

# Standardization of Hemagglutination Inhibition Assay for Influenza Serology Allows for High Reproducibility between Laboratories

Mary Zacour,<sup>a,b</sup> Brian J. Ward,<sup>a</sup> Angela Brewer,<sup>a</sup> Patrick Tang,<sup>c</sup> Guy Boivin,<sup>d</sup> Yan Li,<sup>e</sup> Michelle Warhuus,<sup>f</sup> Shelly A. McNeil,<sup>f,g</sup> Jason J. LeBlanc,<sup>f,h</sup> Todd F. Hatchette,<sup>f,g,h</sup> on behalf of the Public Health Agency of Canada and Canadian Institutes of Health Influenza Research Network (PCIRN)

Research Institute of the McGill University Health Centre, Montreal, QC, Canada<sup>a</sup>; BioZac Consulting, Pointe Claire, QC, Canada<sup>b</sup>; British Columbia Centre for Disease Control, Vancouver, BC, Canada<sup>c</sup>; CHU de Quebec and Laval University, Quebec City, QC, Canada<sup>d</sup>; National Microbiology Laboratory, Winnipeg, MB, Canada<sup>e</sup>; Canadian Center for Vaccinology, Dalhousie, Halifax, NS, Canada<sup>f</sup>; Departments of Medicine<sup>g</sup> and Pathology and Laboratory Medicine,<sup>h</sup> Nova Scotia Health Authority, Halifax, NS, Canada

Standardization of the hemagglutination inhibition (HAI) assay for influenza serology is challenging. Poor reproducibility of HAI results from one laboratory to another is widely cited, limiting comparisons between candidate vaccines in different clinical trials and posing challenges for licensing authorities. In this study, we standardized HAI assay materials, methods, and interpretive criteria across five geographically dispersed laboratories of a multidisciplinary influenza research network and then evaluated intralaboratory and interlaboratory variations in HAI titers by repeatedly testing standardized panels of human serum samples. Duplicate precision and reproducibility from comparisons between assays within laboratories were 99.8% (99.2% to 100%) and 98.0% (93.3% to 100%), respectively. The results for 98.9% (95% to 100%) of the samples were within 2-fold of all-laboratory consensus titers, and the results for 94.3% (85% to 100%) of the samples were within 2-fold of our reference laboratory data. Low-titer samples showed the greatest variability in comparisons between assays and between sites. Classification of seroprotection (titer  $\geq 40$ ) was accurate in 93.6% or 89.5% of cases in comparison to the consensus or reference laboratory classification, respectively. This study showed that with carefully chosen standardization processes, high reproducibility of HAI results between laboratories is indeed achievable.

The hemagglutination inhibition (HAI) assay is the primary method for determining quantitative antibody titers for influenza virus and is widely used both for licensure of vaccines and for seroepidemiologic studies examining protection in populations (1–3). The assay relies on the ability of the hemagglutinin protein on the surface of influenza virus to bind to sialic acids on the surface of red blood cells (RBCs) (4). If the patient's serum contains antibodies that block viral attachment, this interaction is inhibited. Direct comparison of results between studies has been problematic, as the reproducibility of HAI assays between laboratories has historically been poor (5–12). These studies have shown that HAI titers reported for identical specimens in different laboratories can vary as much as 80-fold or 128-fold (9, 11), with the geometric coefficient of variation (GCV) as high as 803% (5). This variability in results may relate to differences in biological reagents, protocols, and personnel training. The use of international standards (IS) may reduce interlaboratory variability (6, 11, 13), but such reagents currently exist only for influenza A H1N1 and H5N1 clade 1 viruses (14, 15). The need for standardization of HAI assays and other laboratory methods (e.g., microneutralization [MN] assays) has been highlighted as a priority of CONSISE, the Consortium for the Standardization of Influenza Seroepidemiology (16, 17). CONSISE collaborators have recently published data showing that a standardized MN protocol improves the comparability of serologic results between laboratories (18); however, the consortium has not yet assessed the effect of standardization on HAI assay variability.

As part of a large multicenter influenza research network (Public Health Agency of Canada/Canadian Institutes of Health Research Influenza Research Network [PCIRN]), we attempted to rigorously standardize HAI testing across Canada via common

training with a shared, consensus protocol and the use of common reagents and seed virus at all five participating academic and public health laboratories. This report shows that, with a commitment to this level of coordination and standardization, results of HAI testing can indeed be comparable across different laboratories.

## MATERIALS AND METHODS

**Serum samples.** Institutional review board (IRB)-approved informed consent for this study's use of residual sera from human studies conducted across the PCIRN was obtained at all participating sites prior to the original studies. For this study, residual sera were selected based on their original HAI titer estimations and pooled to create large-volume, standardized human serum panels of 10 samples per virus at antibody titers that were negative, low, medium, and high. The specimens were deidentified and divided into aliquots by a single site of the PCIRN, frozen at  $-80^{\circ}\text{C}$ , and shipped on dry ice to the other four participating laboratories for HAI determination.

Received 26 October 2015 Returned for modification 1 December 2015

Accepted 14 January 2016

Accepted manuscript posted online 27 January 2016

Citation Zacour M, Ward BJ, Brewer A, Tang P, Boivin G, Li Y, Warhuus M, McNeil SA, LeBlanc JJ, Hatchette TF, on behalf of the Public Health Agency of Canada and Canadian Institutes of Health Influenza Research Network (PCIRN). 2016.

Standardization of hemagglutination inhibition assay for influenza serology allows for high reproducibility between laboratories. *Clin Vaccine Immunol* 23:236–242.

doi:10.1128/CVI.00613-15.

