Academic and Nonacademic Laboratories Perform Equally on CIQC Immunohistochemistry Proficiency Testing

Zhongchuan Will Chen, MDCM,1 Heather Neufeld,2 Maria A. Copete, DDS, MS,3 John Garratt, RT,4 C. Blake Gilks, MD,5 and Emina E. Torlakovic, MD, PhD6

Key Words: Academic center; External quality assessment; Immunohistochemistry; Proficiency testing

DOI: 10.1309/AJCPHBL59MWBVYRD

Abstract

Objectives: To test whether academic centers (ACs) are more successful than nonacademic centers (NACs) in immunohistochemistry (IHC) external quality assessment challenges in the Canadian Immunohistochemistry Quality Control (CIQC) program.

Methods: Results of 9 CIQC challenges for breast cancer marker (BM) and various non–breast cancer marker (NBM) tests were examined. Success rates were compared between AC/NAC laboratories and those located in small or large cities. Performance was also correlated with annual IHC case volumes.

Results: There was no statistically significant difference in performance in any of the comparisons. However, overall performance on BM was significantly better (P < .0001, t test) than on NBM tests regardless of AC/NAC nature or city size. The mean failure rate on NBM was approximately twice that of BM tests.

Conclusion: Our results suggest that recent emphasis on breast hormone IHC quality assurance has led to improved test quality.

Academic hospitals in large population centers are typically regarded as regional or national centers of excellence in health care in Canada. These centers are expected to have larger patient volumes and higher surgical pathology caseloads than smaller community-based hospitals. Studies have shown that the concept of “practice makes perfect” holds true in the clinical setting, with physicians having experience with larger patient caseloads performing better in terms of clinical outcome.1-5 However, to our knowledge, studies specific to the pathology laboratory comparing performance in immunohistochemistry (IHC) based on medical center type (ie, academic centers [ACs] vs nonacademic centers [NACs]) have not been published. Participation in external quality assurance (EQA) programs is mandated by many laboratory accreditation programs.6-8 Although no studies have definitively proven that performance on proficiency testing (PT) directly reflects day-to-day performance, it is an objective measurement that enables comparison of a laboratory’s performance with that of reference laboratories and the peer group.9 We hypothesized that the trend of greater caseloads resulting in better performance seen in physician-based clinical studies also applies to the pathology laboratory in the setting of IHC testing.1-5 Thus, laboratories with larger caseloads, such as those of academic tertiary hospitals or hospitals based in large population centers, may be more successful in IHC PT than community-based hospitals in smaller population centers. Using participant results obtained from the Canadian Immunohistochemistry Quality Control (CIQC) program, a voluntary and free IHC EQA program available to all Canadian laboratories, we examined IHC PT performance in various breast and non–breast cancer markers and compared the results of laboratories in ACs with those
in NACs, as well as laboratories located in large population centers with those in small centers.

Materials and Methods

Data Set

Data from 18 CIQC breast marker (BM) IHC challenges and 19 non-breast marker (NBM) IHC challenges performed by participating laboratories from 2008 through 2011 (CIQC assessment runs 3, 4, 5, 6, 7, 8, 11, 13, and 14) were included in this study. The BM tests included 6 separate estrogen receptor (ER), progesterone receptor (PR), and HER2 challenges. The NBM tests included the following challenges: 1 CD45, 1 CD20, 1 CD3, 2 cyclin D1, 1 Bcl-2, 1 Bcl-6, 2 Ki-67, 2 pan-keratin, 2 low-molecular-weight keratin, 1 high-molecular-weight keratin, 2 cytokeratin (CK) 7, 2 CK20, and 1 CK5.

