ORIGINAL ARTICLE

External quality assurance of HER2 fluorescence in situ hybridisation testing: results of a UK NEQAS pilot scheme

John M S Bartlett, Merdol Ibrahim, Bharat Jasani, John M Morgan, Ian Ellis, Elaine Kay, Hilary Magee, Sarah Barnett, Keith Miller

J Clin Pathol 2007;**60**:816–819. doi: 10.1136/jcp.2006.040840

See end of article for authors' affiliations

Correspondence to:
Dr J M S Bartlett, Endocrine
Cancer Group, Edinburgh
Cancer Research Centre,
Western General Hospital,
Crewe Road South,
Edinburgh EH4 2XR, UK;
John.Bartlett@ed.ac.uk

Accepted 10 August 2006
Published Online First
8 September 2006

Background and Aims: Trastuzumab provides clinical benefit for advanced and early breast cancer patients whose tumours over-express or have gene amplification of the HER2 oncogene. The UK National External Quality Assessment Scheme (NEQAS) for immunohistochemical testing was established to assess and improve the quality of HER2 immunohistochemical testing. However, until recently, no provision was available for HER2 fluorescence in situ hybridisation (FISH) testing. A pilot scheme was set up to review the performance of FISH testing in clinical diagnostic laboratories.

Methods: FISH was performed in 6 reference and 31 participating laboratories using a cell line panel with known HER2 status.

Results: Using results from reference laboratories as a criterion for acceptable performance, 60% of all results returned by participants were appropriate and 78% either appropriate or acceptable. However, 22.4% of results returned were deemed inappropriate, including 13 cases (4.2%) where a misdiagnosis would have been made had these been clinical specimens.

Conclusions: The results of three consecutive runs show that both reference laboratories and a proportion of routine clinical diagnostic (about 25%) centres can consistently achieve acceptable quality control of HER2 testing. Data from a significant proportion of participating laboratories show that further steps are required, including those taken via review of performance under schemes such as NEQAS, to improve quality of HER2 testing by FISH in the "real world".

At the ASCO 2005 annual meeting, data from a number of multinational trials in early breast cancer showed dramatic survival advantages for patients receiving trastuzumab (Herceptin), a humanised murine monoclonal antibody directed against the HER2 growth factor receptor, over those receiving polychemotherapy alone, providing strong evidence for extending trastuzumab treatment for early HER2 positive breast cancers.^{1 2}

Extensive data support the use of fluorescence in situ hybridisation (FISH) in testing for HER2 gene amplification as a valid predictor of HER2 expression and Herceptin response.^{3–8} Concerns have been raised about the quality of HER2 FISH testing delivered outside specialist centres. Given that the number of cases tested will increase up to four-fold as testing is applied to early invasive breast cancers, there is a clear need for external assessment of quality of HER2 testing using FISH technology.

The UK NEQAS initiated a multinational pilot scheme to assess the technical quality of laboratories using HER2 FISH testing. Data from six reference laboratories were used to define acceptable criteria for HER2 FISH testing based on a panel of four samples tested in duplicate (www.ukneqasicc.ucl.ac.uk). Results from 31 active participants from 12 countries (United Kingdom, Austria, Denmark, Eire, France, Norway, Portugal, Slovenia, Switzerland, The Netherlands, Hong Kong, P.R. China) were collected and validated against these results on three separate occasions over a period of 12 months. These results provide the first multinational assessment of reproducibility of FISH based analysis over an extended period and provide valuable insight into factors influencing the quality of HER2 FISH testing.