Editor: R. L. Hodinka

Address correspondence to Todd F. Hatchette, todd.hatchette@cdha.nshealth.ca.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

**TABLE 1** Reproducibility of HAI titers is high within laboratories

Laboratory	Intra-assay precision <sup>a</sup> [no. of samples with results differing >2-fold from duplicate-sample results/total no. of sample pairs (%)]			Interassay reproducibility <sup>b</sup> [no. of assays with results differing >2-fold from repeat assay results/total no. of sample assays (%)]			Median interassay % GCV [minimum–maximum]
	H1N1	H3N2	All viruses	H1N1	H3N2	All viruses	
A	NR	NR	NR	0/40 (0)	0/40 (0)	0/80 (0)	41.4 [0–49.2]
B	0/60 (0)	0/60 (0)	0/120 (0)	0/60 (0)	0/60 (0)	0/120 (0)	22.7 [0–43]
C	0/60 (0)	ND	0/60 (0)	4/60 (6.7)	ND	4/60 (6.7)	40.6 [0–68.5]
D	0/60 (0)	1/60 (1.7)	1/120 (0.8)	2/60 (3.3)	5/60 (8.3)	7/120 (5.8)	18.1 [0–91.2]
E	0/60 (0)	0/60 (0)	0/120 (0)	0/60 (0)	0/60 (0)	0/120 (0)	0 [0–46.2]
Overall	0/240 (0)	1/180 (0.6)	1/420 (0.2)	6/280 (2.1)	5/220 (2.3)	11/500 (2.2)	22.7 [0–91.2]

<sup>a</sup> NR, single replicates tested; ND, no data (laboratory C analyzed only H1N1 samples and was excluded from the H3N2 analyses).

<sup>b</sup> A total of 10 samples/virus were analyzed in 6 assays by laboratories B to E to give a total of 60 sample assays/virus; laboratory A returned data from just 4 assays.

**Influenza viruses.** Influenza A viruses included H1N1 (California-like) and H3N2 (Perth-like) viruses, provided by the National Microbiology Laboratory in Winnipeg, Manitoba, Canada. Working stocks were grown at each test site in MDCK cells and maintained using standard procedures. Viral stocks were quantified by analysis of hemagglutination activity (HA) against (final) 0.25% turkey RBCs (Lampire Biologicals, Pipersville, PA) and adjusted to 4 HA units in the HAI assay.

**HAI assay.** Participating laboratories were requested to have a single experienced operator perform six independent HAI assays per virus/sample panel, with each assay run in duplicate on a separate day. All sites followed a common HAI assay protocol (19), including interpretative criteria, and reagents were obtained from common suppliers when possible. To remove nonspecific inhibitors of HA, sera were incubated at 37°C overnight ( $19 \pm 1$  h) at a 1:4 dilution with receptor-destroying enzyme (RDE; Denka Seiken, Tokyo, Japan) followed by a 30-min inactivation step at 56°C and further dilution to 1:10 with phosphate-buffered saline (PBS). Potential nonspecific HA activity in sera was verified using 0.25% turkey RBCs. If present, it was removed by hemadsorption against (final concentration) 4.5% turkey RBCs for 1 h at 4°C. Just prior to HAI assay, the virus inoculum was back-titrated to verify the accuracy of the HA units (variability of  $\pm 0.5$  units was accepted; otherwise, viral stocks were dilution adjusted), and then 2-fold serial dilutions (starting from 1:10) of 25 µl of RDE-inactivated sera in V-well microtiter plates were incubated for 30 min at room temperature (RT) with 25 µl of 4 HA units of virus. A 50-µl volume of 0.5% turkey RBCs was added, and the reaction mixture was incubated for a further 30 min at RT or until cells in the RBC control wells had fully settled. Wells were examined visually for inhibition of HA, as indicated by the appearance of well-defined RBC “buttons” or teardrop formation upon plate tilting. HAI titers were the reciprocal of the highest dilution of serum that completely prevented HA.

**Data analysis.** Results from all laboratory sites were compiled and analyzed at an independent site (M. Zacour). Data consisted of replicate titers expressed as the reciprocal of serum dilutions. Titers of <10 (below the quantitation range) were assigned a value of 5 for calculations. Similarly to previous studies (5, 6, 9, 11), results for each sample were compared within and between laboratories, with titers considered equivalent if they varied by no more than a single dilution (i.e., 2-fold). Endpoint ratios between comparators were classified in contingency tables according to whether they were within or exceeded a 2-fold difference in value, with the significance of differences assessed by the chi-square ( $\chi^2$ ) test, using Prism 4 statistics software (GraphPad Software Inc., San Diego).

Precision (or reproducibility) was defined as the percentage of samples in a given data set yielding results within 2-fold of the results from the comparators. Intra-assay precision was assessed using two aliquots of the same sample in the same assay. For interassay precision, comparators were the geometric mean titers (GMTs) determined at a laboratory on each repeated assay day. Since specimens were patient sera of unknown “true” values, the interlaboratory reproducibility assessment compared

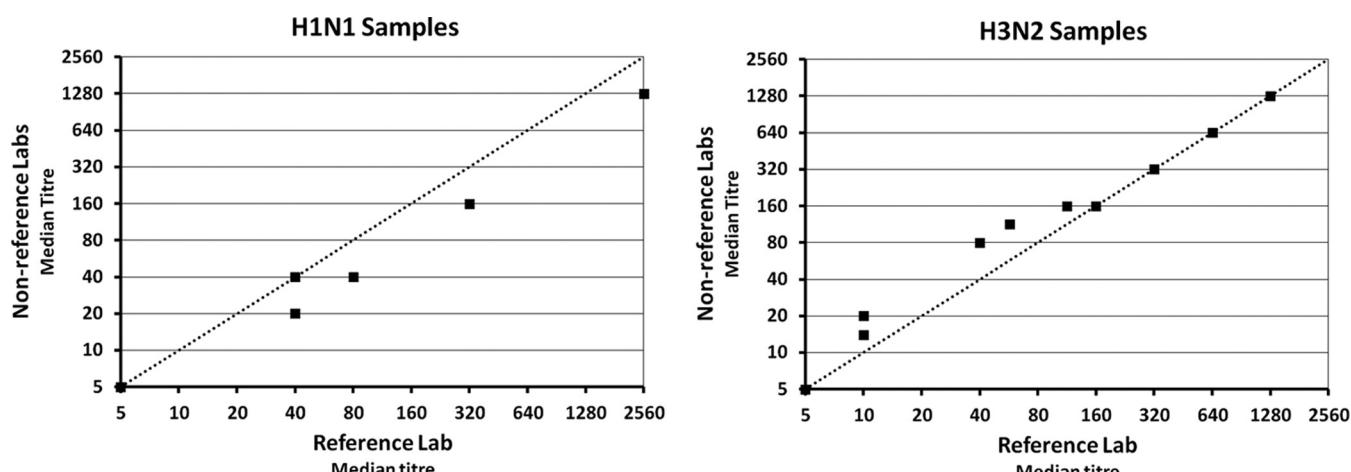
the titers determined for each sample at different laboratory sites (medians or GMTs over repeated assays) both to the titers determined at the network’s reference laboratory (laboratory E) and to consensus titers. Consensus titers represented the GMTs across all participating laboratories for a given sample. Interassay and interlaboratory titer variations were also quantified by determination of the geometric coefficient of variation (GCV [5]).

