**Article 25fa pilot End User Agreement**

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) ‘Article 25fa implementation’ pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication other than authorised under this licence or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library  
Radboud University  
Copyright Information Point  
PO Box 9100  
6500 HA Nijmegen

You will be contacted as soon as possible.
Molecular Diagnostics

Accreditation, setting and experience as indicators to assure quality in oncology biomarker testing laboratories

Véronique Tack¹, Ed Schuuring², Cleo Keppens¹, Nils 't Hart², Patrick Pauwels³, Han van Krieken⁴ and Elisabeth M.C. Dequeker¹

BACKGROUND: Predictive biomarkers allow clinicians to optimise cancer treatment decisions. Therefore, molecular biomarker test results need to be accurate and swiftly available. To ensure quality of oncology biomarker testing, external quality assessments (EQA) for somatic variant analyses were organised. This study hypothesised whether laboratory characteristics influence the performance of laboratories and whether these can be imposed before authorisation of biomarker testing.

METHODS: Longitudinal EQA data from the European Society of Pathology were available over six (metastatic colorectal cancer) and four years (non-small cell lung cancer), including the percentage of analysis errors and technical failures, and information on laboratory characteristics (accreditation status, laboratory setting, number of samples analysed and detection method). Statistical models for repeated measurements were used to analyse the association between the EQA results and the laboratory characteristics.

RESULTS: Laboratory accreditation was associated with fewer analysis errors in early stages of biomarker introduction into the laboratory. Analysing more samples, or university and research laboratories showed better performance. Changing the detection method did not have an effect.

CONCLUSION: The indicators support the clinicians in choosing molecular pathology laboratories by improving quality assurance and guaranteeing patient safety. Accreditation of laboratories, centralisation of biomarker testing or a university and research setting should be stimulated.


INTRODUCTION
Pathology laboratories are challenged to maintain high quality assurance due to the constant pressure of newly developed technical and medical expertise. Research on predictive biomarkers has gained considerable momentum, highlighting the need for regular updates of test strategies.¹,²

Predictive biomarkers allow clinicians to predict clinical effects of cancer treatments.³ For this purpose, biomarker test results need to be very accurate and swiftly available. For metastatic colorectal cancer (mCRC), the confirmation of a wild-type KRAS and NRAS gene is required before anti-EGFR monoclonal antibody therapies.⁴–⁷ In the context of non-small-cell lung cancer (NSCLC), treatment with EGFR tyrosine kinase inhibitors requires the presence of an activating variant in the EGFR gene. Moreover, ALK and ROS1 gene tyrosine kinase inhibitors (TKI) treatment can only be administered to patients with a confirmed gene rearrangement in the mentioned genes.⁸–¹¹ Recently, the immune checkpoint inhibitor, pembrolizumab, was approved by the US Food and Drug Administration (FDA) in first line therapy for patients with ≥50% positive PD-L1 expression in NSCLC, as assessed by immunohistochemistry (IHC).¹² The immune checkpoint inhibitor nivolumab can be given to patients with NSCLC without confirmation of PD-L1 positivity.

To ensure quality of oncology biomarker testing, external quality assessments (EQA) are organised by the European Society of Pathology (ESP) according to international standards.¹³ Remarkably, each year a significant number of laboratories are not able to correctly identify the relevant variants.¹⁴,¹⁵

Previous research by our group in 2014 demonstrated that laboratories were not able to rapidly introduce new accurate biomarker testing for the use of cetuximab and panitumumab, two anti-EGFR therapies for patients with mCRC.¹⁴ On the other hand, multiple participations in EQA programs showed improved performance rates.¹⁵,¹⁶ However, the number of genotyping errors remains high.¹⁴,¹⁵ Laboratories are also challenged by the large variation in mutation types, for example targeted variants in the EGFR gene include point mutations, insertions and deletions, which evolve over time.¹⁷ With the large number of ongoing phase-III clinical trials, and the rapid developments in biomarker testing, several new testing strategies and relevant biomarkers are expected to be implemented each year.¹⁴,¹⁸ The insufficient performance of laboratories in EQA together with the increasing number of biomarker tests are subjects of concern. Laboratories...
should be stimulated to maintain a high degree of quality assurance by means of continuous education and development of expertise in oncology biomarkers. Several longitudinal studies were already performed on data from EQA in oncology biomarker testing, but the influence of specific laboratory characteristics on the performance of oncology biomarker testing has, to date, not yet been analysed.19

Laboratory characteristics, such as accreditation, experience and laboratory setting or organisational structure, are recognised as important elements that could influence the performance of laboratories.20–22 How can the hypothesis be justified that accreditation or a minimal level of experience can be imposed before authorisation of biomarker testing?

