

Original Paper

Real-World Performance of COVID-19 Antigen Tests: Predictive Modeling and Laboratory-Based Validation

Miguel Bosch^{1,2}, PhD; Dawlyn Garcia², MSc; Lindsey Rudtner², BSc; Nol Salcedo², MSc; Raul Colmenares¹, MSc; Sina Hoche², PhD; Jose Arocha², MSc; Daniella Hall², PhD; Adriana Moreno¹, MSc; Irene Bosch², PhD

¹Info Analytics Innovations, Houston, TX, United States

²IDX20 Inc, Brookline, MA, United States

Corresponding Author:

Irene Bosch, PhD
IDX20 Inc
166 Clinton Rd
Brookline, MA 02445
United States
Phone: 1 6177347745
Email: ibosch@idx20.us

Related Articles:

Preprint (medRxiv): <https://www.medrxiv.org/content/10.1101/2024.10.21.24315762v1>

Peer-Review Report by Helena de Puig (Reviewer BH): <https://med.jmirx.org/2025/1/e83476>

Peer-Review Report by Gerald Kost (Reviewer FZ): <https://med.jmirx.org/2025/1/e83479>

Authors' Response to Peer-Review Reports: <https://med.jmirx.org/2025/1/e83474>

Abstract

Background: Rapid and safe deployment of lateral-flow antigen tests, coupled with uncompromised quality assurance, is critical for outbreak control and pandemic preparedness, yet real-world performance assessment still lacks laboratory and quantitative approaches that remain uncommon in current regulatory science. The approach proposed here can help standardize and accelerate early phase appraisal of antigen tests in preparation for clinical validation.

Objective: The aim of this study is to present a quantitative, laboratory-anchored framework that links image-based test line intensities and the population distribution of naked-eye limits of detection (LoD) to a probabilistic prediction of positive percent agreement (PPA) as a function of viral-load-related variables (eg, quantitative real-time polymerase chain reaction [qRT-PCR] cycle thresholds [Ct]). Using dilution-series calibrations and a Bayesian model, the predicted PPA-vs-Ct curve closely tracks the observed PPA in a real-world self-testing cohort.

Methods: The proposed methodology combines: (1) a quantitative evaluation of the test signal response to concentrations of target protein and inactive virus or active virus, (2) a statistical characterization of the LoD using the observer's visual acuity of the test band, and (3) a calibration of a gold-standard method (eg, qRT-PCR cycles) against virus concentration. We elaborate these quantitative methods and unfold a Bayesian-based predictive model to describe the real-world performance of the antigen test, quantified by the probability of positive agreement as a function of viral-load variables like qRT-PCR Ct.

Results: We applied the methodology by characterizing each brand of COVID-19 antigen test and estimating its real-world probability of agreement with qRT-PCR. We aligned protein and inactivated-virus standard curves at matched signal intensities and fit a linear calibration linking protein to viral concentrations. Using logistic regression, we modeled the PPA as a continuous function of qRT-PCR Ct, then integrated this curve over a predefined reference Ct distribution to obtain the expected sensitivity. This standardization enables consistent performance comparisons across sites.

Conclusions: Modeling performance under real-world conditions requires coupling laboratory evaluation with the population's ability to perceive the test's visual signal. We represent observer capability as a probability density function of the LoD over the signal-intensity domain. Rather than reporting bin-based sensitivity, we summarize performance with the PPA as a continuous function of qRT-PCR Ct. Our framework produces PPA-Ct curves by composing (1) normalized signal-to-concentration models from the laboratory, (2) the observer LoD distribution, and (3) a Ct-to-viral-load calibration. The resulting inferences are inherently context-bound—disease-, assay-, and setup-specific. External validity depends on

the particular antigen lateral-flow test, the user population (visual acuity and interpretation), and cross-laboratory qRT-PCR calibration. Comprehensive clinical studies under intended-use conditions are still required before making generalized claims.

Trial Registration: ClinicalTrials.gov NCT05884515; <https://clinicaltrials.gov/study/NCT05884515>

JMIRx Med 2025;6:e68376; doi: [10.2196/68376](https://doi.org/10.2196/68376)

Keywords: COVID-19; SARS-CoV-2 antigen test; lateral flow assay; point-of-care diagnostics; real-world performance; Langmuir–Freundlich isotherm; Bayesian regression (Monte Carlo); probability of positive agreement; limit of detection; image-based signal quantification

Introduction

Quantifying the performance of antigen lateral flow tests (Ag-LFT) according to the US regulatory science standards commonly requires the calculation of test performance statistics—for example, sensitivity or positive percent agreement (PPA) of an antigen test's (AT) binary assessments with reference to the quantitative real-time polymerase chain reaction (qRT-PCR) gold standard [1-3]. These statistics are based on human clinical samples, requiring paired qRT-PCR cycle thresholds (Cts) to generate performance data and corresponding test validations as described in a comprehensive literature review for COVID-19 studies [4]. The clinical performance of Ag-LFTs increases at higher viral loads (low Ct values), which present early on in the symptom window, and test performance declines with low viral loads (high Ct values) at the end of the acute disease window [2,5-10]. These realities motivate a fast, laboratory-anchored, model-based appraisal that anticipates PPA-vs-Ct before large trials. That is, AT performance is closely related to the sample viral load and the performance statistics are dependent on several factors such as viral load distribution, specific virus variants [8,9,11], and symptomatic versus asymptomatic cases [7,12], as well as the observer's training [6,13-17]. Hence, the regulatory process is typically a lengthy process and includes appropriate sample size and requires a viral load distribution to cover the spectrum of target concentrations.