Seroprotection (defined as titers of  $\geq 40$ ) was classified on a per-sample, per-assay basis at each laboratory, and diagnostic accuracy was expressed as the percentage of agreement of the results from each PCIRN laboratory with the reference laboratory or consensus results. Reference or consensus seroprotection status was considered to be positive for a sample when  $>50\%$  of its assays produced a titer of  $\geq 40$ . Similarly, assay specificity (percentage of seronegative samples found seronegative by PCIRN laboratories) and sensitivity (percentage of seropositive samples found seropositive by PCIRN laboratories) were based on titers of  $<10$  defining seronegativity and titers of  $\geq 10$  defining seropositivity, respectively.

## RESULTS

**Intra-assay precision within laboratories.** Laboratory A was excluded from this assessment since it performed assays using a single replicate of each sample. Precision of duplicates ranged from 99.2% to 100% over the four participating laboratories, with an overall precision of 99.8% ( $n = 420$  duplicate pairs, assayed over 42 separate assays). There was only one instance where duplicate pairs differed by more than 2-fold (Table 1). The intra-assay precision did not differ between PCIRN laboratories ( $\chi^2 = 2.5$ ,  $P = 0.47$ ).

**Interassay precision within laboratories.** Overall, of a total of 90 sample aliquots that were compared over different assays within all laboratories, eight (8.9%) varied by greater than a doubling titer in at least one of the repeated assays at the same laboratory. Using a stringent definition of interassay precision as being the percentage of samples demonstrating no more than a 2-fold variation over all six assay repeats, the five sites ranged from 60% to 100%. Three laboratories (laboratories A, B, and E) demonstrated 100% precision over all samples and all assays, and the other two (laboratories C and D) showed variability beyond a doubling in titer in at least one of the six repeated assays, for 4/10 and 4/20 samples tested, respectively ( $\chi^2$  between sites = 20.85,  $P = 0.0003$ ). Although statistically significant, such variability from the 2-fold limit was infrequent, occurring in just 6.7% and 5.8% of assays at sites C and D, respectively (Table 1); hence, the interassay reproducibility in laboratories ranged from 93.3%



**FIG 1** Of samples assayed in nonreference laboratories, 100% showed median titers that were within 2-fold of the reference laboratory titer. Ten sera were tested per virus. Reference laboratory titers represented the median of 12 HAI determinations. Nonreference laboratory titers represented the median of 40 (H1N1; left panel) or 28 (H3N2; right panel) determinations. Dotted lines show the line of equality.

to 100% (97.8% over all laboratories). Median values for GCV between titers from each assay were also modest in laboratories C and D (40.6% and 18.1%, respectively; **Table 1**). The GCV provides a signal size-normalized measure of variability between the endpoint titers, with 0% GCV indicating identical titers over all comparisons. Since variations of 2-fold are not considered different, even laboratories with 100% precision had appreciable GCVs in some cases (e.g., 49.2% maximum GCV between assays; **Table 1**). Laboratory E had the lowest median GCV of all laboratories (0%), reflecting its central tendency to produce identical titer results over repeated assays of the same sample. In laboratory E, 60% of the samples assayed demonstrated 0% GCV between all repeated assays (compared to 25% to 40% of samples at the other four laboratories). Even the most variable sample at laboratory E showed a GCV of 46.2%, compared to 43% to 91.2% in other laboratories (**Table 1**). Laboratory E's consistent and precise HAI assay performance provides the rationale for its designation as the reference laboratory in our analyses.

**Interlaboratory reproducibility.** Reproducibility between laboratories was assessed by three different methods.

(i) **Comparison of median titers over all replicates and assays: reference laboratory versus other laboratories.** The median titer of all assays and replicates across nonreference laboratories was within 2-fold of the reference laboratory titer for all samples (**Fig. 1**). Compared to the reference laboratory (laboratory E), other laboratories demonstrated an overall tendency to underestimate H1N1 titers and overestimate low-to-medium-range H3N2 titers; however, the magnitude of these differences did not exceed 2-fold (the mean fold differences were 1.4 and 1.3 for H1N1 and H3N2 samples, respectively). This outcome of 100% reproducibility in comparisons between pooled data from the nonreference and reference laboratories remained consistent when modes (data not shown) or GMts (**Table 2**) were used instead of medians.

(ii) **Comparison of each laboratory's GMts to the reference laboratory GMts and consensus titers.** These comparisons allow identification of potential variability on a per-laboratory basis and mimic typical approaches used in interlaboratory proficiency testing. Overall, 94.3% of the titers determined for the 70 samples

assayed at laboratories A to D were within 2-fold of the laboratory E titers, with reproducibility in individual laboratories ranging from 85% to 100% (**Table 3**) ( $\chi^2$  between laboratories = 5.038,  $P = 0.1690$ ). Similarly, titers of 98.9% of the 90 samples assayed at laboratories A to E were within 2-fold of the consensus titers and ranged from 95% to 100% in individual laboratories (**Table 3**) ( $\chi^2$  between laboratories = 3.549,  $P = 0.47$ ). As seen in **Fig. 2**, results from all individual samples suggested systematic bias in most laboratories (as illustrated with downshifting or upshifting of all sample titers at a given laboratory, relative to the 100% consensus line); however, the degree of bias was modest, with only one sample at one laboratory showing an endpoint titer that exceeded a 2-fold difference from the consensus titer.

