

EXPERT OPINION

Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment I

R A Walker

Department of Cancer Studies and Molecular Medicine, Leicester Royal Infirmary, Leicester, UK

Walker R A

(2006) *Histochemistry* **49**, 406–410

Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment I

Immunohistochemistry is no longer a technique used only for research but is employed increasingly for diagnosis and for the assessment of therapeutic biomarkers. The latter, in particular, often require a semiquantitative evaluation of the extent of their presence. There are many factors that can affect this that relate to the method: fixation of tissue, duration and type of antigen retrieval, antibody specificity, antibody dilution and detection systems. Other complexities relate to assessment. Different scoring systems are used for either the same or different antigens. Cut-off levels for assessing whether a tissue is 'positive' or 'negative' can vary for

the same antigen. Whilst there are quality assurance schemes for the methodology that have improved standards of staining, there are no similar schemes that relate to interpretation, although errors here can create as many problems. There have been improvements in automated analysis but availability is limited and it is still predominantly a research tool. In order for quantification of immunohistochemistry to be a reliable and reputable tool, there must be easy to use, reproducible, standardized protocols for assessment which are international. Improvements in automated analysis with wider applicability could lead to standardization.

Keywords: automated image analysis, HER-2, immunohistochemistry, oestrogen receptor, quantification

Introduction

Immunohistochemistry has been used increasingly over the last 30–40 years, initially for research but now for diagnosis and, highly relevant to this article, for assessing therapeutic biomarkers. With this has been the change from 'is protein X present?' to 'how much of protein X is present?'. The number of laboratories undertaking the technique has markedly increased, along with the number of people undertaking the interpretations, but this has not solved the problems that there can be with a method which can vary for each antigen and its method of assessment.

Address for correspondence: Rosemary A Walker, Department of Cancer Studies and Molecular Medicine, Robert Kilpatrick Building, Leicester Royal Infirmary, Leicester, LE2 7LX, UK.
e-mail: raw14@le.ac.uk

Methodology

The first critical component relates to the tissue and type and duration of fixation. This is particularly important for some antigens, e.g. oestrogen receptor (ER), and has led to the recommendation¹ that breast specimens are received as soon as possible after surgery and sliced to allow rapid and even fixation. The rapid fixation achieved with core biopsies is of benefit. If all that is available for analysis is suboptimally fixed tissue then there should be caution about interpretation and any semiquantitative assessment. Although proteins appear to be stable in wax blocks, there is evidence of deterioration once paraffin sections have been cut.^{2,3} In a routine laboratory there should be no problem in using sections that are cut and stained promptly, but it will create problems for those clinical trials that have requested sections that are then stored for subsequent

analysis months to years later. Will any data obtained be meaningful?

The next stage in the methodology is antigen retrieval. Although many protocols relate to heating in an acid buffer (either by pressure cooking, microwaving or water bath), there are some antigens where enzymatic treatment is preferable.⁴ The duration of antigen retrieval needs to be optimized by laboratories in relation to different antigens and fixation schedules. Interlaboratory comparisons have identified insufficient antigen retrieval as a major contributory factor to variations in extent of staining of test sections,^{5–7} with microwave antigen retrieval causing particular problems.⁵ This can result in low-level expressing tumours being assessed as being negative, which for antigens such as ER can affect therapeutic decisions and have an impact on patients.

A recent study has demonstrated how variation of antibody dilutions can affect results.⁸ When a high antibody concentration (1 : 500) was used to detect HER-2, low expression was associated with poorer survival (contrary to published data), but when a low antibody concentration (1 : 8000) was employed, high expression was associated with increased mortality. This relationship was detected because an automated quantitative analysis system was used. The effect of antibody dilution when assessing the proportion of positive cells has clearly been demonstrated for proliferation markers⁹ and p53,¹⁰ emphasizing the problems that relate to the method and the difficulties that can arise in comparing findings from different studies unless there is complete standardization in methodology.

Different antibodies directed against the same antigen may give different results. This is due to differences in the epitopes that they detect. Whilst in research it could be of value to use several different antibodies against the same protein, in routine practice, when information is being provided that could influence or determine treatment, it is necessary to know the sensitivity and specificity of antibodies. This is possible when immunohistochemical results can be related to a 'gold standard' such as comparing HER-2 staining with gene amplification status. An early study by Press *et al.*¹¹ demonstrated wide variations in both specificity and sensitivity between 38 HER-2 antibodies, so explaining differences in findings. The affinities of different antibodies directed against p53 can influence the extent of detection of the protein,^{12,13} making comparisons between publications and potential clinical utility difficult.¹⁴

The other factors in methodology that can result in variations between different laboratories include detection systems. Automated immunostaining should

overcome some of the variability that can be due to technical factors, but cannot overcome many of the above factors—and this is before methods of assessment, scoring systems and cut-off levels are considered.