MATERIALS AND METHODS

Unstained sections from formalin-fixed paraffin processed blocks of human breast carcinoma cell lines (MDA-MD-231,

MDA-MB-453, BT-20/MDA-MB-175, and SKBR3) were circulated to participants within the scheme on three separate occasions over a 12-month period. Previous FISH analysis on these cell lines showed the SKBR3 and MDA-MB-453 cell lines to have HER-2/*neu* gene amplification, while the cell lines BT-20, MDA-MB-175 and MDA-MB-231 lines were negative². Sections (4 µm) cut from paraffin-embedded cell pellets provided by Novacastra (Newcastle upon Tyne, UK) were assayed by different methodology in the reference laboratories (PathVysion (Abbott Molecular, Des Plaines, Illinois, USA), the Pharm Dx (Dako, Cambridge, UK) or the Ventana Inform FISH assay (Ventanc Medical System X.S.A. Illkired, France)). Cell lines were selected to span the critical diagnostic region of low-level amplification (2–3 copies/chromosome). The reference laboratories returned results for each cell line relating to both HER2:chromosome 17 ratio and HER2 copy number (in order to assess laboratories using HER2 copy number only). Results from reference laboratories and participants were then assessed for each cell line and over time.

In total, 78 sets of data were returned from participants, representing 312 HER2 FISH results. In addition 29 complete sets of results were returned from the reference centres (116 FISH results). Data presented here are therefore from a total of 428 analyses of these 5 cell lines.

RESULTS

Table 1 summarises the results from reference laboratories (five using HER2:chromosome 17 ratio and one using HER2 copy number) for each cell line over three separate runs. The mean and range of results obtained from reference laboratories over

Abbreviations: FISH, fluorescence in situ hybridisation; NEQAS, National External Quality Assessment Scheme

Table 1 Reference laboratory results from the UK National External Quality Assessment Scheme (NEQAS) pilot HER2 fluorescence in situ hybridisation (FISH) scheme

	HER2:chromosome 17 ratio			HER2 copy number		
	Average	Range	Borderline	Average	Range	Borderline
MDA-MB-231						
Run 1	1.09 (0.05)	0.87–1.43	0.78–1.57	2.41 (0.12)	1.55–2.85	1.40–3.14
Run 2	1.16 (0.08)	0.88–1.44	0.79–1.58	2.25 (0.25)	1.53–2.80	1.38–3.08
Run 3	1.04 (0.05)	0.71–1.19	0.64–1.31	2.62 (0.22)	1.45–3.65	1.31–4.02
BT20/MDA-MB-175*						
Run 1	0.93 (0.04)	0.74–1.15	0.67–1.27	3.05 (0.14)	2.50–4.15	2.25–4.57
Run 2	1.19 (0.08)	0.92–1.34	0.83–1.47	2.97 (0.20)	2.60–3.55	2.34–3.91
Run 3	1.21 (0.02)	1.13–1.29	1.02–1.42	3.32 (0.26)	1.70–4.17	1.53–4.59
MDA-MB-453						
Run 1	2.37 (0.06)	2.12–2.72	1.91–2.99	6.72 (0.24)	5.70–8.20	5.13–9.02
Run 2	2.30 (0.13)	2.08–2.85	1.87–3.14	6.21 (0.18)	5.78–6.90	5.20–7.59
Run 3	2.39 (0.09)	2.10–2.90	1.89–3.19	7.45 (0.42)	5.00–10.0	4.50–11.0
SK-BR-3						
Run 1	3.52 (0.14)	2.58–4.09	2.32–4.50	14.57 (0.34)	13.3–17.1	12.0–18.8
Run 2	3.53 (0.20)	3.01–4.06	2.71–4.47	13.53 (0.80)	12.1–16.1	10.9–17.7
Run 3	3.88 (0.12)	3.37–4.41	3.03–4.85	16.23 (0.78)	11.7–20.7	10.5–22.8

Mean (SEM) of results from reference laboratories relating to three separate quality assurance "runs" during 2004/05. Results are expressed either as copies of HER2 gene per chromosome 17 (HER2/chromosome 17 ratio) or mean copies of HER2/cell (HER2 copy number). Five separate cell lines were used to represent either non-amplified or amplified breast carcinomas.

*BT20 was used in run 1 and MDA-MB-175 was used for runs 2 and 3.

29 assays for each cell line shows that on no occasion did any of the reference laboratories misdiagnose the HER2 status of any of these cell lines (table 1). Results were maintained within a very narrow band for each cell line over the three assessments during the pilot scheme. This supports previous evidence^{9–11}

that reference laboratories can deliver high quality diagnostic testing. However, we show, for the first time, that such performance is also sustained over a prolonged period and can be used to provide standards against which other centres can be assessed.