(iii) **Magnitude of titer variability across laboratories.** The variation between the titers across all laboratories is quantified by the per-sample GCVs shown in **Table 2**. The median GCV between laboratories was 26.9% (range, 0% to 156.4%). As assessed by influenza A virus subtype, the median GCV for H1N1 samples was 38.8% (range, 0% to 156.4% across five laboratories); for H3N2 samples, it was 17.2% (range, 0% to 75.5% across four laboratories; one laboratory did not perform assays for this virus subtype). The GCV between laboratories tended to be highest at lower titers (**Fig. 3**). The single sample with titer values that differed from the consensus titers by more than 2-fold in one laboratory produced the highest interlaboratory GCV of 156.4%. The GMT for that sample ranged from 20 to 57 across four laboratories but was below the quantitation range (i.e., was assigned a value of 5) at the fifth laboratory, resulting in an 11-fold difference from lowest to highest titer across laboratories. No other sample approached this level of variability between laboratories. In fact, 70% of all samples showed differences of  $\leq 2$ -fold from lowest to highest GMT across all laboratories; 90% showed differences that were less than 4-fold, and 95% showed  $\leq 4$ -fold differences.

**Diagnostic accuracy.** Using an HAI titer of  $\geq 40$  as a serologic correlate of protection for healthy adults (1, 4, 20, 21),  $n = 500$  HAI results were classified for seroprotection over all samples, assays, and laboratories. Of the sera, 74.4% showed perfect concordance, being classified identically over every assay at every laboratory (including 70% of H1N1 and 80% of H3N2 samples).

TABLE 2 Sample titers over all replicates and assays<sup>a</sup>

Virus	Sample no.	Geometric mean titer (median titer)								GCV between laboratories (%)
		Consensus	Reference laboratory E	All nonreference laboratories	Laboratory B	Laboratory C	Laboratory D	Laboratory A		
H1N1	H1-1	5	5 (5)	5 (5)	6	5	5	5	10.9	
	H1-7	5	5 (5)	5 (5)	5	5	5	5	0.0	
	H1-4	6	6 (5)	6 (5)	6	6	5	5	10.9	
	H1-5	22	40 (40)	19 (20)	18	22	13	28	58.3	
	H1-10	24	36 (40)	21 (20)	5	38	20	57	156.4	
	H1-8	43	40 (40)	44 (40)	24	57	40	67	49.5	
	H1-2	64	80 (80)	61 (40)	40	80	45	95	49.5	
	H1-6	237	285 (320)	226 (160)	170	254	226	269	25.7	
	H1-9	234	254 (320)	230 (160)	160	240	160	453	53.1	
	H1-3	1,325	2,032 (2,560)	1,191 (1,280)	959	1,280	1,280	1,280	28.0	
H3N2	H3-3	5	5 (5)	5 (5)	6	ND	5	5	5.9	
	H3-8	11	10 (10)	11 (14)	17	ND	12	7	43.4	
	H3-6	12	10 (10)	13 (20)	27	ND	13	7	75.5	
	H3-9	50	40 (40)	54 (80)	101	ND	48	34	62.1	
	H3-1	92	57 (57)	109 (113)	135	ND	101	95	43.4	
	H3-5	123	113 (113)	127 (160)	160	ND	113	113	18.9	
	H3-10	174	160 (160)	180 (160)	151	ND	202	190	14.7	
	H3-4	293	285 (320)	296 (320)	302	ND	320	269	11.1	
	H3-2	772	640 (640)	822 (640)	761	ND	806	905	15.5	
	H3-7	1,280	1,280 (1,280)	1,280 (1,280)	1,280	ND	1,280	1,280	0.0	

<sup>a</sup> Laboratories E, B, and D,  $n = 12$  per sample; laboratory C,  $n = 12$  per H1N1 sample; laboratory A,  $n = 4$  per sample; nonreference laboratories together,  $n = 40$  per H1N1 sample and  $n = 28$  per H3N2 sample; consensus titer,  $n = 52$  per H1N1 sample and  $n = 40$  per H3N2 sample; ND, no data.

When diagnostic accuracy was defined as the percentage of titers from all individual assay results that agreed with the reference laboratory classification for seroprotection, accuracy was 89.5% over all samples (98.1% [157/160] for H3N2 samples and 83.2% [183/220] for H1N1 samples). When the consensus classification was used to define a “true” determination, accuracy was somewhat higher: 93.6% over all samples (98.6% [217/220] and 89.6% [251/280] for H3N2 and H1N1 samples, respectively). Differences between reference laboratory and consensus classifications of seroprotection occurred with only two H1N1 samples, corresponding to consensus titers of 22 and 24.

Sixteen of the 20 samples were seropositive and 4 were seronegative, by both reference laboratory and consensus determinations. The four negative samples were also classified as seronega-

tive at all individual laboratories, indicating 100% assay specificity across all PCIRN laboratories. Sensitivity (defined here as the percentage of seropositive samples that were identified as seropositive at a laboratory) was 100% at laboratories C, D, and E and ranged from 87.5% to 93.8% over all samples at laboratories A and B. Considered on a per-virus basis, sensitivity ranged from 90% to 100% and from 80% to 100% across laboratories for H1N1 and H3N2, respectively.