The aim of this study was to correlate laboratory characteristics with EQA results in biomarker testing for NSCLC and mCRC in order to identify important quality indicators.

MATERIALS AND METHODS

The EQA schemes of the ESP were coordinated by the Biomedical Quality Assurance (BQA) Research Unit of the KU Leuven according to international guidelines.13,23 The schemes were open to all laboratories worldwide (Table 1). The participants had to report their results in an online datasheet, together with questions on laboratory characteristics and testing strategies. It was mentioned in this datasheet that data could be used for further research. In addition, mock diagnostic reports were requested based on the information provided by the BQA Unit of the KU Leuven. Medical and technical experts assessed the data provided by the participants in collaboration with the EQA coordination centre. A laboratory received full marks (2 points) when the genotype of the requested biomarker was correctly assigned; points were deducted in case of clerical errors, technical failures, analysis errors or nomenclature errors. Each sample result was evaluated by two assessors independently, and equivocal results were discussed during a meeting. Samples for which more than 25% of the laboratories could not obtain a reliable result were not considered for the analyses.24 These were defined as educational samples.

The EQA schemes of the ESP can be categorised into two groups: those for variant analysis (RAS genes, including KRAS and NRAS, in mCRC and the EGFR gene in NSCLC) and those for gene rearrangement and expression analyses (ALK and ROS1 rearrangements in NSCLC).

For variant analyses, participants should use their routine procedures to analyse the EQA samples. The sent samples were formalin-fixed paraffin-embedded (FFPE) material from the resection specimens, commercial FFPE reference standards (Horizon Diagnostics, Cambridge, UK, http://www.horizondx.com and Thermo Fisher Scientific, Fremont, California, www.thermoscientific.com/qc) or cell line material.19 The cell line material was chosen as a pilot for standardisation of homogeneous samples for testing. The percentages of neoplastic cells varied between 10 and 100%.

For gene rearrangement and expression analyses, both IHC and fluorescent in situ hybridisation (FISH) techniques were included in the EQA schemes. The samples included FFPE cell lines on glass slides as well as tissue microarrays from several FFPE blocks from either the cell line material or the resection specimen.19

The association study included the EQA results as the dependent variable and the laboratory characteristic as the independent variable (Table 2). For the EQA results, the total EQA scores were taken into account, but also, the isolated percentage of analysis errors or the number of technical failures were considered. Other specific error types were not analysed individually, but remain integrated in the EQA scores. Analysis errors consisted of false positive (reported a variant, rearrangement or aberrant expression in a wild-type sample), false negative (wildtype was reported in a tumour containing a variant, rearrangement or aberrant expression) or wrongly reported variants (a variant was found, but it was an incorrect variant).

Technical failures arose when no conclusive results were reported by the participant, either due to a limited amount of DNA recovered from the sample or a doubtful result.