Participant Categorization

A participating laboratory was designated as an AC if it was affiliated with a teaching university having an anatomic pathology residency program. All other laboratories were designated as NACs. A laboratory was designated as “large city” if it was located in a city with a population greater than or equal to 300,000 and “small city” if located in a city with a population of less than 300,000. A survey was also sent to all participants with the question, “What is the approximate number of IHC slides stained in your laboratory per year?” Participants then responded by choosing one of the following ranges: less than 5,000, 5,000 to 10,000, greater than 10,000 to 30,000, greater than 30,000 to 50,000, and greater than 50,000.

CIQC Testing Method

Although there were slight variations in the exact procedures of each CIQC challenge, in general, participants were mailed unstained slides with 10% formalin-fixed, paraffin-embedded tissue, often in tissue microarray (TMA) format, to be stained with their routine in-house IHC protocol for each marker requested in the test. Stained slides were returned to the CIQC to be evaluated by a panel of expert assessors. Participants were also required to provide details of the protocol used for each IHC test, including primary antibody clone and manufacturer, dilution, incubation time, epitope retrieval method, and detection system used.

In brief, BM tests were performed on TMAs containing up to 54 tissue cores. Participants’ results were correlated with those generated by CIQC reference laboratories, generally regarded as regional centers of excellence for IHC. ER and PR results were determined to be positive if 1% or more of the cells showed nuclear positivity. HER2 IHC was scored on a 0 to 3+ scale per American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines. When a participant core was assessed as 2+ (equivocal), it was by default determined to be concordant with the reference value regardless of the actual reference value because these patients would be referred for HER2 in situ hybridization and appropriately stratified for therapy based on those results. All reference HER2 IHC results were correlated with fluorescence in situ hybridization (FISH) results to ensure concordance. The PathVysion HER2 DNA Probe kit (Abbott Molecular, Abbott Park, IL) was used for FISH. Each core was examined and a count made of Cep17 and HER2 signals. A HER2/Cep17 ratio of less than 1.8 was considered nonamplified, greater than 2.2 was considered amplified, and 1.8 to 2.2 was considered equivocal. For the purposes of this study, the percentage of correct results achieved by each participant in each BM test was converted to a binary score as unsatisfactory and satisfactory, with a cutoff of at least 90% agreement with the designated reference value.

The NBM tests were performed by using TMAs or whole-tissue sections. The tissue sections or cores on each slide were selected to evaluate the expected staining pattern not only in the neoplastic or diseased tissue targeted by the IHC test but also in normal tissue. Expert assessors were chosen from a panel of pathologists and senior IHC medical laboratory technologists from Canadian academic centers with extensive experience in interpreting IHC results for the test of interest. They also received on-site training in applying the CIQC scoring system to examine the participants’ results. Slides were evaluated using criteria specific to each IHC test. Generally, these criteria evaluated appropriate staining of the antigen targeted by the IHC test and the presence of false-positive or false-negative staining in either diseased or normal tissue. The quality of staining and the proper calibration of staining protocols were also assessed by comparing the signal-to-noise ratio in staining to biologically expected results and that achieved by the reference laboratory. A detailed method for one of the tests has been described. Participants were assigned a 3- or 4-tier score for each NBM test, with 1 being the least optimal result and 3 or 4 being the most optimal result. For this study, the tiered score was converted to a binary score as follows: tier 1 was unsatisfactory; all other tiers were satisfactory.

Statistical Analysis

A percent unsatisfactory (%UNS) rate was calculated for each participating laboratory by dividing the number of unsatisfactory tests (following binary conversion of the original score) by the total number of tests it performed. The %UNS was compared between AC and NAC and large-city and small-city laboratories using SPSS version 19 (IBM, Armonk, NY). Means in each category were compared using the t test. The %UNS was also compared between the various volume
ranges indicated in the volume survey to explore any possible correlation between volume and performance using P values generated from the Spearman correlation coefficient.