Because Ag-LFTs are powerful tools for transmission control and epidemic mitigation, the fast and safe deployment of tests without compromising quality assurance in the evaluation process is key for outbreak control and pandemic preparedness. Home testing and easy access to Ag-LFTs enables them to be used for serial testing, as has been recommended. For COVID-19 Ag-LFTs, serial testing increases effective sensitivity [18], including in home testing [7,13,19].

We have developed a methodology for the quantitative evaluation of SARS-CoV-2 ATs based on laboratory measurements of the regions of interest (ROIs), including the test's regions and corresponding normalized signal intensity and binary naked-eye user assessments for positive or negative results. In both cases, we characterized the test performance according to the sample concentration of the target recombinant protein, with heat-inactivated virus as well as biologically active virus, and we used human samples self-collected 2 times per week, under a prospective clinical protocol (ClinicalTrials.gov: NCT05884515). To support accurate self-reporting, participants received a brief,

standardized orientation with visual aids on test interpretation, photo capture, and upload procedures. Including the statistical characterization of the user population's limit of detection (LoD) in the signal intensity domain, we developed a predictive model for the probability of positive agreement in real-world conditions.

Our method involves (1) characterizing the AT signal intensity with protein and inactivated virus dilutions, (2) calibrating the qRT-PCR cycles with virus dilutions, (3) characterizing the signal intensity LoD of the user population for the AT, and (4) predicting the real-world probability of a positive agreement signal response of the AT. Our methods have the advantage of being formulated using continuous variable analysis and probability models instead of plain discrete analysis and sample statistics.

We demonstrate our methodology capabilities when comparing the predicted probability of positive agreement with that generated using real-world data collected through an institutional review board (IRB)-approved study for frequent antigen testing to monitor COVID-19 in an underserved population (ClinicalTrials.gov: NCT05884515). Participants consisted of individuals from vulnerable populations in low-income and assisted-living facilities located in the city of Chelsea, MA. Recruitment included people living in state-regulated, independent senior living communities and other residents of Chelsea. The consented participants were provided with ATs to routinely self-test for COVID-19 at home or in community centers, 2 times per week, uploading the test results and photos to the project informatics platform. We obtained confirmatory qRT-PCR data for all positive results detected by the home AT and for a random number of negative results from an independent Clinical Laboratory Improvement Amendments laboratory. The certified Clinical Laboratory Improvement Amendments laboratory procedures were approved by the IRB and the qRT-PCR data shared included the Ct values for each submitted test.

We describe the quantitative analysis of the AT for signal intensity and naked-eye binary data, the characterization of the user's LoD, the calibration of the qRT-PCR, and the formulation of the predictive model. In the Results section, we illustrate the application of the described methodology with the characterization of an AT in the common cassette device presentation and compare the predicted result with the real-world probability of positive agreement.