Four laboratory characteristics were used for association with EQA results of the laboratories, namely the accreditation status of a laboratory, their setting, the number of samples tested each year and the change in used methods. The corresponding subcategories and number of observations per characteristic can be found in Table 2. Each observation of a characteristic denotes the participation of a specific laboratory in a specific EQA scheme. For the accreditation status, different national and international standards were considered. The ISO 15189 and the ISO 17025 standards were included as recognised international accreditation standards.25,26 Several national accreditation standards, such as CAP 15189 (College of American Pathologists), and some widely used national standards were also taken into account for accreditation. Examples of the latter are the national standard in the Netherlands (CCKL) and the standards of the Clinical Pathology Accreditation in the United Kingdom.27–30 For good readability, both recognised accreditation standards and their equivalents are hereafter referred to as ‘accreditation’. No distinction was made between accreditation standards and their equivalents for further analysis. The reported accreditation status was retrieved from the websites of the relevant national accreditation body. As there is still a lot of confusion in laboratories about accreditation status, these were verified retrospectively.20 In case no archived data were available, these data were defined as ‘missing data’ in Table 2. Also, the effect of the setting of a laboratory was studied. The reported setting of the laboratories in the EQA questionnaires was verified on the laboratory’s website. Different categories were assigned and this terminology will be used throughout the manuscript (Table 2). Industry denotes those laboratories involved in the development of diagnostic commercial kits. Private laboratories are laboratories that are not present within a hospital’s infrastructure. Hospital laboratories included private and public hospital laboratories. The setting defined as ‘university and research’ included education and research hospitals, university hospitals, university laboratories and anti-cancer centres, all with a clinical demand for biomarker testing. The number of samples tested each year and the used methods were requested during EQA participation and were not further validated. As not all participants provided information on the number of samples, there are some missing data for this characteristic too. Examples of commercial methods (excluding commercial methods based on next-generation sequencing (NGS)) were the Therascreen Pyro Kit (Qiagen), the RAS mutation analysis kit (EntroGen) or the Idylla Mutation Test (Biocartis). Non-NGS-based laboratory-developed techniques (LDT) included Sanger sequencing or TaqMan allelic discrimination assays (Fig. 2).

Longitudinal data between 2010 and 2016 were analysed using statistical models for repeated measurements. Because of the skewed distributions of the EQA scores, a binary outcome was used: <90% (unsuccessful participation) and ≥90% (successful participation). In case an EQA scheme only contained nine samples that were eligible for scoring, because of an educational sample, ≥88.9% was also considered as a successful participation. For binary outcomes (successful participation versus unsuccessful participation), generalised estimating equations (GEE) were used to account for clustering. For continuous outcomes (percentage of analysis errors or technical failures), an unstructured residual covariance matrix was modelled to account for clustering.

All tests were two-sided and a p-value of <0.050 was considered significant. Pairwise comparisons were only performed in case of significant results. Analyses were performed with SAS software for Windows (version 9.4).
RESULTS
This study investigated the hypothesis that whether laboratories would have to comply to specific laboratory characteristics before the authorisation of biomarker testing. Three factors, namely, the accreditation status of a laboratory, their setting and their experience with oncology biomarker testing in mCRC and NSCLC were considered. The latter included the number of samples tested each year and the change in used methods.

Accreditation
Based on a total dataset of 691 observations (2010–2016) for variant testing in mCRC, no association was found between the accreditation status of laboratories and their EQA score for variant analysis in KRAS or NRAS genes (Table 3). However, there was a positive effect of accreditation on the percentage of analysis errors for the NRAS biomarker (incidence rate ratio (IRR) = 0.52, p = 0.023) in the first years of NRAS implementation in the laboratory (2013–2014). Pairwise comparisons of different accreditation categories in this group showed that accredited laboratories have 47% fewer analysis errors compared with laboratories without accreditation (IRR = 0.53, p = 0.030). This significant effect disappears (p = 0.080; 2013–2016) over time, indicating that accredited laboratories have a better implementation procedure (Table 3).

The number of years that a laboratory was accredited had no effect on the EQA result (p = 0.152) or the percentage of analysis errors (p = 0.871) for NRAS and KRAS testing laboratories between 2012 and 2014.

For EGFR testing (2012–2016), the total number of observations for accreditation characteristic was 453 (Table 2). An association was found between accreditation and successful EQA scores (p = 0.018) (score ≥90%), and between accreditation and fewer analysis errors (p = 0.002). The percentage of technical errors showed no association with the accreditation status (p = 0.422) (Table 3).

The EQA schemes for rearrangement analysis included the ALK (2012–2015) and the ROS1 (2014–2015) biomarker (Table 1). For rearrangement and expression analysis of ALK and ROS1 genes by FISH or IHC, there was no effect between the EQA score or the percentage of analysis errors and the accreditation status (Table 3). However, accreditation is associated with less technical failures for ALK IHC (p = 0.009) and more technical failures for ROS1 FISH (p = 0.031). Pairwise comparison of different categories showed evidence that the coverage of ALK is not specifically required in the scope (Table 3), however, for ROS1 FISH, ROS1-specific accreditation is important (p = 0.017).

Figure 1 shows the statistical results of different characteristics as an illustrative summary of the study.

