



OPEN ACCESS

Getting controls under control: the time is now for immunohistochemistry

Emina Emilia Torlakovic,^{1,2,3} Søren Nielsen,^{4,5} Mogens Vyberg,^{4,5} Clive R Taylor⁶

¹Department of Laboratory Hematology, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada

²Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

³Canadian Immunohistochemistry Quality Control (CIQC), Vancouver, BC, Canada

⁴Department of Clinical Medicine, Institute of Pathology, Aalborg University Hospital, Aalborg, Denmark

⁵Nordic Immunohistochemical Quality Control (NordiQC), Aalborg, Denmark

⁶Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

Correspondence to

Dr Emina E Torlakovic, Department of Laboratory Hematology, Toronto General Hospital/UHN, 11th Floor, R411, 200 Elizabeth Street, Toronto, Ontario, Canada M5G 2C4; emina.torlakovic@uhn.ca

Received 22 June 2015

Revised 24 July 2015

Accepted 27 July 2015

ABSTRACT

For several decades, immunohistochemistry (IHC), more specifically diagnostic IHC (dIHC), has been considered an art rather than a laboratory test. There was no clarity about what test performance characteristics are relevant to dIHC, test performance characteristics were not fully defined for dIHC and partly as a consequence of that, there were no standardised controls or reference standards. Herein, we discuss the role of standardisation of external controls for test performance characteristics and the role of standardised controls and reference standards for overall standardisation of IHC.

Controls for immunohistochemistry (IHC) have taken centre stage following rapidly increasing use of IHC testing as prognostic and predictive markers, as ‘companion diagnostics’, providing a window into the molecular biology of tumours and in some instances facilitating effective targeted therapies, and also what may be considered as ‘next-generation IHC’ (NG-IHC).^{1–2} For example, the coupling of mass spectroscopy and IHC, so-called ‘mass spectroscopy IHC’, calls for greatly improved control of the IHC process. In this context, the requirement for greatly improved standardisation and reproducibility of IHC, along with true quantification, becomes paramount.^{3–5} Accordingly, there are higher expectations of accuracy for IHC testing, which inevitably translates to radically improved control systems, and adoption of the principles and concepts of quality assurance that apply to quantitative assays used in clinical laboratories for the measurement of analytes in blood and other body fluids.

An increased focus on controls for IHC also resulted from an action of the College of American Pathologists (CAP), whereby a longstanding CAP requirement for a large number of negative reagent controls was replaced by the option (at the discretion of the pathologist) that none are required if polymer/multimer detection systems are used for testing.⁶ This somewhat surprising reversal gave more dynamism to discussions already ongoing between global leaders in the field, about the need for standardisation of controls in context with changing clinical applications of IHC testing. Thus, the seemingly simple question of how to select appropriate controls for IHC testing finally is getting the spotlight it deserves. For decades, IHC has been an orphan test in the anatomical pathology (histology) laboratory, regarded as a variety of ‘special stain’ rather than, in essence, an immunoassay (although imperfectly controlled) closely akin to ELISA, which is a powerful quantitative clinical laboratory test, with many different

clinical and research applications. The orphan status of IHC is changing because anatomical pathologists are finally coming to terms with the notion that the principles of clinical laboratory testing and quality assurance may be employed to convert a qualitative IHC test, which is basically a descriptive in situ test with non-linearity and no universal reference standards, into something that may approach a tissue-based immunoassay, with a quantitative potential akin to ELISA, a prospect that has been termed ‘in situ proteomics’.⁷