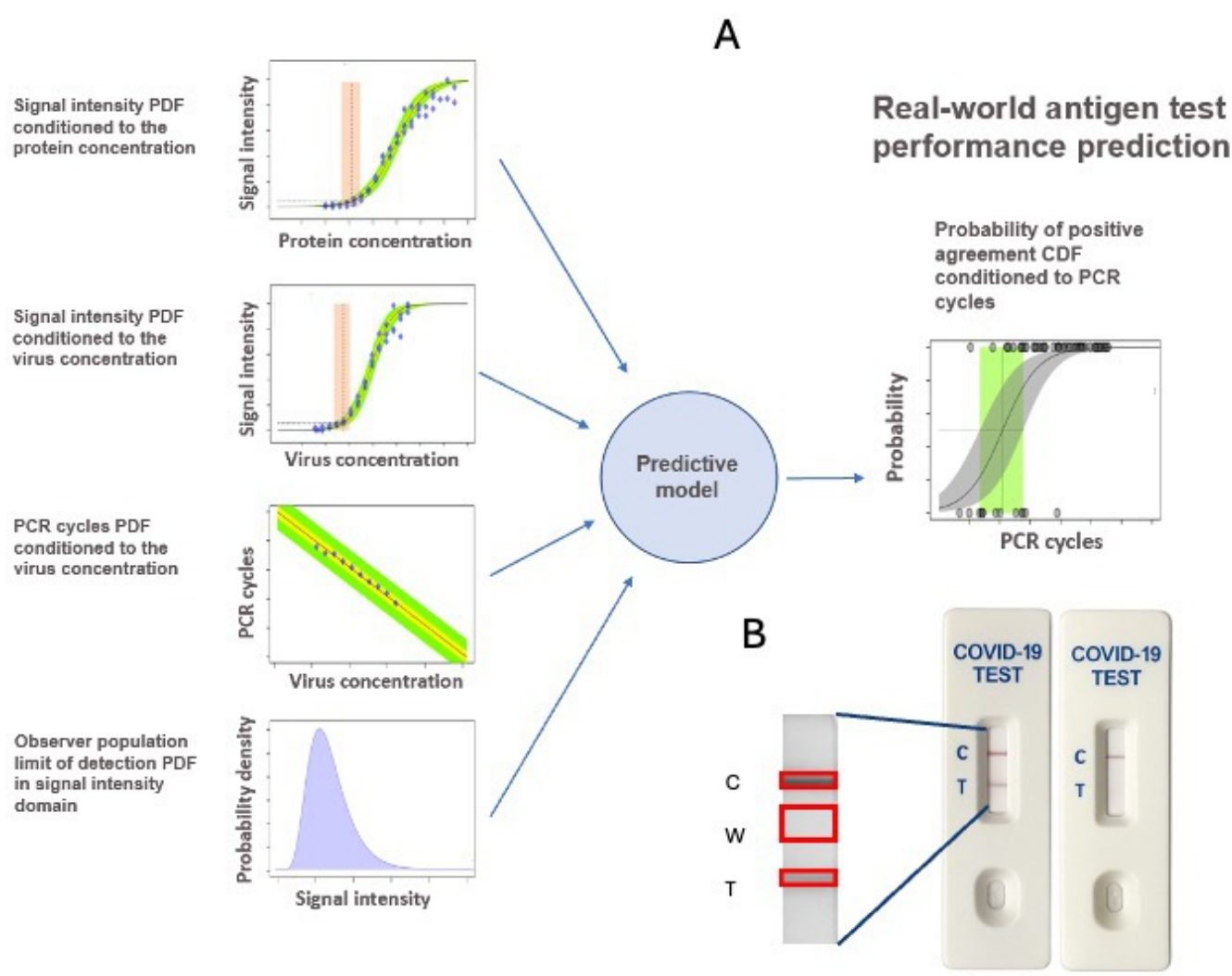
Methods

Overview

The present methodology combines (1) a quantitative evaluation of the test signal response to concentrations of target protein and inactive or active virus, (2) a statistical characterization of the LoD using observers' visual acuity

of the test band, and (3) a calibration of a gold-standard method (ie, qRT-PCR cycles) against virus concentration. We elaborate these quantitative methods and unfold a Bayesian-based predictive model to describe the real-world performance of the AT, quantified by the probability of positive agreement as a function of viral-load variables like qRT-PCR Cts. Figure 1A describes the different types of information involved in the predictive model.

Figure 1. (A) Schematic of the inference model. Input components are based on various laboratory calibrations of the antigen test—normalized signal intensity against protein and/or virus concentrations, qRT-PCR Cts against virus concentrations, and the limit of detection PDF of the observer population's visual assessment of the test result. The model predicts performance for real-world conditions of the antigen tests, quantified as the probability of agreement function. (B) Schematic of the antigen test results as positive or negative (recombinant antigen and/or inactivated virus or patient sample), according to the manufacturer's protocol and read at the specified development time. The cassette image was generated using a cell phone camera, and the annotated regions of interest were the control band (C), local white background (W), and test band (T) areas. Mean gray scale values from these regions of interest are used to compute normalized signal intensity for model fitting. Ct: cycle threshold; PDF: probability density function; PPA: positive percent agreement; qRT-PCR: quantitative real-time polymerase chain reaction.



AT Signal Response Characterization

In a selected lateral flow test, we measured the signal intensity of the test, white background, and control band, captured digitally in a photograph using a cell phone camera (Figure 1B). We processed this image and evaluated the average pixel intensity in 3 ROIs on the nitrocellulose strip. The resulting gray scale-normalized signal intensity is a continuous variable independent of the observer. There

are existing methodologies to assess AT performance, such as test band signal intensity [20] and LoD studies [21]. We calculated the signal intensity by subtracting the white background and test band average pixel brightness and normalizing by the largest signal intensity present in the dataset. The software was designed to provide use instructions, obtain gray scale pixel intensity of the ROI, compute normalized pixel intensities, and generate a written report [22].

We analyzed the signal intensity response of an AT, calculating the signal intensity across a dilution series of the target recombinant protein. To fit these data, we used techniques based on isotherm modeling [23,24] and used the Langmuir-Freundlich adsorption model [25,26],

$$I = \frac{k C^b}{1 + k C^b} \quad (1)$$

with I being the normalized signal intensity, C the concentration, k the adsorption equilibrium constant, and b an empirical exponent close to one. The model parameters to estimate by fitting the normalized intensity data corresponded to the adsorption constant k and the exponent b . We used a Bayesian regression solved with Monte Carlo sampling, which provided a description of the model uncertainties. We followed a similar procedure to characterize the relationship of the normalized signal intensity with other variables in addition to that of using recombinant protein concentrations (eg, with a known plaque-forming unit/mL of SARS-CoV-2 followed by chemical inactivation of the virus).

Probability of Agreement Function in Naked-Eye Assessment

A common use of ATs involves human naked-eye interpretation of the result. The outcome of each assessment is a binary variable, either positive or negative (1 for positive or 0 for negative for mathematical analysis). We considered that for the naked-eye analysis, individuals required training to follow specific protocols to properly report test results. Before self-testing, participants completed a brief, standardized orientation (10-15 min) delivered in English or Spanish according to specifications of the Chelsea study. The module covered (1) correct nasal specimen collection and adherence to the manufacturer's instructions for use, (2) strict timing of the development/read window, (3) interpretation of positive, negative, and invalid outcomes, and (4) photographing and uploading results from a cell phone camera or tablet to the study's digital platform in real time. Participants received a 1-page pictorial quick-start guide. In addition, refresher prompts were available on the study platform. Thereby, efforts were made to stabilize the observer LoD distribution data used in our model.