Setting
There is no evidence of an association between the setting of an institute and its EQA score or percentage of analysis errors in KRAS of NRAS testing (n = 711). For RAS analysis between 2013 and 2016 (n = 348), there is a significant higher probability to obtain a successful EQA score (≥90%) in the setting of university and research versus hospitals and (private) laboratories without the research incentive (p = 0.010 and p = 0.037, respectively). University and research settings were associated with less analysis errors than hospitals and (private) laboratories, and an industry setting showed less analysis errors compared to hospitals and (private) laboratories (p = 0.013 and p = 0.012, respectively). In EGFR testing, (n = 460) no association with the EQA scores was found, however, less analysis errors were made in a university and
research background, compared to (private) laboratories and industry laboratories \((p = 0.016)\) and \((p = 0.012)\), respectively.

For the rearrangement analyses, \((n = 494, n = 368, n = 124\) and \(n = 61\), for ALK FISH, ALK IHC, ROS1 FISH and ROS1 IHC, respectively) (Table 2), no significant associations were identified.

Experience with oncology biomarker testing

Table 4 shows an increasing probability to have a successful EQA score with increasing number of samples tested per year for KRAS, NRAS or EGFR \((p = 0.002; p = 0.009; p = 0.009)\), respectively. Similar results were obtained for the association with the percentage of analysis errors, except for the number of samples tested per year for EGFR (Table 4). In case the number of samples was considered as an ordinal variable, we observed a decreasing number of analysis errors with increasing category of the number of samples. For KRAS, an IRR of 0.72 indicates a decrease of 28% of genotype mistakes for every one-level increase, for example from <10 samples tested to 10–99 or from 10–99 to 100–249 samples tested. With number of samples as categorical variable, the number of analysis errors is significantly lower in laboratories that test 10–99 samples compared with laboratories that test more than 99 samples, both for KRAS and NRAS.

For Lung ALK FISH, a higher number of samples tested per year was associated to an improved EQA score and a decreased percentage of analysis errors (Table 4). For an increase in level of the samples tested per year, 32% fewer analysis errors were made with ALK FISH analysis (IRR \(= 0.68\), \(p = 0.005\)). This is in contrast with results for ROS1 rearrangement testing, where no association could be observed between the ROS1 FISH EQA scores or analysis errors and the number of samples tested for ROS1.

For IHC tests, there was evidence of variation between EQA scores and the categories of number of samples tested, however, there is no conclusive evidence that the EQA score improved with an increasing number of samples tested per year. The percentage of analysis errors showed a difference between the number of ALK IHC testing categories, but no conclusive results could be made that less analysis errors were made when more samples were tested for ALK. For ROS1 IHC, however, a 36% increase of ROS1 IHC analysis errors was seen for every increase in the level of number of samples tested per year.

In addition to the number of samples tested in a laboratory, changing a routinely applied method can also influence the experience of a laboratory with a specific biomarker. The methods used in molecular pathology laboratories vary greatly. Figure 2
gives an overview of the evolution of the distribution of different types of methods over the years for the RAS and EGFR biomarkers. In general, less LDTs were used over the years, while non-NGS-based commercial methods and NGS-based techniques gained more weight. At the time that a new predictive biomarker was introduced in the routine, a limited number of non-NGS-based commercial test kits was available. The implementation of a new method for KRAS variant analysis (2010–2014) showed no significant association with the EQA score or the number of analysis errors ($p = 0.471$, $p = 0.871$, respectively). As not enough data were available on other biomarkers, no further analyses could be performed.

**DISCUSSION**

The changing era of personalised medicine challenges laboratories to offer quality-assured biomarker testing and maintain a high performance level for their tests. Different laboratory characteristics are accepted as important tools for increased quality assurance, but the question remains whether these should be mandatory or not. This study investigated four characteristics in laboratories testing for biomarkers in NSCLC and mCRC: accreditation status of a laboratory, their setting, the number of samples tested each year and the change in used methods.

Less analysis errors were observed in accredited laboratories in the early stages of biomarker introduction. This was demonstrated by the positive effect for NRAS variant analysis, which is a new biomarker introduced in laboratories in 2013, and EGFR variant analysis. These two biomarkers (EGFR and NRAS) are essential for therapy decision-making and will challenge the laboratories because of the need of fast introduction in the routine or the large spectrum of variants that needs to be tested. 