Assessment

Immunohistochemical reactivity can be within cells—membrane, cytoplasmic and/or nuclear—or in stroma. Assessment of reactivity can range from being either present or absent; in relation to the extent of staining; in relation to intensity of staining; a combination of extent and intensity. If intensity is being assessed then there has to be a positive control having different levels of staining that is included with each staining batch and intensity in test cases is compared with this. Having this standard should overcome the problem of batch variation and allow comparisons. An arbitrary 'weak, moderate and strong', depending on how the assessor feels, is not appropriate.

Assessing the extent of reactivity can also range from careful counting to 'the quick eye-ball'. It can involve counting up to 1000 cells for reactivity to give a percentage or counting number of positive cells in 10 high-power fields (HPF). Less precise is assessment of proportion of cells staining throughout the tissue section, with subdivisions into percentage ranges. If tissue microarrays are being used with a smaller area per sample, a more formal count would seem appropriate. There will be variations between antigens; counting of nuclear reactivity is easier than cytoplasmic. Deriving a percentage as a method of assessment of factors relating to cell cycle, e.g. MIB-1, p53, p21, is appropriate to allow comparisons between tissues of the same type and of different types (provided the methodological factors considered above are standardized). A problem with dividing assessments into percentage ranges is that different research groups can have different ranges for the same antigen. These may relate to the median and standard deviations, or quarters or thirds. Comparison of results between different studies then becomes difficult.

Scoring systems have been introduced in order to overcome variations, particularly for markers that are used for making therapeutic selections. The H-score¹⁵ was introduced for assessment of ER but has been used for a range of other markers.^{16,17} It takes into account extent of reactivity and intensity and is the sum of the percentage of cells with weak, moderate and strong staining with a range from 0 to 300. It can be time consuming to do if done accurately, and so other scoring systems have been introduced. The quick score¹⁸ was based on percentage range of cells staining

from 1 to 4 and overall intensity as 1–3, which were added, but for ER has generally been replaced by the Allred score,¹⁹ which has expanded the lower end of the scale giving a range of 1–5. This was derived by assessing a group of breast cancers that had previously been analysed for ER by biochemical assays. One criticism of these scoring systems is that extent and intensity scores are added and that multiplying them would more relevant.²⁰ In practice, all scoring systems have been shown to give statistically significant correlations with clinical outcomes when used by those with experience in this area.^{21,22} Although there has been much effort to improve the quality of immunohistochemical staining by the use of external quality assurance schemes such as the National External Quality Assessment Service, there is no equivalent check on the ability of those undertaking assessment. One study in Germany of 172 pathologists highlighted the problems that can occur with interpretation, with 24% of ER staining having a false-negative assessment.²³

A completely different system has been introduced for another marker used for therapeutic selection, HER-2. This is the HercepTest United States Food and Drug Administration-approved grading system of 0, 1+, 2+, 3+. To be 3+ there must be strong membrane staining of the complete cell in >10% of tumour cells; 2+ is moderate staining, whilst 1+ is weaker staining and the complete cell is not stained. Cytoplasmic staining is not considered. Since this was introduced there have been numerous papers concerned with HER-2 testing, comparing immunohistochemical assessment with HER-2 gene amplification as determined by *in situ* hybridization (ISH) (either fluorescence ISH or chromogenic ISH). Identification of a very strongly staining tumour as 3+ is not difficult. The problem comes with a proportion of the 1+ and the 2+,^{24–26} and this is more likely to arise with laboratories doing smaller numbers of tests.²⁷ Even reference centres can have a variation in the number of 2+ cases from 7% to 19%.²⁸ Selection for treatment is based on the presence of 3+ staining and 2+ with gene amplification as determined by ISH, so false negatives and false positives could have the wrong management.

What is positive?

This is not just a problem for immunohistochemistry, but applies to any system that uses semiquantitative or quantitative approaches such as ELISA or ligand binding assays. It helps if the immunohistochemical results can be compared with another analysis method

for which clinical data are available, e.g. ER immunohistochemistry with dextran-coated charcoal assays for the same tumour¹⁹ or for HER-2 with ISH for gene amplification.²⁸ This is not always the case and so different studies can have different cut-off points between positive and negative, e.g. moderate/strong staining of 10% of breast cancer cells for p53²⁹ or anything above 0 using the Allred score.³⁰ Comparison of findings from different studies of the same molecule can then become difficult, and researchers have to be careful in considering what is positive.