The PPA or sensitivity of a test is a well-known measure of test performance. The PPA is strongly dependent on the viral concentration distribution of the tested samples (eg, sensitivity improves with higher concentrations of the target).

For an accurate description of the naked-eye performance of the test, we estimated the PPA as a *function* of the nucleoprotein concentration or other viral concentration-related variable, such as qRT-PCR Cts [8]. We modeled the PPA with a logistic function,

$$p(x) = \frac{1}{1 + e^{-(a + bx)}} \quad (2)$$

with x being the viral-load-related variable and $p(x)$ being the probability of positive agreement function. The model parameters to estimate fitting the binary naked-eye data are the intercept a and the slope b . The PPA function is commonly described against qRT-PCR cycles; similarly, we applied this method for other viral-load-related variables (like concentration and normalized signal intensity).

Predictive Model for the Probability of Positive Agreement

The formulation of the model followed a probabilistic approach, meaning that variables and relationships across the model were randomized and described by probability density functions (PDFs). We defined random variables used in this formulation. There was a group of continuous positive variables that were related to the viral load: the recombinant protein concentration x_{prot} , the virus concentration x_{vir} , the test-normalized signal intensity x_{int} , the observer LoD-normalized signal intensity x_{lod} , and the qRT-PCR cycles x_{cycle} . In addition, we had the binary agreement variable A , which indicated the observer assessment of the test outcome, with values of 0 (for negative) or 1 (for positive).

For the purpose of this analysis, the LoD did not represent an exact value. We analyzed the LoD associated with a group of observers (ie, a certain population) or a single observer; LoD depends on different environmental circumstances (eg, illumination, visual context) and individual abilities. Hence, we consider the LoD as a random variable defined by a PDF $p_{\text{LoD}}(x_{\text{lod}})$ in the domain of the normalized signal intensity x_{int} . The probability of positive agreement (ie, the conditional positive agreement PDF) was the corresponding cumulative distribution function of the LoD PDF,

$$p(A = 1 | x_{\text{int}}) = \int_0^{x_{\text{int}}} p_{\text{LoD}}(x_{\text{lod}}) dx_{\text{lod}} \quad (3)$$

Correspondingly, the LoD PDF in the signal intensity domain was the derivative of the probability of positive agreement in the same domain. It summarized the process of observation and assessment of the testing device by the observer or the observer population.

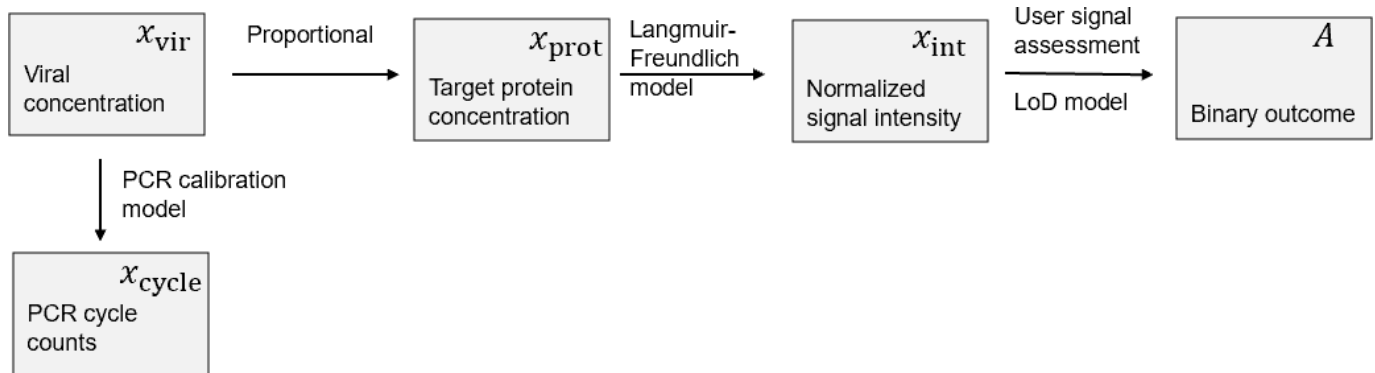
Probability of Positive Agreement Across Viral-Load-Related Domains

To follow, we transformed the probability of agreement in the signal intensity domain (expression 3) to the rest of the viral-load-related domains: recombinant protein concentration, viral concentration, and qRT-PCR cycles. The defined random variables and their causal dependencies are described in the Bayesian network of Figure 2. Let us consider, first, the case of the protein concentration domain. As per the intensity analysis described previously, we experimentally estimated a conditional probability for the signal intensity given the recombinant protein concentration $p(x_{\text{int}} | x_{\text{prot}})$. We used this information to propagate the probability of positive agreement from the signal intensity domain to the

protein concentration domain. For this purpose, we applied the probability chain rule,

$$p(A, x_{\text{int}} | x_{\text{prot}}) = p(A | x_{\text{int}}, x_{\text{prot}}) p(x_{\text{int}} | x_{\text{prot}}) \quad (4)$$