Implementation of new tests in the accredited laboratories should follow strict standards, including the introduction of well-planned and validated standard operating procedures. This can be seen as a drawback of accreditation, as it could lead to postponing the introduction of new developments. However, in the context of biomarkers, laboratories are forced to introduce these changes, as they become mandatory before therapy decision. In that case, accreditation forms a welcomed basis during introduction.

Once the biomarker is successfully implemented in the laboratory, the association between good performance and laboratory accreditation disappears. While accredited laboratories have addressed the difficulties during the extensive validation procedure, laboratories without accreditation have gained experience during the routine execution of biomarker testing and
perform as well as accredited laboratories in the long term. The positive association continues for the \textit{EGFR} biomarker. The LUNG \textit{EGFR} EQA scheme remains a challenging scheme in the long term because of the use of different types of sample material (cell line material versus resection specimens), which was not always included by the laboratories in their initial validation and the large spectrum of variants.

Nevertheless, in the long term, the effect of accreditation seems to gain more influence again, as there is a trend towards less analysis and technical errors in accredited laboratories for the routinely used \textit{KRAS} biomarker. A laboratory with accreditation must remain attentive and should be able to guarantee a high quality assurance, even after many years of routine analysis.

The hypothesis that accreditation status of a laboratory has a positive impact on the correctness of results also has a role in reporting conclusive results for gene rearrangement and expression analyses. Accredited laboratories were more careful in reporting conclusive results for \textit{ROS1} FISH analysis, but no association could be found with \textit{ALK} FISH. At the time of the analyses, in 2012, \textit{ROS1} was a recently discovered biomarker, not yet included in any European label for anti-EGFR tyrosine kinase inhibitors (TKI).\textsuperscript{33,34} This might be an explanation for the...
Laboratories that work within a setting of university and research made less analysis errors in variant analysis testing than laboratories in another setting. It appears that university and research laboratories have more extensive experience and are used to innovative methodology, while private and industry laboratories are more stimulated to work in a cost-effective way. It is expected that this significant result will disappear within a few years, when new or complex biomarkers are fully integrated into the routine work of the laboratories, like what was observed with the accreditation indicator. There were no signs that these two indicators were linked to each other, as the university and research laboratories have an almost equal distribution of accreditation or no accreditation.

Table 4. Statistical results for the association between the EQA results (EQA score, analysis errors) and the number of samples tested per year