Table 2 Summary of results from participating laboratories for the UK National External Quality Assessment Scheme (NEQAS) pilot HER2 fluorescence in situ hybridisation (FISH) scheme

	Results	Appropriate	Acceptable	Inappropriate	Misdiagnosis
SK-BR-3					
Run 1	22 (9)	10	8	4	0
Run 2	27 (5)	9	10	8	0
Run 3	29 (2)	16	3	10 (1NR)	1
Total	78 (16)	35	21	20	1
MDA-MB-453					
Run 1	22 (9)	13	3	6	2* (<1.8–2.0)
Run 2	27 (5)	17	3	7	6
Run 3	29 (2)	15	6	8	3
Total	78 (16)	45	12	21	11
BT20/MDA-MB-175*					
Run 1	22 (9)	16	2	4	1
Run 2	27 (5)	14	5	8	0
Run 3	29 (2)	10	10	9 (3 NR)	0
Total	78 (16)	40	17	21	1
MDA-MB-231					
Run 1	22 (9)	20	2	0	0
Run 2	27 (5)	25	1	1	0
Run 3	29 (2)	20	4	5 (2 NR)	0
Total	78 (16)	65	7	6	0

Distribution of results from participating laboratories over three runs during 2004/05 as assessed by results from reference laboratories.

Results = number of participating laboratories submitting results. Of participating laboratories, 22 participated in run 1, 25 in run 2 and 29 in run 3; figures in brackets represent those participating laboratories which did not submit results for that run (e.g. in run 1, 22 laboratories submitted results and 9 did not).

Results were scored as either "appropriate" (falling within the range identified by the reference laboratories), "acceptable" (falling within 10% of the range identified by the reference laboratories), "inappropriate" (falling outside the range $\pm 10\%$ identified by reference laboratories but not resulting in a misdiagnosis), or "misdiagnosis" (falling outside the above range and resulting in a possible "misdiagnosis" of the case in question).

*BT20 was used in run 1 and MDA-MB-175 was used for runs 2 and 3.

Results from reference laboratories were used to define an acceptable range (based on both copy number and gene:chromosome ratio) as follows: samples from participants falling within the upper and lower limits of results obtained from reference laboratories were regarded as “appropriate” (see table 1). Samples within 10% of the lower or upper limits of results obtained from reference laboratories were defined as “acceptable”; this is based on previous evidence that inter-observer error for FISH scoring is around 10%.^{5 12–14} Results from participating laboratories outside these limits were regarded as “inappropriate”. Notwithstanding these definitions, any sample which resulted in a “misdiagnosis” (e.g. a non-amplified sample scored as amplified) was identified as such (table 2).

Using these criteria, 60% of all results returned by participants were appropriate and 78% either appropriate or acceptable (see table 2). However, 22.4% of results returned were deemed inappropriate, including 13 cases (4.2%) where a misdiagnosis would have been made had these been clinical specimens. These data suggest that, in approximately 1 in 5 cases analysed within this external quality assessment scheme, laboratories returned results markedly outside optimal performance criteria. In addition the majority of misdiagnosed cases were related to the MDA-MB-453 cell line, which exhibits low-level HER2 amplification which was falsely diagnosed as “non-amplified” in 14% of cases. This result reflects only those cases in which a ratio below 1.8 copies/chromosome was reported, since in many diagnostic laboratories HER2 ratios between 1.8 and 2.2 are reported as “borderline amplified” or “borderline non-amplified”. For this cell line a “borderline non-amplified” result was not scored as a “misdiagnosis” since this clearly falls within the technical limits of the variation within the assay for this sample. The amplification ratio for the SK-BR3 line was over-reported in over 25% of cases and in one case this cell line was reported as non-amplified. There appears to be equal variation in reporting non-amplified cases (table 2). Although only one laboratory reported amplification of the BT20 cell line (gene:chromosome ratio of 5.95), 27% of the results for both this cell line and the MDA-MB-175 cell line and 8% of results for the MDA-MB-231 cell line were outside the technical limits set for the external quality assessment scheme.