Results

Participants

A total of 72 laboratories participated in at least 1 IHC challenge. Of these, 32 laboratories (44%) were ACs, 40 (56%) were NACs, and 47 (65%) were large-city and 25 (35%) small-city laboratories. The number of participants in each challenge ranged from 17 to 53. Of the laboratories, 67 participated in at least 1 BM test and 70 in at least 1 NBM test. Thirty-eight participants responded to the IHC slide volume survey (53% response rate). In general, ACs were significantly more likely to be associated with higher IHC volumes (P = .008; \( \chi^2 \) test), and large-city laboratories were more likely to report volumes greater than 10,000, although the \( \chi^2 \) test did not reach statistical significance (P = .30) \( \chi^2 \) Table 1.

Statistical Analysis

There was no statistically significant difference in the mean and range of %UNS rates when comparing ACs and NACs for all tests, NBM tests, or BM tests. However, the mean %UNS rate was 2 times higher in NBM tests (ACs, 20%; NAC, 21%) than in BM tests (ACs and NACs, both 10%) when comparing within each category of participants (ACs vs NACs). This comparison of means and ranges reached statistical significance (P < .0001). Likewise, a comparison of large-city and small-city laboratories yielded similar results, with similar mean %UNS rates across test categories. However, when BM and NBM tests were compared within each participant category (large vs small city), the mean %UNS rate of NBM tests (24%) was 2.7 times higher than that of BM tests (9%) among small-city laboratories (P < .0001) \( \chi^2 \) Table 2 and \( \chi^2 \) Figure 1. In addition, comparison of performance to volume based on the results of the survey did not reveal significant differences or trends in the analysis of NBM, BM, or all tests \( \chi^2 \) Figure 2.

Discussion

The results of this study show that there was no significant difference in the performance of ACs vs NACs and large-city vs small-city laboratories on various CIQC IHC tests, with further analysis of results of a case volume survey also confirming the absence of any significant difference in performance based on volume of IHC slides processed alone. These findings refute our initial hypothesis that ACs and large-city laboratories would be expected to perform better. This hypothesis was based not only on the previously mentioned studies showing better clinical outcomes in physicians with larger caseloads but also on studies in clinical pathology testing showing that laboratories doing cholesterol testing that performed more than 25 tests per day had lower error rates on a PT involving a cholesterol specimen with a standardized value.12 The phenomenon of better performance by laboratories with higher caseloads has also been described in gynecologic cytology PT.13,14 These studies led us to hypothesize that ACs and large-city laboratories, which are more likely to have large caseloads (a trend also seen in our survey of IHC case volumes), would perform better on IHC PT than their

Tables:

**Table 1** Results of Immunohistochemistry Case Volume Survey With Comparison of Laboratory Type to Volume

<table>
<thead>
<tr>
<th>Volume (in Thousands)</th>
<th>No. of Academic Laboratories (n = 21)a</th>
<th>No. of Nonacademic Laboratories (n = 17)b</th>
<th>No. of Large-City Laboratories (n = 24)b</th>
<th>No. of Small-City Laboratories (n = 14)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5-10</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>&gt;10-30</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>&gt;30-50</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>&gt;50</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

\( \chi^2 = 13.7; P = .008. \)

\( \chi^2 = 4.92; P = .30. \)

**Table 2** Unsatisfactory Results for Various Laboratory and Test Categories

<table>
<thead>
<tr>
<th>Laboratory Type</th>
<th>Mean (Range) of Unsatisfactory Result, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Tests</td>
</tr>
<tr>
<td>Academic center</td>
<td>14 (0-38)</td>
</tr>
<tr>
<td>Nonacademic center</td>
<td>14 (0-35)</td>
</tr>
<tr>
<td>Small city</td>
<td>15 (0-35)</td>
</tr>
<tr>
<td>Large city</td>
<td>14 (0-38)</td>
</tr>
</tbody>
</table>

© American Society for Clinical Pathology
counters. In fact, ASCO/CAP guideline recommendations for HER2 testing also hint at the fact that laboratories with larger case volumes may perform better and advise laboratories performing few HER2 tests annually to refrain from testing and refer specimens to a “central laboratory with more experience and volume.” However, the guidelines do note that there is “no systematic evidence for a relationship of volume to test accuracy.” In practice, certain national guidelines, such as those of Canada and the United Kingdom, restrict HER2 testing to laboratories performing at least 250 IHC tests annually.