## DISCUSSION

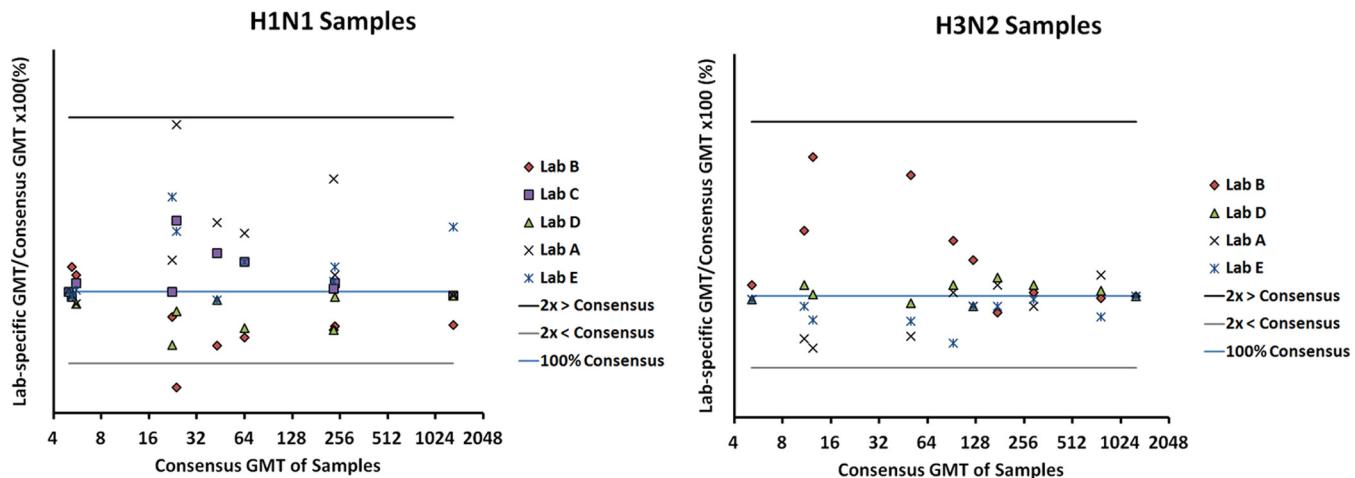
Our results show that, by careful planning, training, and standardization of reagents and protocols, a transnational laboratory network has the capacity to produce consistent and comparable HAI assay results. The five public health and academic laboratories participating in this study were part of PCIRN, which was developed to increase Canada’s national capacity to support vaccine and other influenza research in the event of a pandemic. In a pandemic situation, with high-volume, high-pressure, and geographically dispersed influenza testing, standardization of the HAI assay will be of paramount importance to ensure comparability of results between laboratories. The PCIRN laboratories showed a high standard of within-laboratory reproducibility, with 99.8% precision of duplicates (99.2% to 100% per laboratory), 97.8% of titers within a 2-fold range over repeated assay days (93.3% to 100% per laboratory), and 91.1% of the test sera showing titer reproducibility over every single repeat of the six-assay protocol (60% to 100% per laboratory). Interlaboratory reproducibility was also high, with median titers across all replicates from the four nonreference laboratories within 2-fold of the reference laboratory for 100% of the sera tested. On a per-laboratory basis, the titers of 94.3% and 98.9% of samples quantified within a single dilution of the reference laboratory and consensus titers,

TABLE 3 Reproducibility of HAI titers is high between laboratories

Laboratory	Interlaboratory reproducibility		
	No. of samples with GMT differing >2-fold from reference (laboratory E) GMT/total no. of samples (%)	No. of samples with GMT differing >2-fold from consensus GMT/total no. of samples (%)	
H1N1			
A	0/10 (0)	0/10 (0)	0/20 (0)
B	1/10 (10)	2/10 (20)	3/20 (15)
C	0/10 (0)	ND <sup>b</sup>	0/10 (0)
D	1/10 (10)	0/10 (0)	1/20 (5)
E	NA <sup>a</sup>	NA	NA
Overall	2/40 (5)	2/30 (6.7)	4/70 (5.7)
H3N2			
A	0/10 (0)	0/10 (0)	0/20 (0)
B	1/10 (10)	2/10 (20)	3/20 (15)
C	0/10 (0)	ND	0/10 (0)
D	1/10 (10)	0/10 (0)	0/20 (0)
E	NA	NA	NA
Overall	0/40 (0)	0/40 (0)	0/20 (0)
All viruses			
A	0/10 (0)	0/10 (0)	0/20 (0)
B	1/10 (10)	2/10 (20)	3/20 (15)
C	0/10 (0)	ND	0/10 (0)
D	1/10 (10)	0/10 (0)	0/20 (0)
E	NA	NA	NA
Overall	1/50 (2)	0/40 (0)	1/90 (1.1)

<sup>a</sup> NA, not applicable.

<sup>b</sup> ND, no data.



**FIG 2** Of 10 samples per virus assayed at each laboratory, the titers of all but 1 at one laboratory were within 2-fold of the corresponding consensus titers. Titers per sample, per laboratory, were expressed as a percentage of the sample's consensus titer (vertical axes) and plotted against the  $\log_2$ -scaled titers on the horizontal axes for the H1N1 (left panel) and H3N2 (right panel) samples.

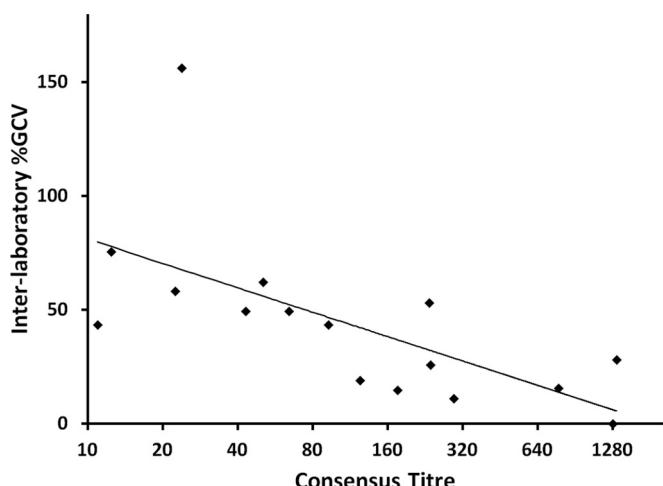
respectively (ranges, 85% to 100% and 95% to 100%). These percentages meet typical acceptability thresholds for HAI assay variability (i.e., at least 80% of samples being within 1 dilution of nominal values and/or all-assay GMTs [22, 23]). HAI assay specificity (100%) and sensitivity (87.5% to 100%) across PCIRN laboratories were similarly well within the acceptance criteria of previous studies (i.e., 100% of negative samples should quantify as below levels of quantitation, and at least 80% of positive samples should quantify as positive [22, 23]).

