to actual differences between test methods compared. We agree with the author's conclusion that "there was no strong evidence that any one mutation test had greater accuracy than any other test". In our opinion no clear recommendations on the relative clinical and cost effectiveness of different test methods can be made based on the findings of this report. We support the conclusion that additional research is required and suggest that research should focus on studies allowing direct comparisons between tests, e.g. studies comparing results on a common set of samples. In addition, appropriate methods need to be considered that allow the incorporation of such direct comparisons into a cost-effectiveness analysis that does not introduce bias.	
	2.	34	3.1.6	Comparative assessment of accuracy based on 'objective response': The use of a common diagnostic "gold standard" across the cited studies was not employed to assess clinical sensitivity and specificity for each reviewed EGFR test. The use of a common test performed in conjunction with the evaluated tests on all samples in the report would allow assessment of true diagnostic test sensitivity and specificity, independent of clinical outcome that is also significantly influenced by other variables. In the absence of	Comment noted. The stakeholder notes the limitations of the available data and states that 'The use of a common test performed in conjunction with the evaluated tests on all samples in the report would allow assessment of true diagnostic test sensitivity and specificity, independent of clinical outcome that is also significantly influenced by other variables.' However, as discussed in the report, even if

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
				a common test between studies, the surrogate endpoint of objective response was used. However, there is considerable demographic variability across the studies identified that could potentially impact objective responses. Although each study supports the use of EGFR activating mutation detection for the selection of patients eligible for anti-EGFR TKIs, the context in which these tests were used was highly variable. For example, there was no adjustment for diagnostic test performance relative to objective response based on gender, age, smoking history or ethnicity. Also, mutations in other genes involved in EGFR signalling may impact patient response but not impact accuracy for EGFR mutation detection.	available such data would not be conclusive; each test identifies a subtly different group of 'mutation positives' and it would therefore remain unclear which test would best predict response to treatment. For this reason, 'true sensitivity and specificity' for EGFR mutation tests can only be determined with respect to response to treatment.
	3.	38	3.2.1	Technical performance of tests in the literature: The DAR report identified only one publication reporting technical performance characteristics, such as failure-rates, on the TheraScreen® Kit. However, for CE marked comparator tests data on technical performance is also available from the manufacturer's information that should be considered as a key source of information. An additional study that was published after the completion of the DAR report in *The Journal of Clinical Pathology* (Lopez-Rios F, Angulo B, Gomez B, *et al. J Clin Pathol Published* Online First: 05 Feb, 2013 doi:10.1136/jclinpath-2012-201240) provides key evidence on the comparative technical performance of different test methods. This study demonstrated that the overall agreement between the	Studies of this type were included as a protocol extension, with the aim of collecting data similar to that provided by the survey of laboratories (i.e. data on the real practical experience of UK laboratories in processing clinical samples). Manufacturers' information from kit inserts is included in the background to the report, but does not provide information on laboratory experience with clinical samples. The additional study cited does not meet the inclusion criteria for the External Assessment Group's review. As noted in response to comment 2, studies of this type are not informative as they only show agreement

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
				TheraScreen® EGFR29 Mutation kit and the cobas® EGFR Mutation Test was 99.2% for mutations that were detectable by both tests, supporting the premise that both PCR tests can equally provide actionable results as demonstrated by the evaluation of tests in the DAR. The cobas® EGFR Mutation Test was also compared *directly* to Sanger sequencing *on the same set of samples.* Out of 133 samples tested 7 were discordant between the two tests. However, all 7 samples were resolved with massively parallel pyrosequencing in favour of the cobas® EGFR Mutation test. Clinically, 6 cases would have not been eligible for anti-EGFR therapy by Sanger sequencing although an EGFR mutation was confirmed. The other case was a false positive by Sanger sequencing and this patient would have been exposed to ineffective therapy. Also in this study, the cobas® EGFR Mutation kit resulted in 1/248 (0.4%) invalid samples and Sanger sequencing resulted in 23/124 (18.5%) invalid samples after one run and 5/124(4.0%) after a second run. As part of the analytical validation of the cobas® EGFR Mutation test kit (using 2 lots of reagents), for CE-IVD approval, there were 48 invalid samples with Sanger sequencing and 5-6 invalid samples with cobas® EGFR Mutation kit (Table 6 of CE-IVD package insert). For the 152 valid samples between both methods, there were 5 discordant samples. After resolution of discordant cases with pyrosequencing, 3-4 samples resolved in favour of the	between two tests which essentially have different definitions of a positive mutation (different target mutations and limits of detection). If tests results are not related to clinical outcome, then it is not possible to determine whether a mutation detected by one test and not the other would in fact have resulted in more appropriate treatment.