Fourteen non-reference laboratories completed all three assessments within this pilot external quality assurance scheme; a further 9 joined part of the way through the scheme and 7 laboratories did not complete all the assessments. Of the non-laboratories who completed three assessments; while none performed to as high a standard at the reference laboratories, 5 consistently reported all cell lines within adequate ranges (reference range $\pm 10\%$), 6 achieved this standard on 2/3 attempts, 2 achieved this standard on 1/3 attempts and 1 did not attain this level of reporting, for all cell lines, on any of the three runs (data not shown).

DISCUSSION

The assessment of tumour expression or amplification of the HER2 oncogene is increasingly requested for patients diagnosed with breast cancer who are considered for either adjuvant therapy or therapy for advanced disease (NCCN guidelines). It is of central importance that these diagnoses provide an accurate and reliable indication for both clinicians and patients of the true HER2 status of breast carcinomas. Recent reports, relating to both immunohistochemical and molecular (FISH) based tests have suggested that in the “real world” performance of such tests may be sub-optimal. However, to date such evidence has been related only to laboratories testing for clinical trials assessed at a single time point. We have now collated evidence from six reference laboratories and over 30

other laboratories, which perform diagnostic HER2 testing for local practices participating within an international scheme for external quality assurance. Most importantly, as data were collected over a prolonged time period, a dynamic assessment of the quality of HER2 FISH testing over time was possible.

These data support previous evidence suggesting that so called “reference laboratories”, those with particular expertise and experience, can provide consistently high quality analysis of HER2 status over time. The six laboratories which were selected to act as “reference laboratories” for the UK NEQAS FISH pilot scheme provided consistently high quality results over time, irrespective of the testing modality used (whether assessing HER2 copy number only or combined HER2:chromosome 17). This does not imply that one or other method is superior, but rather that both can be adequately quality controlled in a routine diagnostic environment.

Results from laboratories participating in the scheme, as previously reported, were more variable than those achieved by the reference laboratories. However, performance among participants, while variable, was not uniformly poor. A significant number of laboratories consistently achieved acceptable results (about 35% of those completing all assessments) while just over half showed variability between acceptable and sub-optimal performance and about 10% showed consistently poor performance. It should be noted that during this period of assessment, laboratories were effectively blinded to the results of the assessment, and therefore these results reflect the “real-world” performance of these tests. Conversely, these results leave open the opportunity to improve performance by regular review of standards; this is currently underway in the second year of the pilot scheme.

We conclude from these results that while quality of diagnostic testing, in particular relating to FISH testing for HER2 status in breast cancer, remains highly variable between centres, high quality testing is not solely confined to “reference” laboratories. A significant proportion of routine diagnostic centres achieve consistently acceptable results. However, it still appears that, for the majority of centres, further attention to provide consistently high quality performance is required. We believe that by participating in external quality assurance schemes such as the UK NEQAS scheme, improvements in diagnostic quality will be achieved for HER2 FISH testing in a similar manner as that previously shown for ER immunohistochemical testing.^{15 16}

Take-home messages

- Quality assurance is an essential component of delivery of consistent and accurate diagnostic results and should be performed both internally and externally.
- FISH, as with any diagnostic test, can and should be rigorously evaluated before use in the diagnostic setting.
- Many laboratories show acceptable quality control of FISH testing for HER2; others need to assess procedures and improve.
- Every laboratory performing diagnostic FISH based assays should participate in an appropriate external quality assurance scheme.
- The NEQAS FISH scheme provides an accredited and robust means of external quality assurance for laboratories performing FISH based diagnosis.