For the intralaboratory component of HAI assay variability, repeatability within PCIRN laboratories appears similar to that shown in other studies (5, 22). One caveat of these previous studies was that no more than three repeats were examined, despite common immune assay validation practices of examining six or more repeated assays (24–27). As such, it is not clear if their data are directly comparable to the results determined over six assays in this study. Nonetheless, our finding of 0% to 6.7% (median, 0%)

of assays with results differing by more than 2-fold from those of other assay repeats seem comparable to the 4.8% to 7.1% noted previously in one laboratory (22) and the 0% to 15% (median, 2%) in nine laboratories (5).

Our findings of high HAI titer reproducibility between geographically dispersed PCIRN laboratories contrast with previous reports showing substantial variability between international laboratories performing this assay (5, 6, 9–12). For example, data from a large collaborative study (11 laboratories) found per-sample GCVs ranging from 84% to 803% (median GCV per virus of 138%, 155%, and 261% [5]). Other studies reported GCVs for H1N1 and H5N1 samples that spanned 95% to 345% (11) and 22% to 582% (6), and GMTs differed between laboratories by up to 80-fold (11) or 128-fold (9). While previous reports often included higher numbers of participating laboratories than the current study, even when just six laboratories contributed data, median interlaboratory GCV was 83% for five H1N1 samples, with 80% of them exceeding a 4-fold difference between laboratories (11). In contrast, the current study showed median GCVs per virus of only 39% and 17%, with 95% of samples within a 4-fold difference between laboratories. One sample exceeded the 4-fold difference; the titer was not quantifiable for this sample at one laboratory, and hence the lowest value was assigned rather than measured. If that sample were omitted from analyses, the maximum GCV would be 75.5%, which would represent a value lower than the minimum GCV in other studies. Although that sample showed higher variability than others, this low measurement range appeared in general less precise than higher ranges. For example, GCV ranged from 43.4% to 156.4% at titers between 10 and 80 compared to 11.1% to 53.1% at higher titers (Fig. 3). Greater variability with lower titer samples has also been previously described (23) and likely contributed to incongruent seroprotection classifications in the current study for the three H1N1 samples with titers closest to the threshold of 40 (consensus titers of 22, 24, and 43). Despite this, diagnostic accuracy of seroprotection across PCIRN laboratories remained high overall (93.6% when consensus titers were used to define seroprotection and 89.5% compared to reference laboratory values).

High agreement of interlaboratory HAI titers in PCIRN labo-



**FIG 3** Variability between laboratories in titer estimations (expressed as GCV on the y axis) tends to be higher for samples in low measurement ranges and to decrease as the measurement range increases;  $n = 16$  (4 samples that were below the lower limit of titer quantitation were excluded).

ratories may relate to the careful harmonization of HAI assay procedures and reagents across the PCIRN. Participants in previous studies that showed high interlaboratory variability generally used HAI assay methods that varied between laboratories. Assay variables thought to contribute to variability in results included the amount of virus added, the viral culture system (e.g., embryonated chicken eggs versus tissue culture), the duration of RDE treatment, the initial serum dilution, the type and age of the RBCs used, and the time allowed for RBC settling (5, 11, 12). Surprisingly, however, when Wood et al. provided commercial and public laboratory participants involved in seasonal influenza vaccine testing in Europe with a detailed standard operating procedure to follow and common reagents/materials, they reported that interlaboratory variation was not significantly improved (10). The authors speculated that persistent “local technical variations,” as well as different RBC suppliers, might have affected reproducibility between laboratories but noted that further standardization was beyond the limits of feasibility for their study. Later publications interpreted these findings as indicating that HAI assay standardization between laboratories is, in general, not a feasible approach for harmonizing HAI results across different laboratories (8, 12), and at least one group has tried to develop a modified and more robust HAI assay to generate more reliable and reproducible data (22).

In the current study, the PCIRN laboratories adhered strictly to the standard, publicly available World Health Organization HAI assay protocol, in which points subject to technical or interpretative variations were harmonized between laboratories prior to study commencement. Key reagents (e.g., RDE, RBC, virus stocks) were obtained from the same suppliers as much as possible. In addition, restricting the testing to a single technologist at each test site permitted increased standardization. The possible effect of using multiple personnel at each laboratory was not tested in this study. While the use of multiple operators could potentially increase variability, the only study that examined this possibility for HAI testing did not show any significant difference between the results from two operators in the same laboratory (22).

Potential variations between laboratories in the viral strains grown for use in the assay were minimized between PCIRN laboratories by using viral seed stocks from a single source that were grown in the same cell line under similar conditions at all laboratories. We speculate that interlaboratory reproducibility might be amenable to even further improvements were viral working stocks to be produced centrally and distributed to participating laboratories. The importance of viral strain variations for HAI endpoint titers is well illustrated by the 2-to-3-fold-lower titers demonstrated in one study when the HAI assay used an A/Cal pH1N1 strain [A(H1N1)pdm09] compared to the X179A reassortant that still possessed equivalent hemagglutinins and neuraminidases (11). Unfortunately, stocks were not sequenced as a part of this study, and so we cannot verify that all laboratories used viruses with identical hemagglutinin sequences. Even if minor sequence variations had been introduced by local propagation of virus stocks, the variability between laboratories in this study was small, and so it is hard to imagine how the use of stocks with identical HA sequences might further reduce variability. Although the differences did not exceed 2-fold in this small sample set, minor variations in HA sequence introduced by local propagation of viral

stocks could be one possible explanation for the trends to underestimation of H1N1 titers and overestimation of H3N2 titers.