Figure 2. Bayesian network showing (boxes) the model random variables, (arrows) their causal relations and (annotations) relation models. LoD: limit of detection; PCR: polymerase chain reaction.



$$p(A | x_{\text{prot}}) = \int p(A | x_{\text{int}}) p(x_{\text{int}} | x_{\text{prot}}) dx_{\text{int}} \quad (5)$$

The functions within the integral involve the PPA in the signal intensity domain (expression 3) and the PDF of signal intensity conditioned to the protein concentration. We determined the integral by Monte Carlo integration. Similarly, we transformed the probability of agreement to the virus concentration domain and qRT-PCR cycles domain as

$$p(A | x_{\text{vir}}) = \int p(A | x_{\text{int}}) p(x_{\text{int}} | x_{\text{prot}}) p(x_{\text{prot}} | x_{\text{vir}}) dx_{\text{int}} dx_{\text{prot}} \quad (6)$$

and

$$p(A | x_{\text{cycle}}) = \int p(A | x_{\text{int}}) p(x_{\text{int}} | x_{\text{prot}}) p(x_{\text{prot}} | x_{\text{vir}}) p(x_{\text{vir}} | x_{\text{cycle}}) dx_{\text{int}} dx_{\text{prot}} dx_{\text{vir}} \quad (7)$$

To model the probability of agreement in the domain of qRT-PCR cycles, as is common for real-world testing, we solved Equation 7 by Monte Carlo integration. For this purpose, we needed to estimate models for the 4 conditional probabilities within the right-hand integrand, which involve the information represented in Figures 1 and 2. In our implementation, we first integrated in the virus concentration domain to have a relationship between the protein concentration and the qRT-PCR cycles, $p(x_{\text{prot}} | x_{\text{cycle}})$. Thus,

$$p(A | x_{\text{cycle}}) = \int p(A | x_{\text{int}}) p(x_{\text{int}} | x_{\text{prot}}) p(x_{\text{prot}} | x_{\text{cycle}}) dx_{\text{int}} dx_{\text{prot}} \quad (8)$$

The conditional probability of positive agreement $p(A | x_{\text{cycle}})$ fully describes the test performance in the qRT-PCR cycle domain. For a given PDF of the sample polymerase chain reaction (PCR) cycles distribution, the resulting sensitivity $p(A)$ is by integration,

$$p(A = 1) = \int p(A = 1 | x_{\text{cycle}}) p(x_{\text{cycle}}) dx_{\text{cycle}} \quad (9)$$

Integrating in the x_{int} domain and taking into account that the observer assessment is only dependent on normalized signal intensity (see Figure 2)

$$p(A | x_{\text{int}}, x_{\text{prot}}) = p(A | x_{\text{int}}),$$

For a given collection of N real-world samples with PCR cycles $x_{\text{cycle}} = \{x_1, x_2, \dots, x_n, \dots, x_N\}$, the previous integral is approximated by the average of the PPA function evaluated at the sample qRT-PCR cycles,

$$p(A = 1) \approx \frac{1}{N} \sum_{i=1}^N p(A = 1 | x_n) \quad (10)$$

We have made software available to calculate the probability of agreement [22].

Ethical Considerations

This study was reviewed and approved by Advarra IRB for the protocol “Center of Complex Interventions – IDx20-001, Community frequent antigen testing to monitor COVID-19 in senior public housing setup (Pro00059157).” The most recent continuing review approval was granted on November 13, 2023, with an approval period through November 13, 2024. Advarra attests compliance with the US Department of Health and Human Services 45 CFR 46 and Food and Drug Administration 21 CFR 50/56 and is registered with OHRP/FDA (IRB number 00000971). All procedures adhered to the ethical standards of the responsible institutional/national committees and the World Medical Association Declaration of Helsinki. All personnel received certification and training through the Collaborative Institutional Training Initiative for human subjects research protection. There was no compensation to participants.

All study personnel completed Collaborative Institutional Training Initiative coursework prior to engaging in any human-subjects activities. This training is required by our IRB and institutional policy and included role-appropriate modules in Biomedical Human Subjects Research, Good Clinical Practice, Responsible Conduct of Research, Conflicts of Interest, and Health Insurance Portability and Accountability Act Privacy/Security. Certificates were verified by the private investigator and maintained on file. The curriculum covers the Belmont Report principles and applicable

regulations (45 CFR 46 and Food and Drug Administration 21 CFR Parts 50/56), informed consent and documentation, recruitment and equitable selection, protection of vulnerable populations, adverse event and deviation reporting, data privacy/confidentiality, and secure data management.

Because the protocol involves point-of-care antigen testing and handling of respiratory specimens, staff also completed biosafety/Blood borne Pathogen training and followed BSL-2-appropriate standard operating procedures. All participant-facing procedures (screening, consent, anterior-nares swab collection, test execution, results disclosure) were conducted only by trained personnel under IRB-approved standard operating procedures. Study data were coded with limited identifiers, stored in access-controlled databases, and managed according to least-privilege access and audit-trail requirements. This statement documents personnel competence and compliance with human-subjects protections for the conduct of this study.

Prior to study procedures, all participants were informed of the study purpose, procedures, potential risks and benefits, data uses, and their right to withdraw without penalty. Written informed consent was obtained from each participant using IRB-approved consent materials. No identifiable personal information is reported in this manuscript.