Authors' affiliations

J M S Bartlett, Endocrine Cancer Group, Edinburgh Cancer Research Centre, Western General Hospital, Edinburgh, UK

M Ibrahim, K Miller, UK National External Quality Assessment Scheme, University College London, London, UK

B Jasani, J M Morgan, Department of Pathology, School of Medicine, Cardiff University, Cardiff, UK

I Ellis, University of Nottingham, Molecular Medical Sciences, Department of Histopathology, Nottingham City Hospital, Nottingham, UK

E Kay, Department of Histopathology, Beaumont Hospital and The Royal College of Surgeons in Ireland, Beaumont Hospital, Dublin, Ireland

H Magee, Adelaide and Meath Hospital, incorporating the National Children's Hospital, Dublin, Ireland

S Barnett, University College London, Advanced Diagnostics, London, UK

Competing interests: None declared.

REFERENCES

- Vogel C**, Cobleigh MA, Tripathy D, *et al*. First-line, single-agent Herceptin (trastuzumab) in metastatic breast cancer: a preliminary report. *Eur J Cancer* 2001;**37**(Suppl 1):S25–9.
- Slamon DJ**, Leyland-Jones B, Shak S, *et al*. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;**344**:783–92.
- Ellis IO**, Bartlett J, Dowsett M, *et al*. Updated recommendations for HER2 testing in the UK. *J Clin Pathol* 2004;**57**:233–7.
- Bartlett J**, Mallon E, Cooke T. The clinical evaluation of HER-2 status: which test to use? *J Pathol* 2003;**199**:411–17.
- Bartlett JMS**, Going JJ, Mallon EA, *et al*. Evaluating HER2 amplification and overexpression in breast cancer. *J Pathol* 2001;**195**:422–8.
- Mass R**. The role of HER-2 expression in predicting response to therapy in breast cancer. *Semin Oncol* 2000;**27**(Suppl 11):46–52.
- Press MF**, Bernstein L, Thomas PA, *et al*. HER-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas. *J Clin Oncol* 1997;**15**:2894–904.
- Press MF**, Hung G, Godolphin W, *et al*. Sensitivity of HER-2/neu antibodies in archival tissue samples: potential source of error in immunohistochemical studies of oncogene expression. *Cancer Res* 1994;**54**:2771–7.
- Dowsett M**, Bartlett J, Ellis IO, *et al*. Correlation between immunohistochemistry (HerceptTest) and fluorescence in situ hybridization (FISH) for HER-2 in 426 breast carcinomas from 37 centres. *J Pathol* 2003;**199**:418–23.
- Paik S**, Bryant J, Tan-Chiu E, *et al*. Real-world performance of HER2 testing—national surgical adjuvant breast and bowel project experience. *J Natl Cancer Inst* 2002;**94**:852–4.
- Paik S**, Tan-Chiu E, Bryant J, *et al*. Successful quality assurance program for HER2 testing in the NSABP trial for Herceptin. *Breast Cancer Res Treat* 2002;**76**:9.
- Bartlett JMS**, Watters AD, Ballantyne SA, *et al*. Estimation of chromosome 9 copy number using quantitative fluorescence in situ hybridisation fish as a marker of disease recurrence in transitional cell carcinoma of the bladder. *J Pathol* 1997;**182**:A20.
- Watters AD**, Ballantyne SA, Going JJ, *et al*. Aneusomy of chromosomes 7 and 17 predicts the recurrence of transitional cell carcinoma of the urinary bladder. *BJU Int* 2000;**85**:42–7.
- Watters AD**, Going JJ, Cooke TG, *et al*. Chromosome 17 aneusomy is associated with poor prognostic factors in invasive breast carcinoma. *Breast Cancer Res Treat* 2003;**77**:109–14.
- Miller K**, Rhodes A, Jasani B. Variation in rates of oestrogen receptor positivity in breast cancer again. *BMJ* 2002;**324**:298–9.
- Rhodes A**, Jasani B, Barnes DM, *et al*. Reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: interlaboratory variance in the sensitivity of detection and evaluation of scoring systems. *J Clin Pathol* 2000;**53**:125–30.

Save your favourite articles and useful searches

Use the “My folders” feature to save and organise articles you want to return to quickly—saving space on your hard drive. You can also save searches, which will save you time. You will only need to register once for this service, which can be used for this journal or all BMJ Journals, including the BMJ.