Various studies have reported reduced interlaboratory variability in HAI titers by calibrating results against international standards (6, 11–13); however, such standards do not exist for the current seasonal influenza B and influenza A H3N2 virus strains, and the constant antigenic drift of seasonal strains that drive yearly reformulation of the vaccines may well require the continuous development and validation of new standards. In this study, the interlaboratory comparability appears superior to that achieved through use of even the most effective calibration standards. For example, following calibration with IS 09/194, Wood et al. (11) reported that maximal differences between GMTs at different laboratories diminished from 80-fold to 64-fold (i.e., a value still appreciably higher than the maximal 11-fold in the current study) and that ranges of GCV between laboratories diminished from 95% to 345% (median, 105%) to 34% to 231% (median, 109%)—i.e., well over the range of 0% to 156% (median, 27%) seen in this study. Similarly, the use of replacement IS 10/202 diminished the GCV between laboratories from a range of 108% to 157% to a range of 24% to 144% across five samples and maximal differences between laboratories from 43-fold to 21-fold (13), and, using the 07/150 standard, Stephenson et al. (6) reduced the median GCV of 15 NIBRG-14 H5N1 clade 1 samples from 125% (range, 31% to 582%) to 61% (34% to 535%).

While the exact reason(s) for the high level of agreement between the PCIRN laboratories compared to previous studies cannot be ascertained, our success was facilitated by the incentive that we had as a national network to work toward the common goal of establishing reproducible processes. Detailed and rigorous process and reagent harmonization between sites is perhaps the only “new” aspect we brought to this study, compared to the long-standing body of literature showing poor HAI reproducibility between laboratories. However, our data suggest that with careful standardization, interlaboratory variations in HAI titers may be reduced to levels similar to those observed within single laboratories. Potential next steps in minimizing variations in HAI titers in laboratories worldwide would be the mobilization of networks/consortia, the implementation of guidelines and/or regulatory requirements for the use of standard procedures and reagents, standardized training to ensure reproducibility, and the provision of support and/or incentives for the laboratories to embrace the “standardization” concept.

## ACKNOWLEDGMENTS

We thank Melanie Wilcox, May Elsherif, Chantal Rheaume, and Suzana Sabaiduc for technical support to complete the project.

We have no conflicts of interest to declare.

M.Z., B.J.W., P.T., G.B., Y.L., S.A.M., J.J.L., and T.F.H. were involved in the conception and design of the study; M.W. and A.B. were responsible for creation and coordination of specimen panels, specimen testing, and collation of the data; B.J.W., P.T., G.B., Y.L., S.A.M., J.J.L., and T.F.H. supervised the PCIRN laboratories; M.Z., T.F.H., and B.J.W. analyzed and interpreted the data; M.Z. and T.F.H. drafted the manuscript; all authors revised the manuscript critically for important intellectual content; and all authors reviewed and approved the final draft of the manuscript.

## FUNDING INFORMATION

Through PCIRN, Public Health Agency of Canada provided funding to Brian J. Ward, Guy Boivin, Yan Li, Shelly A. McNeil, Jason Leblanc, and Todd F. Hatchette. Gouvernement du Canada | Canadian Institutes of

Health Research (CIHR) provided funding to Brian J. Ward, Guy Boivin, Yan Li, Shelly A. McNeil, Jason Leblanc, and Todd F. Hatchette. The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