Data were collected and stored using IRB-approved procedures designed to protect participant privacy and confidentiality; only deidentified or aggregated data are presented. For studies of internet/digital tools, we complied with applicable local, national, and international regulations on the protection of personal information, privacy, and human rights.

Any protocol amendments, consent-form changes, or substantive reportable events (eg, unanticipated problems, adverse device effects, or protocol violations affecting rights,

safety, or data integrity) were submitted to Advarra in accordance with IRB requirements prior to implementation.

Results

We illustrate the application of the described methodology to characterize the analyzed COVID-19 AT brand and predict the corresponding real-world probability of agreement against qRT-PCR data.

Figure 3 shows the signal intensity data corresponding to protein dilutions prepared in the laboratory for this AT and the corresponding Langmuir-Freundlich regression model; from the analysis, we modeled the conditional PDF $p(x_{\text{int}}|x_{\text{prot}})$. Figure 4 illustrates our modeled relationships across various viral-load-related variables based on our experimental characterization of the AT and qRT-PCR calibration. Figure 4A shows the signal intensity analysis of the AT based on serial dilutions of inactivated virus. The plot is similar to Figure 3, which shows the signal response to protein dilutions. By combining the protein and virus curves for common signal intensity responses, we calibrated a linear model that describes the relationship between protein and inactivated virus concentration (Figure 4B). Figure 4C shows the calibration of the qRT-PCR Ct curve based on PCR-analyzed inactivated virus dilution series. The qRT-PCR analysis of the dilution series was conducted by the same center used for self-testing our ordinary (real-world) testing program. Figure 4D shows the relationship between qRT-PCR cycles and protein concentration, transformed from the viral concentration domain by the protein-virus relationship characterized in Figure 4B. All the relationships shown in Figure 4 are modeled as conditional PDFs; for illustration, the plots show specific confidence limits.

Figure 3. Normalized signal intensity data for 3 series of protein dilution curves for one of the COVID-19 test brands and Langmuir-Freundlich model. The estimation of the LoD (LoD confidence intervals at 95%). LoD in signal intensity was 5% in the normalized signal intensity. LoD: limit of detection.