Automated analysis

Interpretation of immunohistochemistry is usually done manually and depends on the experience and alertness of the interpreter. Since there are problems in this component of the technique as well as the methodology, can automated analysis improve on our ability to quantify results? Computerized image analysis systems have been used since the late 1980s³¹ and have been shown to represent a more accurate means of quantification of ER than the H-score, which showed marked variability.³² One problem with using diaminobenzidine as a chromogen is that there is only a linear relationship between the amount of antigen and staining intensity at low levels of the latter.³³ Image analysis systems assess the amount of staining by measuring absorption, so the non-linear relationship that occurs at higher levels between amount of antigen and intensity can result in inaccurate readings. Recent approaches have used antibody-conjugated fluorophores and fluorescent microscope systems,^{8,34} but are still within the research rather than the diagnostic arena. Image analysis has been proposed as a more reliable method for assessing HER-2 immunostaining,^{35–37} but seems feasible only if assays are centralized to a small numbers of centres due to cost of the systems.

Conclusions

Advances have been made in immunohistochemistry, particularly at the technical level, but there are still issues around interpretation and quantification which need to be addressed. This is of increasing importance since it has become a routine assay for the assessment of therapeutic selection markers, which are likely to increase in number and be applicable to many common cancers. There must be easy to use, reproducible, standardized protocols for assessment which are international. Improvements in automated analysis with wider applicability could lead to standardization.

References

1. *Pathology reporting of breast disease*. NHSBSP Publication no. 58. Sheffield: NHSBSP 2005; 86–87.
2. Jacobs TW, Prioleau JE, Stillman IE, Schnitt SJ. Loss of tumour marker-immunostaining intensity on stored paraffin slides of breast cancer. *J. Natl Cancer Inst.* 1996; **88**: 1054–1059.
3. Bertheau P, Cazals-Hatem D, Meignin V *et al.* Variability of immunohistochemical reactivity on stored paraffin slides. *J. Clin. Pathol.* 1998; **51**: 370–374.
4. Cattoretti G, Pileri S, Parravicini C *et al.* Antigen unmasking on formalin-fixed, paraffin embedded tissue sections. *J. Pathol.* 1993; **171**: 83–98.
5. Rhodes A, Jasani B, Balaton AJ *et al.* Study of interlaboratory reliability and reproducibility of estrogen and progesterone receptor assays in Europe. Documentation of poor reliability and identification of insufficient microwave antigen retrieval time as a major contributory element of unreliable assays. *Am. J. Clin. Pathol.* 2001; **115**: 44–58.
6. von Wasielewski R, Mengel M, Wiese B, Rudiger T, Muller-Hermelink HK, Kreipe H. Tissue array technology for testing interlaboratory and interobserver reproducibility of immunohistochemical estrogen receptor analysis in a large multi centre trial. *Am. J. Clin. Pathol.* 2002; **118**: 675–682.
7. Mengel M, von Wasielewski R, Weise B, Rudiger T, Muller-Hermelink HK, Kreipe H. Inter-laboratory and inter-observer reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. *J. Pathol.* 2002; **198**: 292–299.
8. McCabe A, Dolled-Filhart M, Camp RL, Rimm DL. Automated quantitative analysis (AQA) of *in situ* protein expression, antibody concentration and prognosis. *J. Natl Cancer Inst.* 2005; **97**: 1808–1815.
9. McCormick D, Yu C, Hobbs C, Hall PA. The relevance of antibody concentration to the immunohistological quantification of cell proliferation-associated antigens. *Histopathology* 1993; **22**: 543–547.
10. McCluggage WG, Connolly LE, McGregor G, Hyland PL, Hall PA. A strategy for defining biologically relevant levels of p53 protein expression in clinical samples with reference to endometrial neoplasia. *Int. J. Gynecol. Pathol.* 2005; **24**: 307–312.
11. Press MF, Hung G, Godolphin W, Slamon DJ. Sensitivity of HER-2/neu antibodies in archival tissue samples: potential source of error in immunohistochemical studies of oncogene expression. *Cancer Res.* 1994; **54**: 2771–2777.
12. Wynford-Thomas D. p53 in tumour pathology: can we trust immunocytochemistry? *J. Pathol.* 1992; **166**: 329–330.
13. Hall PA, Lane DP. p53 in tumour pathology: can we trust immunohistochemistry?—revisited! *J. Pathol.* 1994; **172**: 1–4.
14. Hall PA, McCluggage WG. Assessing p53 in clinical contexts: unlearned lesions and new perspectives. *J. Pathol.* 2006; **205**: 1–6.
15. McCarty KS Jr, Miller LS, Cox EB *et al.* Estrogen receptor analyses: correlation of biochemical and immunohistochemical methods using monoclonal anti receptor antibodies. *Arch. Pathol. Lab. Med.* 1985; **109**: 716–721.
16. Gee JM, Borroso AF, Ellis IO, Robertson JF, Nicholson RI. Biological and clinical associations of c-jun activation in human breast cancer. *Int. J. Cancer* 2000; **89**: 177–186.
17. Abd El-Rehim DM, Ball G, Pinder S *et al.* High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *Int. J. Cancer* 2005; **116**: 340–350.
18. Reiner A, Neumeister B, Spona J, Reiner G, Schemper M, Jakesz R. Immunocytochemical localization of estrogen and progesterone receptor and prognosis in human primary breast cancer. *Cancer Res.* 1990; **50**: 7057–7061.
19. Harvey JM, Clark GM, Osborne CK, Allred DC. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J. Clin. Oncol.* 1999; **17**: 1477–1481.
20. Detre S, Saclani Jotti G, Dowsett M. A 'quick score' method for immunohistochemical semiquantitation validation for oestrogen receptor in breast carcinomas. *J. Clin. Pathol.* 1995; **48**: 876–878.
21. Barnes DM, Harris WH, Smith P, Millis RR, Rubens RD. Immunohistochemical determination of oestrogen receptor; comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients. *Br. J. Cancer* 1996; **74**: 1445–1451.
22. Fisher ER, Anderson S, Dean S *et al.* Solving the dilemma of the immunohistochemical and other methods used for scoring estrogen receptor and progesterone receptor in patients with invasive breast carcinoma. *Cancer* 2005; **103**: 164–173.
23. Rudiger T, Hoffer H, Kreipe H-H *et al.* Quality assurance in immunohistochemistry. Results of an interlaboratory trial involving 172 pathologists. *Am. J. Surg. Pathol.* 2002; **26**: 873–882.
24. Bartlett JM, Going JJ, Mallon EA *et al.* Evaluating HER2 amplification and over expression in breast cancer. *J. Pathol.* 2001; **195**: 422–428.
25. Tubbs RR, Pettay JD, Roche PC, Stoler MH, Jenkins RB, Grogan TM. Discrepancies in clinical laboratory testing of eligibility for trastuzumab therapy: apparent immunohistochemical false-positives do not get the message. *J. Clin. Oncol.* 2001; **19**: 2714–2721.
26. Zarbo RJ, Hammond ME. Conference Summary, Strategic Symposium: Her-2/neu testing of breast cancer patients in clinical practice. *Arch. Pathol. Lab. Med.* 2002; **126**: 803–808.
27. 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**: 852–854.
28. 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**: 418–423.
29. Thor AD, Moore DHII, Edgerton SM *et al.* Accumulation of p53 tumour suppressor gene protein: an independent marker of prognosis in breast cancers. *J. Natl Cancer Inst.* 1992; **84**: 845–855.
30. Allred DC, Clark GM, Elledge R *et al.* Association of p53 protein expression with tumour cell proliferation rate and clinical outcome in node-negative breast cancer. *J. Natl Cancer Inst.* 1993; **85**: 200–206.
31. Bacus S, Flowers JL, Press MF, Bacus JW, McCarty KS Jr. The evaluation of estrogen receptor in primary breast carcinoma by computer-assisted image analysis. *Am. J. Clin. Pathol.* 1988; **90**: 233–239.
32. McClelland RA, Wilson D, Leake R, Finlay P, Nicholson RI. A multicentre study into the reliability of immunocytochemical assay quantification. *Eur. J. Cancer* 1991; **27**: 711–715.