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
				cobas® EGFR Mutation kit. Therefore, the overall invalid rate for Sanger sequencing during controlled analytical validation studies was 24% and was 3% for the cobas® EGFR Mutation kit (Table 10 of CE-IVD package insert). The utility of CE-IVD PCR-based EGFR mutation tests such as the TheraScreen® EGFR29 Mutation kit and the cobas® EGFR Mutation Test also relates to less demand on tissue requirements, greater sensitivity and high reproducibility relative to sequencing methods. As current demands on clinical tissue samples continue to increase, careful consideration should be given to test preferences.	As discussed in the Diagnostics Assessment Report, ever increasing sensitivity does not necessarily translate to increased clinical benefit unless it can be shown that tumours with lower levels of mutation benefit from treatment with TKIs.
	4.	43	3.2.1 Table 4	Technical performance characteristics of test as reported by UK laboratories: There are several remarks regarding the data summarized in Table 5 that highlight the difficulties in using this information to compare different test: a) It is not entirely clear how representative the presented data from the thirteen laboratories participating in the survey is for the clinical practice in England and Wales. b) It is not clear what methods were used to estimate the key technical performance criteria, for example the percentage of failed samples. More information on the number of samples analysed at each site and whether numbers were based on a systematic analysis of test	Comments noted. a) Agree, but the External Assessment Group was limited by the responses provided. b) Agree, this issue was not raised during piloting, but might usefully be included in future similar surveys. However, it does not seem likely that inclusion of these data would alone result in the identification of significant differences between testing methods. c) See previous response. A statement from UK NEQAS is included in the Diagnostic Assessment Report: “Error rates are not always method

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
				records or estimates would be useful. c) The pre-analytical variables for each testing method were not clearly defined. Key requirements for each test should be in the comparison table with: 1) numbers of tumour FFPET sections, 2) % of tumour cells, 3) minimum tumour area per extraction and 4) total DNA per testing. Without pre-analytical data the performance criteria reported may not be comparable across tests as they also reflect different laboratory practices across the reporting sites. For example, an analysis of failure rates of the cobas® EGFR Mutation Test is biased if the test is primarily used for samples with low tumour content at the site. Variability of laboratory practice in the UK is also highlighted in the recent UK NEQAS publication (Deans ZC, Bible N, O'Sullivan B *et al. J Clin Pathol* Published Online First February 1st 2013) reporting a genotyping error rate of 6.4% (3/47 submissions) for EGFR mutation testing. d) It would also be important to know whether test failures reported relate to the actual mutation test only or included all pre-analytic steps from biopsy. The latter performance criteria will ultimately determine the percentage of patients receiving the best treatment option. For example, it would be important to know the number of samples not undergoing EGFR mutation testing because of low tumour content at histological examination in different laboratories.	related and it is not always possible to obtain data from all the labs committing critical genotyping errors. Therefore, any data which could be provided would be skewed with processing and reporting issues rather than being method related. There has been no correlation between any method used for EGFR testing and errors since we started providing scheme in 2010." d) See previous responses – Economic modelling distinguishes between analytical and pre-analytical failure, however, The External Assessment Group's ability to do this fully was limited by the available data. As a result, in the analyses the difference between analytical and pre-analytical failure was only incorporated as an impact on costs. However, the total number of unknowns (either due to analytical failure or pre-analytic failure) from a test was modelled to have impact on prognosis in all analyses.

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
	5.	49-51	3.2.2 Table 7	Variability of accuracy of sequencing methods based on 'objective response': The report identified five studies reporting on direct sequencing results from a total of 236 available samples based on objective response or disease control. The "objective response" sensitivity ranged from 60-84% and the specificity ranged from 61-96%. These outcomes were highly variable and could reflect either the disparate populations tested or variability in the tests; standardized methodologies are needed to control analytical and clinical performance across laboratories and between test methods and failure to fully acknowledge this variability would lead to misinterpretation of the results: For all of the studies, the number of invalid samples and tests were either not reported or there was insufficient material for analysis. Additionally, re-testing rates were not presented or consistent across sites. It appears that one study (Fukuoka M, Wu YL, Thongprasert S *et al. J Clin Onol* 2011; 29(21):2866-2874) modified the test cut-off parameters that reduced the rate of false positives. Therefore, it is important to review the pre-analytical methods for each study to confirm that the test performance endpoints are comparable between sites. The authors do not always distinguish clearly between different sequencing methods and there is a possibility the reader treats "Direct Sequencing or Sanger Sequencing" like one method. In fact, the procedure and protocol for Sanger sequencing could be very different from lab to lab, unlike for commercially available CE-IVD marked PCR	Comment noted. The limitations of the available data were highlighted and discussed in the Diagnostics Assessment Report. The External Assessment Group acknowledges that test methods such as 'direct or Sanger sequencing' and pre-analytical sample processing may vary in their application between studies, however, unfortunately, these data are not generally well reported.