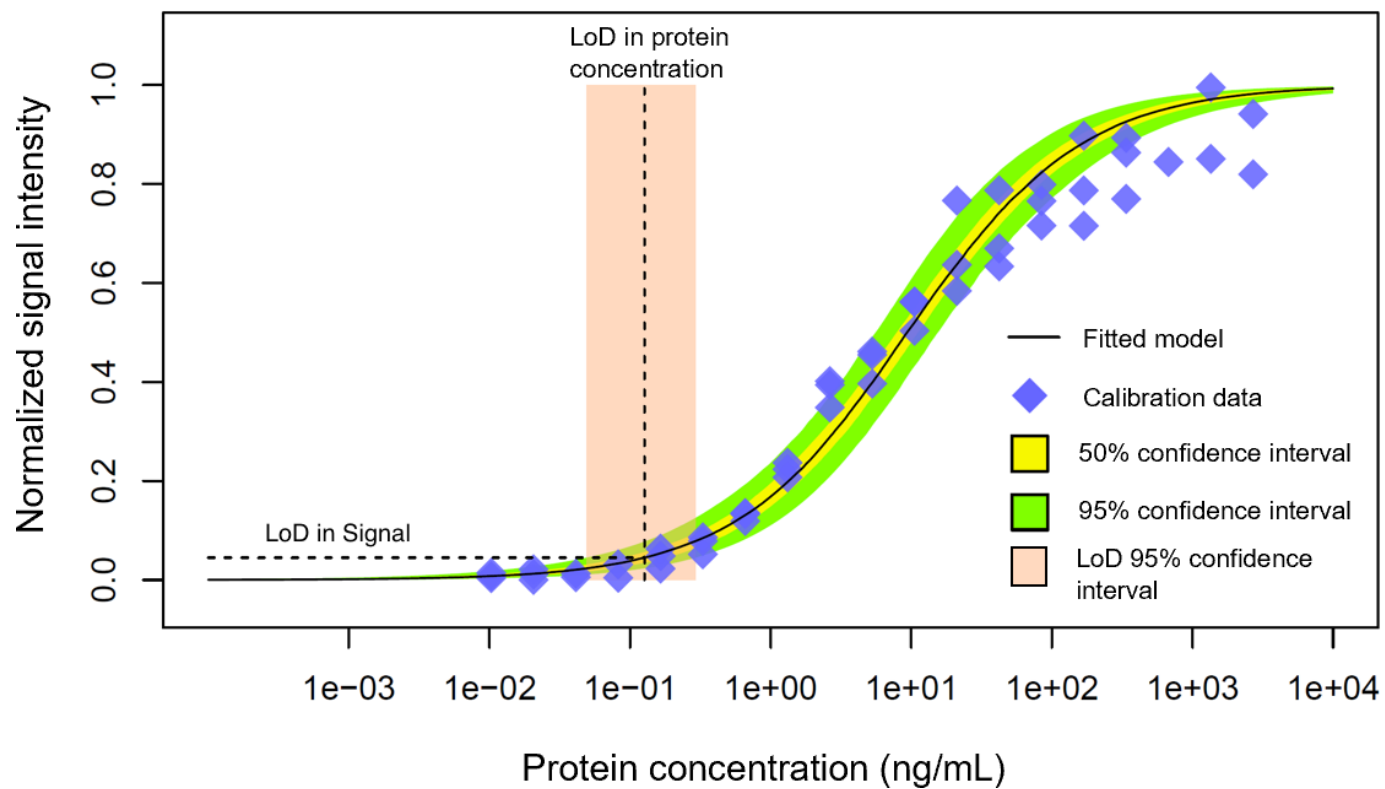
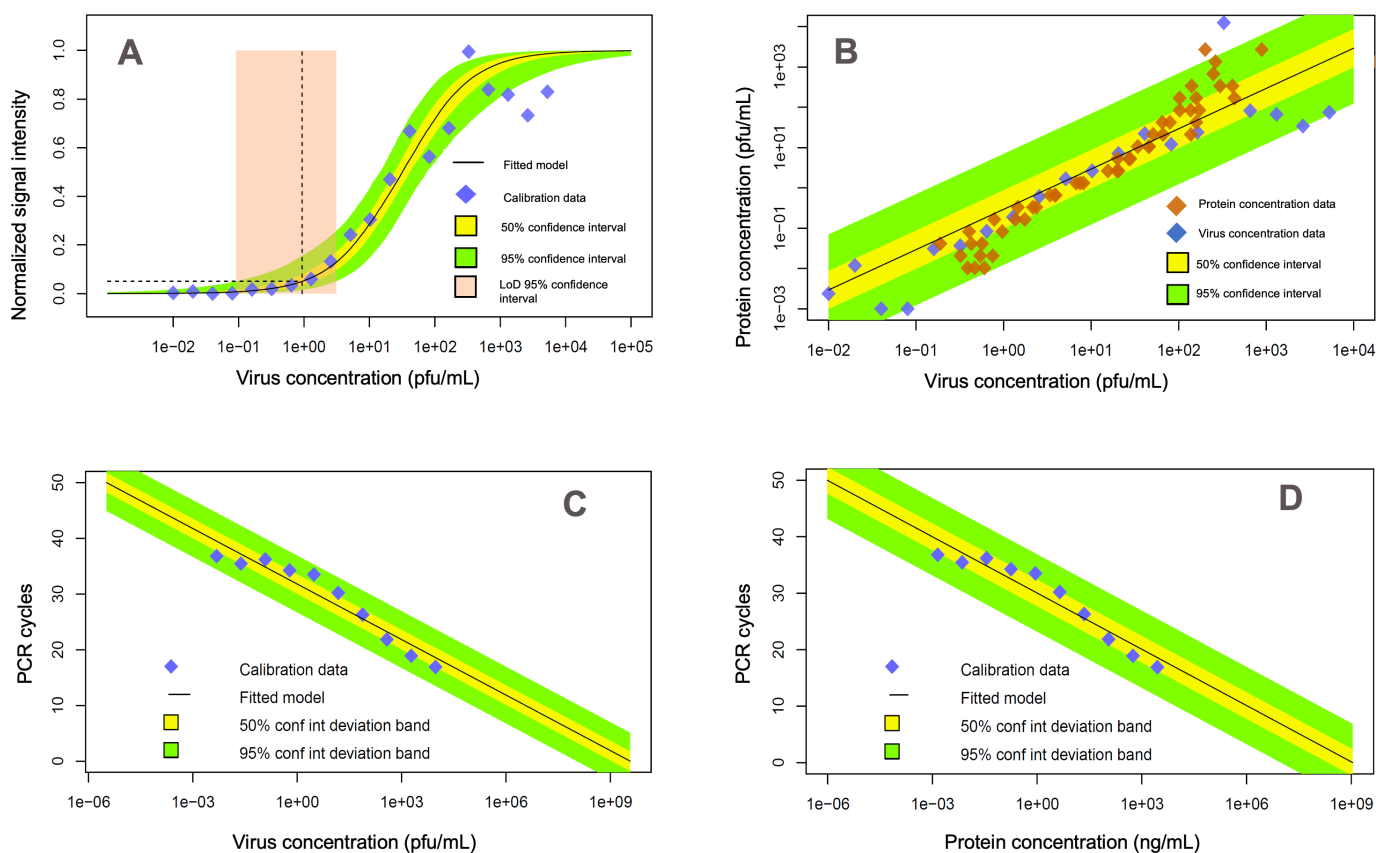