- Food and Drug Administration. 2007. Guidance for industry: clinical data needed to support the licensure of seasonal inactivated influenza vaccines. <http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/Vaccines/ucm074794.htm>.
- Skowronski DM, Hottes TS, Janjua NZ, Purych D, Sabaiduc S, Chan T, De Serres G, Gardy J, McElhaney JE, Patrick DM, Petric M. 2010. Prevalence of seroprotection against the pandemic (H1N1) virus after the 2009 pandemic. *CMAJ* 182:1851–1856. <http://dx.doi.org/10.1503/cmaj.109010>.
- Khan SU, Anderson BD, Heil GL, Liang S, Gray GC. 2015. A systematic review and meta-analysis of the seroprevalence of influenza A (H9N2) infection among humans. *J Infect Dis* 212:562–569. <http://dx.doi.org/10.1093/infdis/jiv109>.
- Reber A, Katz J. 2013. Immunological assessment of influenza vaccines and immune correlates of protection. *Expert Rev Vaccines* 12:519–536. <http://dx.doi.org/10.1586/erv.13.35>.
- Stephenson I, Gaines Das R, Wood JM, Katz JM. 2007. Comparison of neutralising antibody assays for detection of antibody to influenza A/H3N2 viruses: an international collaborative study. *Vaccine* 25:4056–4063. <http://dx.doi.org/10.1016/j.vaccine.2007.02.039>.
- Stephenson I, Heath A, Major D, Newman RW, Höschler K, Junzi W, Katz JM, Zambon MC, Weir JP, Wood JM. 2009. Reproducibility of serologic assays for influenza virus A (H5N1). *Emerg Infect Dis* 15:1252–1259.
- Wood JM, Gaines-Das RE, Taylor J, Chakraverty P. 1994. Comparison of influenza serological techniques by international collaborative study. *Vaccine* 12:167–175. [http://dx.doi.org/10.1016/0264-410X\(94\)90056-6](http://dx.doi.org/10.1016/0264-410X(94)90056-6).
- Girard MP, Katz JM, Pervikov Y, Hombach J, Tam JS. 2011. Report of the 7th meeting on evaluation of pandemic influenza vaccines in clinical trials, World Health Organization, Geneva, 17–18 February 2011. *Vaccine* 29:7579–7586. <http://dx.doi.org/10.1016/j.vaccine.2011.08.031>.
- Wood JM, Newman RW, Daas A, Terao E, Buchheit KH. 2011. Collaborative study on influenza vaccine clinical trial serology—part 1: CHMP compliance study. *Pharmer Bio Sci Notes* 2011:27–35.
- Wood JM, Montomoli E, Newman RW, Daas A, Buchheit KH, Terao E. 2011. Collaborative study on influenza vaccine clinical trial serology—part 2: reproducibility study. *Pharmer Bio Sci Notes* 2011:36–54.
- Wood JM, Major D, Heath A, Newman RW, Höschler K, Stephenson I, Clark T, Katz JM, Zambon MC. 2012. Reproducibility of serology assays for pandemic influenza H1N1: collaborative study to evaluate a candidate WHO International Standard. *Vaccine* 30:210–217. <http://dx.doi.org/10.1016/j.vaccine.2011.11.019>.
- Wagner R, Göpfert C, Hammann J, Neumann B, Wood J, Newman R, Wallis C, Alex N, Pfleiderer M. 2012. Enhancing the reproducibility of serological methods used to evaluate immunogenicity of pandemic H1N1 influenza vaccines—an effective EU regulatory approach. *Vaccine* 30: 4113–4122. <http://dx.doi.org/10.1016/j.vaccine.2012.02.077>.
- Major D, Heath A, Wood J. 2012. Report of a WHO collaborative study to assess the suitability of a candidate replacement international standard for antibody to pandemic H1N1 influenza virus. World Health Organization WHO/BS/2012.2190. [http://www.who.int/biologicals/expert\\_committee/BS\\_2190\\_H1N1\\_IS\\_replacement\\_report\\_final\\_10\\_July\\_2012\(2\).pdf](http://www.who.int/biologicals/expert_committee/BS_2190_H1N1_IS_replacement_report_final_10_July_2012(2).pdf). Accessed 8 August 2015.
- National Institute for Biological Standards and Control (NIBSC). 2009. International Standard for antibody to influenza H5N1 virus; NIBSC code: 07/150; instructions for use. <http://www.nibsc.org/documents/ifu/07-150.pdf>.
- National Institute for Biological Standards and Control (NIBSC). 2012. International Standard for antibody to influenza H1N1 pdm virus; NIBSC code: 09/194; instructions for use. <http://www.nibsc.org/documents/ifu/09-194.pdf>.
- Van Kerkhove MD, Broberg E, Engelhardt OG, Wood J, Nicoll A; CONSISE steering committee. 2013. The consortium for the standardization of influenza seroepidemiology (CONSISE): a global partnership to standardize influenza seroepidemiology and develop influenza investigation protocols to inform public health policy. *Influenza Other Respir Viruses* 7:231–234. <http://dx.doi.org/10.1111/irv.12068>.
- Van Kerkhove M, Wood J; CONSISE. 2014. Conclusions of the fourth CONSISE international meeting. *Euro Surveill* 19:pii=20668. <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20668>.
- Laurie KL, Engelhardt OG, Wood J, Heath A, Katz JM, Peiris M, Hoschler K, Hungnes O, Zhang W, Van Kerkhove MD. 2015. International laboratory comparison of influenza microneutralization assay protocols for A(H1N1)pdm09, A(H3N2), and A(H5N1) influenza viruses by CONSISE. *Clin Vaccine Immunol* 22:957–964. <http://dx.doi.org/10.1128/CVI.00278-15>.
- WHO Global Influenza Surveillance Network. 2011. Manual for the laboratory diagnosis and virological surveillance of influenza. World Health Organization. ISBN 978 9241548090:43-62. [http://whqlibdoc.who.int/publications/2011/9789241548090\\_eng.pdf](http://whqlibdoc.who.int/publications/2011/9789241548090_eng.pdf).
- de Jong JC, Palache AM, Beyer WE, Rimmelzwaan GF, Boon AC, Osterhaus AD. 2003. Haemagglutination-inhibiting antibody to influenza virus. *Dev Biol (Basel)* 115:63–73.
- Black S, Nicolay U, Vesikari T, Knuf M, Del Giudice G, Della Cioppa G, Tsai T, Clemens R, Rappuoli R. 2011. Hemagglutination inhibition antibody titers as a correlate of protection for inactivated influenza vaccines in children. *Pediatr Infect Dis J* 30:1081–1085. <http://dx.doi.org/10.1097/INF.0b013e3182367662>.
- Morokutti A, Redlberger-Fritz M, Nakowitz S, Krenn BM, Wressnigg N, Jungbauer A, Romanova J, Muster T, Popow-Kraupp T, Ferko B. 2013. Validation of the modified hemagglutination inhibition assay (mHAI), a robust and sensitive serological test for analysis of influenza virus-specific immune response. *J Clin Virol* 56:323–330. <http://dx.doi.org/10.1016/j.jcv.2012.12.002>.
- Noah DL, Hill H, Hines D, White L, Wolff MC. 2009. Qualification of the hemagglutination assay in support of pandemic influenza vaccine licensure. *Clin Vaccine Immunol* 16:558–566. <http://dx.doi.org/10.1128/CVI.00368-08>.
- DeSilva B, Smith W, Weiner R, Kelley M, Smolec J, Lee B, Khan M, Tacey R, Hill H, Celniker A. 2003. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res* 20:1885–1900. <http://dx.doi.org/10.1023/B:PHAM.0000003390.51761.3d>.
- Viswanathan CT, Bansal S, Booth B, DeStefano AJ, Rose MK, Sailstad J, Shah VP, Skelly JP, Swann PG, Weiner R. 2007. Workshop/conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *AAPS J* 9:E30–E42. <http://dx.doi.org/10.1208/aapsj0901004>.
- Findlay JWA, Smith WC, Lee JW, Nordbloom GD, Das I, DeSilva BS, Khan MN, Bowsher RR. 2000. Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J Pharm Biomed Anal* 21: 1249–1273. [http://dx.doi.org/10.1016/S0731-7085\(99\)00244-7](http://dx.doi.org/10.1016/S0731-7085(99)00244-7).
- Gijzen K, Liu WM, Visontai I, Oftung F, Van der Werf S, Korsvold GE, Pronk I, Aaberge IS, Tütö A, Jankovics I, Jankovics M, Gentleman B, McElhaney JE, Soethout EC. 2010. Standardization and validation of assays determining cellular immune responses against influenza. *Vaccine* 28:3416–3422. <http://dx.doi.org/10.1016/j.vaccine.2010.02.076>.