External proficiency testing (PT) programmes (eg, NordiQC, UK NEQAS, CIQC) have demonstrated a general truth that, year after year, there are about one-third of laboratory participants that achieve optimal results, while one-third are ‘good’ and one-third fail.⁸ The question is posed—‘Why do so many clinical laboratories fail in PT challenges?’ Collective global experience in IHC PT has shown that one of the reasons is selection of inappropriate positive controls with tissues that show only high expression levels of the antigen of interest; this often leads to inadequate optimisation of the IHC tests. Therefore, it is not surprising that two recent papers having focus on standardisation of controls in IHC, one for negative controls and one for positive controls, were largely authored by PT programme leaders throughout the world.^{8–9} Only by standardising controls will it be possible to accomplish the following: (a) determine that the proper antibody was applied; (b) determine that expected technical sensitivity and specificity is achieved; (c) follow reproducibility from one test to another, from one run to another, from one laboratory to another and (d) transfer methodology from published literature to the clinical laboratory and from one laboratory to another. These two papers adopted a fresh and more rigorous approach to old concepts applicable to IHC controls and also introduced some new concepts. The present editorial aims to examine further the following topics: (a) what are IHC controls controlling? (b) What do the terms sensitivity and specificity mean in the context of IHC? (c) The use of IHC critical assay performance controls (iCAPCs). (d) The effective absence of ‘batching’ in automated IHC. (e) The importance of controls with respect to preanalytical conditions. (f) The position of industry and controls; reference standards and package inserts for IHC.

WHAT ARE IHC CONTROLS CONTROLLING?

Details of terminology and specific recommendations with respect to IHC controls may be found in the recently published guidelines.^{9–10} The intent in

To cite: Torlakovic EE, Nielsen S, Vyberg M, et al. *J Clin Pathol* Published Online First: [please include Day Month Year] doi:10.1136/jclinpath-2014-202705

this editorial is to emphasise that there, unfortunately, are no controls that control everything. From the perspective of the 'total test approach',¹¹ 'internal controls' come closest to serving as perfect controls, but only if they are present in the patient's sample (the test section), and only if they also (same or similar to iCAPCs) give information about the test calibration. The internal controls tell the pathologist if the test was successful or not. Internal controls reflect success or failure of the preanalytical and analytical components. However, once internal positive controls show that the test failed, it does not identify the point of failure. It could have been either the preanalytical or analytical phase that failed. In contrast, 'external controls' (as defined in references 9 and 10) control the analytical component. Therefore, when the internal positive control shows a successful result, the external control should also show the same, but when the internal control shows failed testing, the external control provides information as to whether it was the preanalytical or analytical phase that failed. When an internal positive control is not present, or if present is at such a high level of expression that it does not provide useful information about sensitivity and specificity, then it is necessary to rely on external controls as evidence of test analytical performance. In such circumstance, it is important to recognise the heavy reliance that is placed on departmental and institutional procedures to assure a satisfactory tissue processing/preanalytical component; degradation of the test protein by ischaemia or improper fixation will not be revealed by external positive controls alone. From a practicing pathologist's point of view, ideally, all IHC tests would have perfect internal controls. From the point of view of the IHC laboratory, in the common situation of absence of good internal controls, proper selection and evaluation of the external control on a daily basis is key to assurance of quality of performance of IHC tests. External controls in addition to internal controls are essential in monitoring the reproducibility of the IHC methodology.

WHAT DO THE TERMS SENSITIVITY AND SPECIFICITY MEAN IN THE CONTEXT OF IHC?

External controls should be designed so as to provide evidence that there is no major variation in 'technical sensitivity and specificity' of the analytical phase of the IHC test. The new three-tier classification of sensitivity and specificity of IHC distinguishes between technical (synonym: analytical), diagnostic and clinical sensitivity and specificity.⁹ Technical sensitivity and specificity cannot be accurately calculated when IHC is used as a qualitative test because it is merely a descriptive test with no linearity, and no available calibration controls against which to determine the ability to detect the analyte (protein/antigen). However, before we explore the three different types of sensitivity and specificity of the IHC test/protocol, there is also a distinction between 'antibody sensitivity and specificity' and 'IHC protocol sensitivity and specificity'. The distinction between the 'antibody' and 'IHC test/protocol' is often obscured in literature, and often the two are mixed up. It is entirely possible to have suboptimal or poor protocol even with the best, highest quality primary antibody (Ab) and the other way around; even with the low-affinity, 'difficult' Ab, it is often possible to design a very good or even optimal protocol, although using high-sensitivity primary antibody usually enables easier set-up of optimal IHC tests/protocols. For primary antibody, 'high sensitivity' usually implies that the primary antibody can be used at very high dilutions, while specificity infers that the test detects the targeted analyte only, while for IHC test/protocol, 'high sensitivity' usually implies that cells with very low levels of expression of antigen of interest will be