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
				Tests, such as the TheraScreen® EGFR29 Mutation kit and the cobas® EGFR Mutation test, that are validated and run under standard protocols. For example, in the EURTAC trial (Rosell R, Carcereny E, Gervais T *et al. Lancet Oncol* 2012; 13(2):239-246), even though Sanger sequencing was used to confirm mutation the genomic DNA was from the tumor cells that were Laser Captured Microdissected and the amplicon used for Sanger Sequencing were the products after two runs of PCR.	
	6.	71	4.2	Economic analysis: The presented methodology highlights the considerable challenges due to the data available and considered in the economic analysis. If one interprets the assumptions made correctly, *no head-to-head comparative test data was used across all comparators* in the three analyses presented: -The 'evidence on comparative effectiveness available' analysis is in essence an indirect naïve comparison between different trials that used different test methods to stratify patients but also had different patient populations and treatment regimes. -The 'linked evidence' analysis is again based on indirect comparisons of different test methods based on test accuracy derived from studies that had significant heterogeneity in terms of patient population and study design, as pointed out in the comments on the clinical	Thank you for your comments. As described in the Diagnostics Assessment Report, no head-to-head comparative test data were available. As noted in the conclusions of the report, the outcomes of the assessment of cost-effectiveness should be interpreted with extreme caution as a result of the assumptions made to cope with this lack of head-to-head comparative test data. In particular, the assumption that the differences in treatment response and survival between tests as observed between the different studies are solely due to the different tests used. This ignores all other factors that can explain variations in outcomes between the studies.

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
				effectiveness section of this report. - The 'assumption of equal prognostic value' analysis is based on a survey of UK laboratories and differences in reported test performance may be due to differences local practices rather than attributable to the actual test performance. Therefore, results on comparative cost-effectiveness reported are likely to be biased by a variety of variables other than the actual comparative performances of the test methods used because of the indirect comparisons entering the analysis.	
	7.	93	4.3.2	Sensitivity analysis: A more detailed deterministic sensitivity analysis and discussion on the influence of assumptions on the result may be useful: for example, in the 'assumption of equal prognostic value' analysis, exploring the sensitivity of the results on differences in test sensitivity, specificity and failure-rate might have allowed for a discussion of the relative importance of these parameters for cost-effectiveness. For example, one would expect a test with a lower failure rate, such as **cobas® EGFR Mutation Test,** to be more cost-effective because the actual test costs reported in table 22 of the report vary little across different methods and are small compared to the overall cost of treatment.	Comment noted. The External Assessment Group acknowledges that the tests differ in terms of failure rate, and test performance. However, as no information is available with regard to the comparative prognostic value of the tests, it is impossible to model test sensitivity and specificity for all the tests included in the equal prognostic value analyses. The difference in failure rate is part of this analysis. As failure rate and test costs are the only parameters that differ between the tests in this analysis, these parameters are, of course, influential. See also response on comment 4. Also, it should be noted that the cobas® EGFR

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
					Mutation Test is not the test with the lowest failure rate. See Table 5 and Table 23.
RR3.	**1.**	113	5.2.2	*""Progression-free survival and overall survival were modelled using the Weibull regression models based on the IPASS study and a hazard ratio for TKI (based on a meta-analysis and mixed treatment comparison) used in the NICE Technology Appraisal."* It is not clear from the report whether the meta-analysis (MA) and mixed treatment comparison (MTC) conducted by AstraZeneca for TA192 were updated to evaluate the cost-effectiveness of different EGFR mutation tests. If the MA and MTC were not updated we suggest this is acknowledged as a limitation of the cost-effectiveness analysis in the Discussion section 5.2.2	The hazard ratios and odds ratios (as presented in Table 19 of the reported) were retrieved from the updated mixed treatment comparison from NICE Technology Appraisal 192.
	2.	61	3.2.3	*"In the IPASS trial, the OR rates for mutation negative participants were 1% (1/91) for the TKI group and 24% (20/85) for the standard chemotherapy group, and for participants whose mutation status was unknown the OR rates were 43% (167/386) for the TKI group and 29% (115/394) for the standard chemotherapy group. The first-SIGNAL trial reported similar data on OR rates for participants whose tumours tested negative for EGFR mutations (26% (7/27) for the TKI group and 52% (14/27) for the standard chemotherapy group)"*	We do not believe that it would be helpful to extract this statement alone from the discussion of Han 2012. It is true that EGFR mutations 'missed' by the direct sequencing method used in Han 2012 (FN), due to the relatively high proportion of tumour cells and percentage of cells with mutation required by this method, is one possible explanation for the higher OR rate observed in the mutation negative group this trial. However, differences between the trial populations are also important (as noted by Han