33. Fritz P, Wu X, Tuczek H, Multhaupt H, Schwarzmann P. Quantitation in immunohistochemistry. A research method or a diagnostic tool in surgical pathology? *Pathologica* 1995; **87**; 300–309.
34. Szeszel MK, Crisman CL, Crow L *et al*. Quantifying estrogen and progesterone receptor expression in breast cancer by digital imaging. *J. Histochem. Cytochem.* 2005; **53**; 753–762.
35. Wang S, Saboorian MH, Frenkel EP *et al*. Assessment of Her-2/neu status in breast cancer. Automated cellular imaging system (ACIS)-assisted quantification of immunohistochemical assay achieves high accuracy in comparison with fluorescence *in situ* hybridization assay as the standard. *Am. J. Clin. Pathol.* 2001; **116**; 495–503.
36. Bloom K, Harrington D. Enhanced accuracy and reliability of HER-2/neu immunohistochemical scoring using digital microscopy. *Am. J. Clin. Pathol.* 2004; **121**; 620–630.
37. Tawfik OW, Kimler BF, Davis M *et al*. Comparison of immunohistochemistry by automated cellular imaging system (ACIS) versus fluorescence *in-situ* hybridization in the evaluation of HER-2/neu expression in primary breast carcinoma. *Histopathology* 2006; **48**; 258–267.