'stained' or 'positive'. This distinction is essential for proper understanding of the three-tier classification of sensitivity and specificity, which only relates to IHC test/protocol. Primary Ab is only one parameter of the complex multiparameter IHC test/protocol. Of interest, the terms technical sensitivity and specificity of the primary Ab can be measured by the given amount of target antigen in a liquid-based assay such as ELISA, but this is not possible in histological sections *in situ*. Similarly, it is not possible to measure or calculate technical sensitivity and specificity of the IHC test/protocol; thus, both are descriptive and provide only an approximation of 'how much' positivity one can expect. The lack of linearity of the assay is mainly due to the employment of amplification, which is achieved by using detection systems. Detection systems may use various chemical approaches (eg, avidin-biotin, multimers/polymers), but irrespective of the approach, their routine applications are such that they facilitate detection of antigen, which is mostly unrelated to the antigen amount. Therefore, with respect to technical sensitivity, measuring intensities and defining exact cut-off points based on the signal intensity observed in an IHC test does not make sense in a non-linear assay with multiple amplification steps. However, experience in NordiQC has demonstrated the value of establishing practical biological end points reflective of the lower limit of detection of an IHC method, by using 'low expressor' (LE) positive control tissues, that is, tissues/cells that have been shown to reproducibly express only low levels of the protein in question alongside the more usual 'high expressor' (HE) tissues. For example, a modern 'sensitive' IHC test for CD45, performing properly, is needed to convincingly detect Kupffer cells (LE) in the liver and not only lymphocytes and other HE cell types. Achieving desirable and optimal results is of paramount importance in 'total test approach' to IHC testing and ensures that all preanalytical and analytical parameters were satisfactory (which also includes the selection of proper, sensitive primary Ab).¹¹

Technical specificity may be usefully evaluated by using tissues that frequently reveal non-specific reactions secondary to components of the detection systems (such as biotin) or due to a primary antibody concentration that is too high or other less predictable causes; sections of liver, kidney and smooth muscle are useful in this context. However, pathologists should always be aware that unexpected 'specific' cross-reactions with unknown epitopes do sometimes occur, and will not necessarily be revealed by the usual external controls. If the protocol is stable, the results should be reproducible; this also applies to technical sensitivity and specificity. Technical sensitivity and specificity should not vary in a reproducible IHC assay. The technical sensitivity and specificity concept is based on achieving the highest signal-to-noise ratio and is fully applicable to diagnostic (class I) IHC tests. However, this is not directly applicable to predictive (class II) IHC biomarkers where a range of expression levels that reflects the intended clinical application needs to be considered. Hence, the calibration of the controls for prognostic markers needs to reflect this range too.

In the tree-tier classification of sensitivity and specificity, 'diagnostic sensitivity and specificity' is used in reference to the practical utility of the IHC test in a diagnostic setting (eg, S-100 is a sensitive, but non-specific marker for diagnosis of melanoma), for this usage is distinct from what usually is meant by technical sensitivity and specificity of IHC, by which we determine the ability of the test to detect weakly expressed antigens (sensitivity), and do not cross-react with other known or unknown epitopes (specificity). In contrast to technical, diagnostic sensitivity and specificity are calculated from the number of true and false positive/negative cases. Information pertaining to

'diagnostic sensitivity and specificity' may be derived from intra-laboratory experience, from the published literature or sometimes from data sheets provided by the manufacturer for a specified antibody clone, and for specified reagents and protocols (eg, DOG1 is a sensitive and specific marker for diagnosis of gastrointestinal stromal tumour).¹² To a degree, this information is transferable to the laboratory if the same clone is employed, and the same detection reagents and platform, with the important proviso that preanalytical steps may affect performance outcome, and thus diagnostic sensitivity. To be able to transfer information on diagnostic sensitivity and specificity from the literature, if a different clone (to the same protein target or even to the same epitope) is to be used, or different protocols (eg, different automated stainer platform, different detection system, etc), requires a more complete understanding of how differences in technical sensitivity and specificity, and preanalytical steps may affect the overall outcome of the staining reaction in the diagnostic context with the laboratory. It is here that careful description of controls/control tissues (such as iCAPCs, see below) used for test development is most helpful for methodology transfer. Unfortunately, such information is usually not included in the published literature.