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
				Suggest the following text is added after the text above: The authors of First-Signal (Han 2012) commented that the higher OR observed in First-Signal versus IPASS in EGFR mutation –ve patients treated with gefitinib (26% vs 1%, respectively) was likely to be a result of a higher rate of false negatives (FN).	et al.). The External Assessment Groups believes that the Diagnostics Assessment Report provides a balanced discussion of the issues.
	3.	113	5.2.2	The de novo probabilistic model developed to evaluate the cost-effectiveness of EGFR mutation testing in adults with NSCLC has not adjusted for the costs or the QALYs accrued as a result of the treatment cross-over that occurred in both IPASS and FIRST SIGNAL. In IPASS (Fukuoka 2011), 64.3% of EGFR m+ patients in the carboplatin/paclitaxel treatment arm were treated with an EGFR-TKI on discontinuation of their 1st-line treatment. Similarly, in FIRST SIGNAL (Han 2012) 75% of patients given cisplatin/gemcitabine first received an EGFR-TKI on treatment discontinuation. It is likely that the substantial cross-over from 1st-line platinum-based doublet chemotherapy to an EGFR-TKI may explain the reason why no significant difference in OS between gefitinib and platinum-based doublet chemotherapy has been observed.	Comments noted. Indeed, the de novo probabilistic model did not adjust the outcomes for possible cross-over effects in the IPASS and FIRST SIGNAL trials. This is because the de novo model had to be consistent with TA 192 and should not update the results and recommendations from TA 192. Hence it was not indicated to correct the differences between gefitinib and platinum-based doublet chemotherapy for possible cross-over effects since this may lead to inconsistencies with TA 192.

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
	4.	75		Figure 10: Decision tree structure Anti-EGFR-TKI amend to EGFR-TKI	The External Assessment Group will correct this ahead of publication.
	5.	79	Table 17	Please check the ORR for Negative patients using the Direct sequencing of all exon 19-21 mutations. The figure reported in the table is 0.484 but the paper (Han 2012) gives a figure of 0.519 (see page 1125 Subgroup Analysis).	The OR rate has been adjusted (based on the updated mixed treatment comparison from STA 192) because Han 2012 used a different standard chemotherapy regime (Gemcitabine and Cisplatin); this is explained and the original OR rate (0.519) reported in the footnote to Table 17.
RR4.	**1.**	39	3.2.1	We have identified that the following statement refers to the survey response from our laboratory: *"The third [laboratory] use Sanger sequencing, TaqMan/Real Time PCR/Entrogen and Fragment Length Analysis and also cite verification of mutations and insufficient tumour cell as their reason for using multiple tests."* These methods are employed for samples where sequence analysis has failed. This could be due to poor quality DNA (fragmented/degraded) and is not employed to compensate for low tumour load.	Comments noted. Survey data are reported as provided; this respondent selected both 'verification of mutations' and 'insufficient tumour cells' as reasons for using more than one EGFR testing method. The External Assessment Group will add the additional information provided ahead of publication.
	2.	44	3.2.1	A study by UK labs has compared the sensitivity of different methods for the detection of EGFR mutations. The limit of detection was 1-7.5% depending on mutation tested and method used (manuscript attached).	Comment noted. Though interesting, this study does not meet the inclusion criteria for the External Assessment Group's review and does not report the necessary data for inclusion in the

Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer

Diagnostics Assessment Report (DAR) - Comments

Responder reference no.	Comment no.	Page no.	Section no.	Comment	Response
					'equal prognostic value' model.
	3.			Sensitivity and failure rate of mutation detection methods will be influenced by fixative used for sample (biopsy or resection). Our experience suggests that using Cytofix to prepare samples gives less failures than conventional FFPE.	Comment noted. Information on fixatives was not collected in the laboratory survey (not suggested during piloting). Inclusion of this information in future similar surveys could be considered.
	4.	45	3.2.1	Column 2 of table states minimum % tumour cells required. Has this been misinterpreted and the figure given is actually the sensitivity of the methods employed by that laboratory?	The survey question was phrased 'What is the limit of detection of the EGFR test in terms of the % tumour cells?' Data are reported as provided, and the External Assessment Group did not receive any feedback suggesting problems with this question. However, misinterpretation of survey questions is always a possibility.
RR5.	**1.**	n/a	n/a	This report has reached the same conclusion as very many of us within the diagnostic community who are currently testing for EGFR mutations in NSCLC would have advised (and did advise at the initial stakeholders meeting). I therefore hope that the views of the diagnostic community are gauged before similar, lengthy and expensive appraisals of biomarker technologies are undertaken in the future.	Comments noted.