<table>
<thead>
<tr>
<th></th>
<th>EQA score</th>
<th>Analysis errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio</td>
<td>p-value</td>
</tr>
<tr>
<td>Colon EQA scheme 2010–2016 (all laboratories), n = 688</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRAS categorical (global test)</td>
<td>0.034</td>
<td>0.002</td>
</tr>
<tr>
<td>&lt;10 versus more than 10 samples</td>
<td>0.42</td>
<td>0.067</td>
</tr>
<tr>
<td>10–99 samples versus more than 99</td>
<td>0.41</td>
<td>0.003</td>
</tr>
<tr>
<td>100–249 samples versus more than 249</td>
<td>0.69</td>
<td>0.256</td>
</tr>
<tr>
<td>KRAS ordinal: +1 level</td>
<td>1.47</td>
<td>0.002</td>
</tr>
<tr>
<td>Colon EQA scheme 2013–2016 (RAS laboratories), n = 333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRAS categorical (global test)</td>
<td>0.013</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>&lt;10 versus more than 10 samples</td>
<td>0.44</td>
<td>0.170</td>
</tr>
<tr>
<td>10–99 samples versus more than 99</td>
<td>0.22</td>
<td>0.004</td>
</tr>
<tr>
<td>100–249 samples versus more than 249</td>
<td>0.51</td>
<td>0.419</td>
</tr>
<tr>
<td>NRAS ordinal: +1 level</td>
<td>1.58</td>
<td>0.009</td>
</tr>
<tr>
<td>Lung EGFR EQA scheme 2012–2016, n = 457</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR categorical (global test)</td>
<td>0.015</td>
<td>0.041</td>
</tr>
<tr>
<td>&lt;10 versus more than 10 samples</td>
<td>0.43</td>
<td>0.044</td>
</tr>
<tr>
<td>10–99 samples versus more than 99</td>
<td>0.74</td>
<td>0.252</td>
</tr>
<tr>
<td>100–249 samples versus more than 249</td>
<td>0.52</td>
<td>0.010</td>
</tr>
<tr>
<td>EGFR ordinal: +1 level</td>
<td>1.30</td>
<td>0.009</td>
</tr>
<tr>
<td>Lung ALK EQA scheme 2012–2016, n = 384 (FISH) and n = 305 (IHC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALK FISH categorical (global test)</td>
<td>&lt;0.001</td>
<td>0.018</td>
</tr>
<tr>
<td>&lt;10 versus more than 10 samples</td>
<td>0.19</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>10–99 samples versus more than 99</td>
<td>0.40</td>
<td>0.013</td>
</tr>
<tr>
<td>100–249 samples versus more than 249</td>
<td>0.61</td>
<td>0.277</td>
</tr>
<tr>
<td>ALK FISH ordinal: +1 level</td>
<td>1.79</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ALK IHC categorical (global test)</td>
<td>0.022</td>
<td>0.003</td>
</tr>
<tr>
<td>&lt;10 versus more than 10 samples</td>
<td>0.39</td>
<td>0.032</td>
</tr>
<tr>
<td>10–99 samples versus more than 99</td>
<td>0.83</td>
<td>0.647</td>
</tr>
<tr>
<td>100–249 samples versus more than 249</td>
<td>1.91</td>
<td>0.184</td>
</tr>
<tr>
<td>ALK IHC ordinal: +1 level</td>
<td>1.19</td>
<td>0.269</td>
</tr>
<tr>
<td>Lung ROS1 EQA scheme 2014–2015, n = 122 (FISH) and n = 62 (IHC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROS1 FISH categorical (global test)</td>
<td>0.865</td>
<td>0.616</td>
</tr>
<tr>
<td>ROS1 FISH ordinal: +1 level</td>
<td>1.01</td>
<td>0.937</td>
</tr>
<tr>
<td>ROS1 IHC categorical (global test)</td>
<td>0.038</td>
<td>0.234</td>
</tr>
<tr>
<td>&lt;10 versus more than 10 samples</td>
<td>0.47</td>
<td>0.449</td>
</tr>
<tr>
<td>10–99 samples versus more than 99</td>
<td>10.47</td>
<td>0.006</td>
</tr>
<tr>
<td>ROS1 IHC ordinal: +1 level</td>
<td>1.44</td>
<td>0.306</td>
</tr>
</tbody>
</table>

For categorical variables, the global test shows the difference between the categories: <10, 10–99, 100–249, 250–499, 500–999, ≥1000. Post hoc comparisons are only performed when the global p-value is significant. RAS testing laboratories included those laboratories that tested KRAS and NRAS genes. EGFR external quality assessment, IRR incidence rate ratio, NA not applicable.

contradiction. Once the ROS1 biomarker becomes a well-known relevant biomarker, the data should be re-analysed to observe whether the significant effect disappears after few years, similar as with the variant analysis biomarkers. Accredited laboratories reported fewer technical errors for ALK IHC, while there was no association between the percentage of technical failures and accreditation for ALK FISH rearrangement analysis. This test is for most laboratories technically more demanding and difficult to interpret than IHC analysis. Laboratories, independent of their accreditation status, may introduce this test only after profound training of their personnel, such that they are able to give valid and conclusive interpretations. At the time of the analysis, this analysis could not be repeated for ROS1 IHC, as no data were available.
The number of samples analysed per year for RAS and EGFR can be considered as an influential indicator for good laboratory performance. There is no evidence that this indicator was linked to the previously discussed indicators of accreditation and setting. RAS-testing of more than 99 samples, compared to 10–99 samples gave a higher probability of a better EQA score and less analysis errors. Using these results advocates that centralised biomarker testing improves quality. Development of a centralised approach to testing of predictive markers should be promoted soon. Such a centralised approach will increase the experience of the laboratory and will be beneficial for the performance of laboratories performing variant analyses of oncology biomarkers.