The terms 'clinical' sensitivity and specificity have been used recently with a narrow application to IHC testing of prognostic and predictive markers, so-called 'companion diagnostics' or 'advanced personalised diagnostics', and are determined in formal clinical trials (prospectively or retrospectively).

THE USE OF ICAPCS

The concept of iCAPCs unites the ideas of technical sensitivity and specificity of the IHC, methodology transfer and proper reference standards. iCAPCs are selected based upon documented performance of defined LE and HE tissues so as to serve as 'gold standards' or IHC 'primary' controls.^{8,9} The first 18 iCAPCs for frequently used IHC tests were described in the recent positive controls paper.⁹ Others exist in concept and are yet to be developed, a process that ideally will be driven by national and international quality assurance programmes that provide IHC PT in collaboration with the relevant subspecialty societies. The end goal is to develop a system for 'controlling controls'. In the meantime, IHC laboratory directors can use those examples already published (see references and also 'controls' on <http://www.nordiqc.org>) to refine their selection of tissues to serve as controls. Journal editors and reviewers also have a role to play in raising the level of scrutiny of the controls employed in submitted research, especially with respect to diagnostic, prognostic and predictive use of IHC.

THE EFFECTIVE ABSENCE OF 'BATCHING' IN AUTOMATED IHC

Although not timely recognised, a widespread adoption of automated IHC effectively ended the practice of batching in IHC where groups of 10 or 20 or more slides were run together with a single set of 'batch' control. Modern automated technology is such that every slide is in fact a separate run, and, therefore, with absent on-slide controls, there is no control for that particular slide. So-called 'run' or 'batch' controls only show how IHC test is performed on that particular control slide. Hence, on-slide controls should always be used. On-slide controls are external controls, usually tissue controls (occasionally cell line preparations), that incorporate both positive and negative tissue elements, ideally including such carefully selected tissues that they serve as iCAPCs.

THE IMPORTANCE OF CONTROLS WITH RESPECT TO PREANALYTICAL CONDITIONS

Most clinical IHC laboratories use diagnostic tissue (archived or excess diagnostic tissue) for external controls. The presumption is that this tissue has been processed in the same way (ischaemic time, fixation, decalcification, etc) as patient tissue that is being tested. The presumption is, of course, false in a detailed sense; at best, it is similar. Likewise, successive controls that are selected—as the initial tissues are exhausted—also are at best similar, and all may differ in unknown ways that may or may not affect the outcome of an IHC test for any specified targeted protein. TMAs (tissue microarrays) are increasingly used as controls and as reference tissues for clinical studies, but in TMAs, the problem of preanalytical variation is magnified many fold because each tissue core has undergone different, and often unknown, ischaemic and fixation times. Rigorous attention to the details of sample preparation, with documentation of ischaemic time and fixation time, provides valuable information as to the effects of specified proteins,¹³ and may minimise differences, but cannot eliminate them entirely.

The published recommendations for positive controls⁹ state that tissue controls should be fixed similar to patient samples 'within clinically applicable range' and that they optimally should be processed just like clinical samples (decalcified for the purpose of decalcified bone analysis or other as applicable).

But the matter is not that simple. For example, in theory, external controls should have constant time of fixation rather than 'clinically acceptable range' because their purpose is to detect variation in the analytical phase. If successive tissues that are selected as controls are themselves changing, it becomes difficult to determine if any changes observed in the quality (intensity) of the IHC result are due to variation in the analytical phase or variation in the control that is used that day. It may be more appropriate to take the approach that external controls should always be fixed the same time as predetermined by the laboratory (eg, exactly 24 h). At the moment, there are no published data available to support the effort required; nonetheless, it is important to recognise that a compromise has been made.