The fact that ACs/NACs and large-city/small-city laboratories, regardless of volume of IHC slides processed yearly, performed at an equal level is somewhat reassuring as to the uniformity of IHC testing provided across the country in these laboratories. However, what is not reassuring is the quality of test results as measured by the %UNS rate. A mean %UNS rate of up to 24% in the NBM category suggests that there is significant room for improvement in the quality of NBM tests in all laboratories. Interestingly, %UNS rates in NBM tests were obtained after weighting the results heavily in the participants’ favor, as only the lowest-tier score in the initial scoring system was assigned an unsatisfactory score in the binary score conversion system used in this study. Clearly, laboratories should not settle for less than optimal IHC staining quality. Therefore, an even larger number of participants would have achieved an unsatisfactory score had more stringent criteria been used. Studies have shown that only 23% of participants in the Nordic Immunohistochemical Quality Control program and 27% in the United Kingdom National External Quality Assessment Scheme for Immunocytochemistry performed optimal staining for detection of cyclin D1. In our own analysis of one of the cyclin D1 challenges included in this study (data not shown), only 34% of participants achieved the highest score in a 3-tier scoring system.

Interestingly, the somewhat unexpected result that all participants, regardless of AC/NAC or large-city/small-city nature, performed worse on NBM tests than on BM tests may be evidence that recently published guidelines and emphasis on quality assurance in breast hormone receptor testing have translated into tangible improvements in test performance when compared with NBM tests. In addition, the media attention placed on certain breast cancer testing errors in Canada may have pushed Canadian laboratories to adhere more closely to published guidelines or to optimize their BM IHC tests better. However, this study suggests that the heavy focus on BM testing quality has not been applied to NBM tests, resulting in at least a 1.6- to 2.7-fold increase in unsatisfactory NBM tests as compared with BM tests. One limitation of this study is that participants performing BM tests are likely to have annual volumes of more than 250 tests as required by Canadian guidelines. Therefore, this study does not

![Figure 1](image-url) Comparison of the percentage of poor results on non–breast and breast marker challenges among (A) academic centers and nonacademic centers and (B) laboratories located in small cities and large cities.
not provide information about the BM test %UNS rate for laboratories that have IHC volumes below this cutoff.

Although our study suggests that there is no difference in the quality of IHC testing based on laboratory size, academic affiliation, or case volume, it does not endorse the notion that all IHC tests should or can be performed in laboratories of any size. Other important factors independent of testing quality may require allocation of certain tests to certain types of laboratories. One such factor is the ability to properly validate tests and to maintain an appropriate level of competency for the interpretation of results. This is particularly important in rare diseases, in which the number of positive cases for controls and validation or revalidation may be insufficient if cases are not concentrated at a centralized laboratory. Large numbers of small laboratories performing class II IHC testing for rare diseases may also create problems in setting up and maintaining a proper EQA/PT program. For instance, an EQA program may find it difficult to provide a sufficient number of PT samples for all laboratories that wish to perform a test for a rare disease at low volume. Therefore, these types of tests are likely best concentrated in larger laboratories, irrespective of quality of testing in smaller laboratories, due to practical and economic concerns.

**Figure 2** Comparison of the percentage of poor results to the volume of immunohistochemistry (IHC) slides stained yearly on all (A), non-breast marker (B), and breast marker (C) challenges. A, P = .70 (Spearman correlation). B, P = .90 (Spearman correlation). C, P = .19 (Spearman correlation).
In summary, ACs/NACs and large-city/small-city laboratories, irrespective of annual IHC test volume, performed equally on all IHC tests in the CIQC EQA program.

References