There is evidence of an improved EQA FISH performance for a larger number of samples tested per year for ALK, which is absent for analysis of ROS1 rearrangements (Table 4). ROS1 was a new biomarker that was introduced in the laboratory before it became a mandatory biomarker. In this case, the number of samples tested per year seems to have no influence on the EQA results of FISH analysis. Once the biomarker becomes clinically relevant and routinely used, enough samples should be tested per year to

![Longitudinal overview of the distribution of techniques in different EQA schemes for (a) KRAS variant analysis, (b) NRAS variant analysis and (c) EGFR variant analysis. NGS next-generation sequencing, Commercial includes non-NGS-based commercial kits, laboratory includes developed non-NGS-based laboratory developed](image-url)
maintain the high level of experience as every increase in the level of number of samples tested implies a decrease in analysis errors of 32% (Table 4). It should be taken into account that less data were available for ROS1 analysis compared to ALK analysis so this conclusion should be confirmed in the future.

FISH analysis, including the interpretation, is technically more difficult compared to a more straightforward IHC analysis. This can explain why no linear association was found between ALK IHC analysis and EQA results. Regarding ROS1 IHC analysis, there was even an increase in false positive and false negative results when more than 99 ROS1 samples were analysed with IHC per year compared to 10–99 samples. Also, IHC analysis mainly depends on antibody sensitivity while FISH analysis requires interpretation by an experienced operator. It was expected that more experience with FISH samples is thus needed during the introduction of a new biomarker. However, this cannot be demonstrated in this study.

The results show an increased use of NGS-based methods and a decreased use of non-NGS-based LDTs for variant analysis (Fig. 2). Considering the broader spectrum of relevant biomarkers, this is a logical evolution in molecular pathology testing. At the time of introduction of the NRAS biomarker, the availability of non-NGS-based commercial test kits for NRAS variant analysis was limited.\(^{14}\) There was an extensive use of non-NGS-based LDTs for NRAS analysis in the Colon EQA scheme of 2013, which was significantly higher than for KRAS variant analysis (\(p = 0.047\)). Over the following years, the ratio changed as more non-NGS-based commercial NRAS test kits became available. The change in methods does not seem to influence the EQA results. Consequently, this characteristic cannot be an indicator of good quality assurance of biomarker testing.

**CONCLUSION**

The hypothesis to impose accreditation of laboratories, to centralise biomarker testing or to test within a setting of university and research was supported by this study. Results of EQA schemes in mCRC and NSCLC show that improvement of performance in oncology biomarker testing is associated with three of the four tested characteristics. First, laboratory accreditation is needed to ensure high quality and reliable implementation of new diagnostic molecular biomarkers. Second, university and research laboratories reach more optimal results. Third, larger number of samples tested per year proved to be an indicator for good EQA results, implying more centralisation of biomarker testing to reach sufficiently high volumes. Finally, changes in applied methods do not influence laboratory performance.

**ACKNOWLEDGEMENTS**

We want to thank the participating laboratories for analysis of the samples. We thank the assessors for carefully evaluating the EQA results. We thank the European Society of Pathology for their assistance in administration, especially Dr. R. Al Dierie (ESP-QA Foundation, Brussels, Belgium). We thank Dr. Annouschka Laeren from the Leuven Biostatistics and Statistical Bioinformatics Centre for carrying out the statistical analysis. We thank Prof. Dr. Marjolijn Litgenberg from Radboud University Medical Center (Nijmegen, The Netherlands) for her interest in the study and the valuable suggestions that helped us to further optimise this paper.

**AUTHOR CONTRIBUTIONS**


**ADDITIONAL INFORMATION**

**Competing interests:** Ed Schuuring, performed lectures for Illumina, Novartis, Pfizer and BioCartis; is a consultant in the advisory boards for AstraZeneca, Pfizer, Novartis and BioCartis, and received financial support from Roche, Biocartis, BMS and Pfizer (all fees to the Institution). The remaining authors declare no competing interests.

**Consent for publication:** All the authors will observe the terms of the Licence to Publish that they have signed or has been signed on their behalf and will seek the consent of the publishers acting for the owners of the journal, in any other circumstance.

**Ethics approval and consent to participate:** Ethical compliance is not applicable for this study. Informed consent is not a mandatory prerequisite for the use of patient-derived material, since samples for test validation are exempted from research regulations requiring informed consent.

**Note:** This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution 4.0 International (CC BY 4.0).

**REFERENCES**