External controls developed from peptides, cell lines, histoids or xenografts^{9,10} have potential advantages in terms of assurance of conditions of preparation and even the actual amounts present of the protein(s) of interest; it is clear that there are differences in preanalytical treatment between these controls and the tested tissues. However, because of the potential for reproducible production, such preparations do perhaps provide the better option for precise control of the analytical phase.

THE POSITION OF INDUSTRY AND CONTROLS; REFERENCE STANDARDS AND PACKAGE INSERTS FOR IHC

The need for improved and standardised controls is also dictated by the development of personalised medicine, which requires rigorous quality assurance (QA) in order that clinically relevant sensitivity and specificity are achieved. Industry has picked up on this requirement, and has started to develop 'reference standards' for some class II and III tests (US Food and Drug Administration (FDA)) or class 2 (Canadian Association of Pathologists (CAP-ACP)).¹⁴ Such reference standards for IHC must also be defined for specific use. It is not sufficient to 'package' as cell blocks two or more positive and negative cell lines; the performance characteristics of such 'reference standards' must also be studied, defined and made available. FDA-approved commercial 'kits' for 'companion diagnostics' or predictive and prognostic markers typically contain cell line

controls providing some assurance that such controls are 'fit for use' (and what the intended use actually is). On the other hand, tissue controls selected and documented to function as iCAPCs, or as primary IHC controls, are biological IHC reference standards for which experience has shown reproducible performance, to the extent that they can be used to validate other types of (secondary) IHC controls, including engineered cell lines, xenografts or histoids. However, if cell lines are properly validated, they can be very useful for both new IHC protocol development and daily quality control as well as for PT by external QA programmes.^{15 16} Properly validated cell lines may, in fact, be superior to histological tissue as controls for quantitative IHC assays, especially if combined with image analysis.^{17 18}

One other issue should be considered in this context, namely, 'package inserts' or the antibody specification or data sheets that are provided together with primary antibodies. Such specification sheets may include whether the antibody has been shown useful for IHC on formalin-fixed paraffin-embedded tissues and often include 'suggested positive controls'. While such recommendations may serve as guidelines, they in no way substitute for the necessity that each clinical IHC laboratory develops its own appropriate controls. The responsibility is that of the pathologist (s). Selection of appropriate controls is not purely a technical issue, not just about the best signal-to-noise ratio; most importantly, it is about relevant/desirable test calibration that is appropriate for specific clinical use. It requires in-depth knowledge of how the tests are used ('fit for use'). As such, the responsibility for selection and monitoring of appropriate and proper controls is that of the pathologist(s) who will be relying upon these tests in their diagnostic work, which is as it should be.

It cannot be overemphasised that IHC laboratory needs collaboration and proactive involvement of technologists, technicians as well as pathologists. In some countries, pathologists are not actively involved in the day-to-day function of the IHC laboratory. This approach is wrong since IHC laboratories need pathologists (not just in the role of medical director). The knowledge of how the results are used is essential for proper protocol set-up and monitoring. Interpretation of the results of daily positive and negative tissue controls can be performed by

technologists, especially if the controls are standardised; however, interpretation of various types of errors that may affect controls as well as patient samples should be done by the pathologist. To achieve these goals, technologists need to learn some aspects of biology/pathology, and pathologists need to learn some technical aspects and understand IHC methodology and work as a team.

Handling editor Runjan Chetty

Competing interests None declared.

Provenance and peer review Commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- 1 Taylor CR. Predictive biomarkers and companion diagnostics. The future of immunohistochemistry: "in situ proteomics", or just a "stain"? *Appl Immunohistochem Mol Morphol* 2014;22:555–61.
- 2 Gown AM. Next Generation Immunohistochemistry; a window onto the molecular biology of tumors. <http://www.idpctokyohomepage.org/2013abstract/Dr.%20Gown.pdf> (accessed 9 Jun 2015).
- 3 Rimm DL. Next-gen immunohistochemistry. *Nat Methods* 2014;11:381–3.
- 4 Giesen C, Wang HA, Schapiro D, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* 2014;11:417–22.
- 5 Angelo M, Bendall SC, Finck R, et al. Multiplexed ion beam imaging of human breast tumors. *Nat Med* 2014;20:436–42.
- 6 College of American Pathologists. *College of American Pathologists Anatomic Pathology Checklist CAP Accreditation Program*. North-field, IL: College of American Pathologists, 2013:35–7.
- 7 Taylor CR. Quantitative in situ proteomics; a proposed pathway for quantification of immunohistochemistry at the light-microscopic level. *Cell Tissue Res* 2015;360:109–20.
- 8 Nielsen S. External quality assessment for immunohistochemistry: experiences from NordiQC. *Biotech Histochem* 2015;90:331–40.
- 9 Torlakovic EE, Nielsen S, Francis G, et al. Standardization of positive controls in diagnostic immunohistochemistry: recommendations from the International Ad Hoc Expert Committee. *Appl Immunohistochem Mol Morphol* 2015;23:1–18.
- 10 Torlakovic EE, Francis G, Garratt J, et al. International Ad Hoc Expert Panel. Standardization of negative controls in diagnostic immunohistochemistry: recommendations from the international ad hoc expert panel. *Appl Immunohistochem Mol Morphol* 2014;22:241–52.
- 11 Taylor CR. The total test approach to standardization of immunohistochemistry. *Arch Pathol Lab Med* 2000;124:945–51.
- 12 Espinosa I, Lee CH, Kim MK, et al. A novel monoclonal antibody against DOG1 is a sensitive and specific marker for gastrointestinal stromal tumors. *Am J Surg Pathol* 2008;32:210–18.
- 13 Neumeister VM, Anagnostou V, Siddiqui S, et al. Quantitative assessment of effect of preanalytic cold ischemic time on protein expression in breast cancer tissues. *J Natl Cancer Inst* 2012;104:1815–24.
- 14 Press Release: Horizon Discovery Group plc Introduces Range of Immunohistochemistry HDx Reference Standards. <http://www.horizondx.com/News/cat/articles/post/PressRelease06052015/> (accessed 16 Jun 2015).
- 15 UK NEQAS; Immunocytochemistry & In Situ Hybridisation. <http://www.ukneqas.org.uk/content/PageServer.asp?S=1624207&C=1252&Type=N&AD=16&SID=47> (accessed 19 Jul 2015).
- 16 Canadian Immunohistochemistry Quality Control (CIQC). Assessor's report for CIQC Run 22: ER, PR and HER2. <http://cpqa.ca/main/wpcontent/uploads/2014/09/CIQC-Run-22-summary.pdf> (accessed 19 Jul 2015).
- 17 Cheung CC, Neufeld H, Lining LA, et al. The laboratory score/reference method score ratio (LSRSR) is a novel tool for monitoring laboratory performance in immunohistochemistry proficiency testing of hormone receptors in breast cancer: the CIQC experience. *Am J Clin Pathol* 2011;136:67–73.
- 18 Sørensen M, Hansen L, Nielsen S. Cell lines as the basis for quality assurance of immunohistochemical assay for Ki67 in tissue. <http://nordiqc2015.dk/p9-cell-lines-as-the-basis-for-quality-assurance-of-immunohistochemical-assay-for-ki67-in-tissue.aspx> (accessed 24 Jul 2015).

Take home messages

- Unfortunately, there are no IHC controls that control everything.
- There is a difference between technical and diagnostic sensitivity and specificity as well as between primary antibody sensitivity and specificity and IHC protocol sensitivity and specificity.
- The concept of iCAPCs unites the ideas of technical sensitivity and specificity of the IHC, methodology transfer, and IHC reference standards.
- Selection of appropriate controls is not purely a technical issue. It is about relevant/desirable test calibration that is appropriate for specific clinical use. It requires in-depth knowledge of how the tests are used ('fit for use'). As such, the responsibility for selection and monitoring of IHC controls is that of the pathologist(s) who will be relying upon these tests in their diagnostic work.