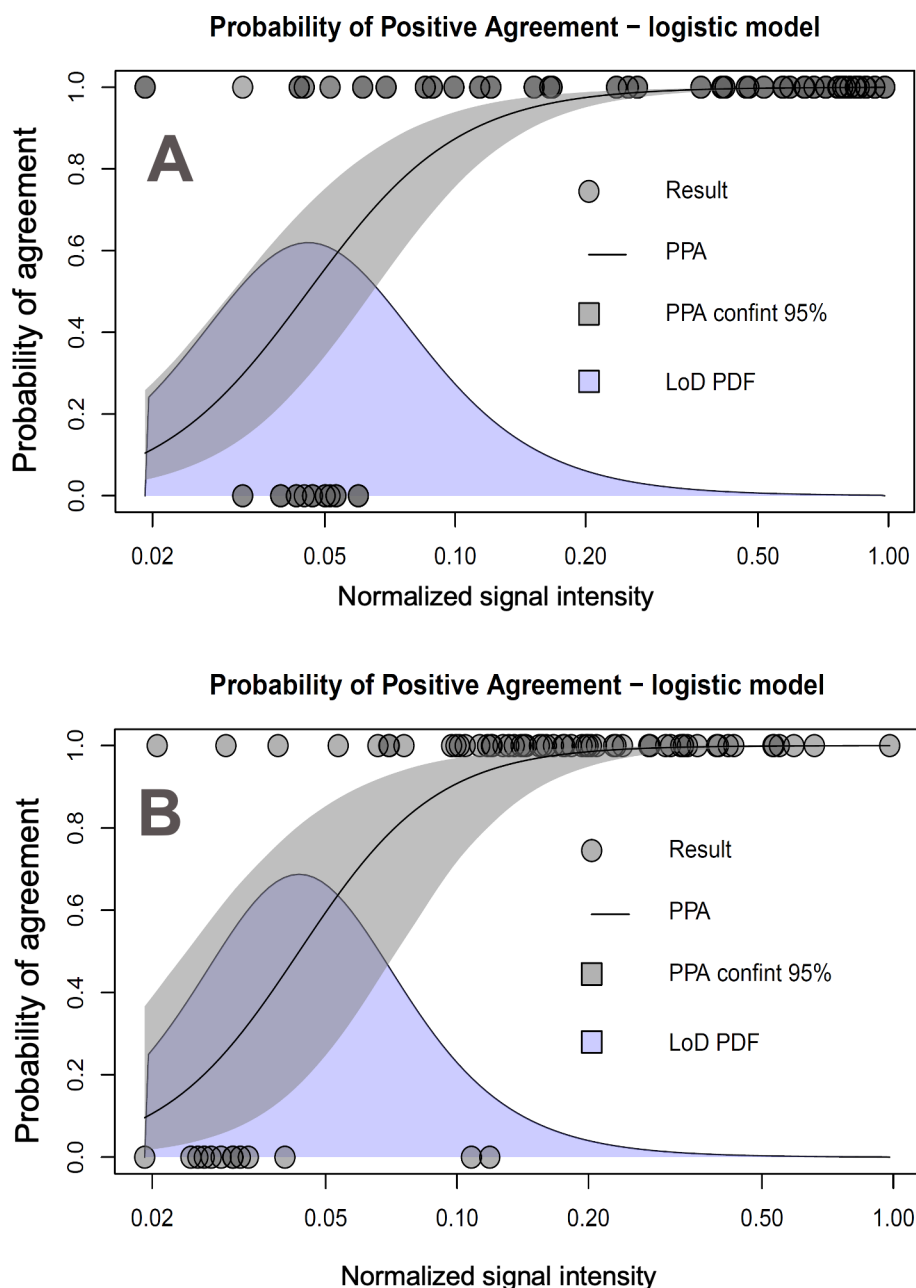
Figure 4. Characterization of the signal response with inactivated virus and domain-related transformation functions calibrated from experimental data. (A) The normalized signal response analysis for serial dilutions of inactivated virus. (B) Protein and virus concentration linear relational model based on the common signal intensity response of the devices, which allows transforming virus concentration to protein concentration and vice versa. (C) PCR cycle response calibration to inactivated virus dilution series. (D) qRT-PCR cycle response related to protein concentration, by combining transformations (B) and (C). LoD: limit of detection; PCR: polymerase chain reaction; pfu: plaque-forming unit; qRT-PCR: quantitative real-time polymerase chain reaction.



An additional component required by our predictive method is the LoD PDF of the observer's population $p_{\text{LoD}}(x_{\text{lod}})$ in the domain of the signal intensity. Figure 5A shows the real-world binary assessment conditioned to the signal intensity (based on Chelsea study participants' AT results and uploaded cell phone camera photos), the corresponding PPA function of the signal intensity estimated by logistic regression, and the LoD PDF in the domain of the signal intensity. The former is the observed LoD in the domain of the signal intensity for the participant population. Although, we can characterize the LoD in the domain of the signal intensity

based on the real-world naked-eye data of the participant population, it is interesting to compare this estimation to one based on easier-to-obtain naked-eye data. Figure 5B shows naked-eye test results, the PPA function, and LoD PDF for trained staff assessing the AT result using a dilution series of recombinant SARS-CoV-2 nucleoprotein (including at a concentration of 0; ie, negatives) with blinded concentration labels. Acknowledging the nonnegligible effect of user heterogeneity [27] and external conditions, the probability functions (cumulative and density) were remarkably similar, as shown in Figures 5A and 5B.

Figure 5. Probability of positive agreement and LoD probability density for naked-eye assessments of antigen tests in common cassette presentation made by two different groups of observers. (A) Trained staff observing results from SARS-CoV-2 nucleoprotein dilutions blinded to the observer and (B) community participants of the study reporting self-tested results. In the case of trained staff, normalized intensity was calculated from laboratory environment photographs using cell phone cameras. The community participants' mobile phone photographs were uploaded to the digital reporting system and the visual assessment was a self-report also recorded in real time through the study's digital platform. LoD: limit of detection; PDF: probability density function.

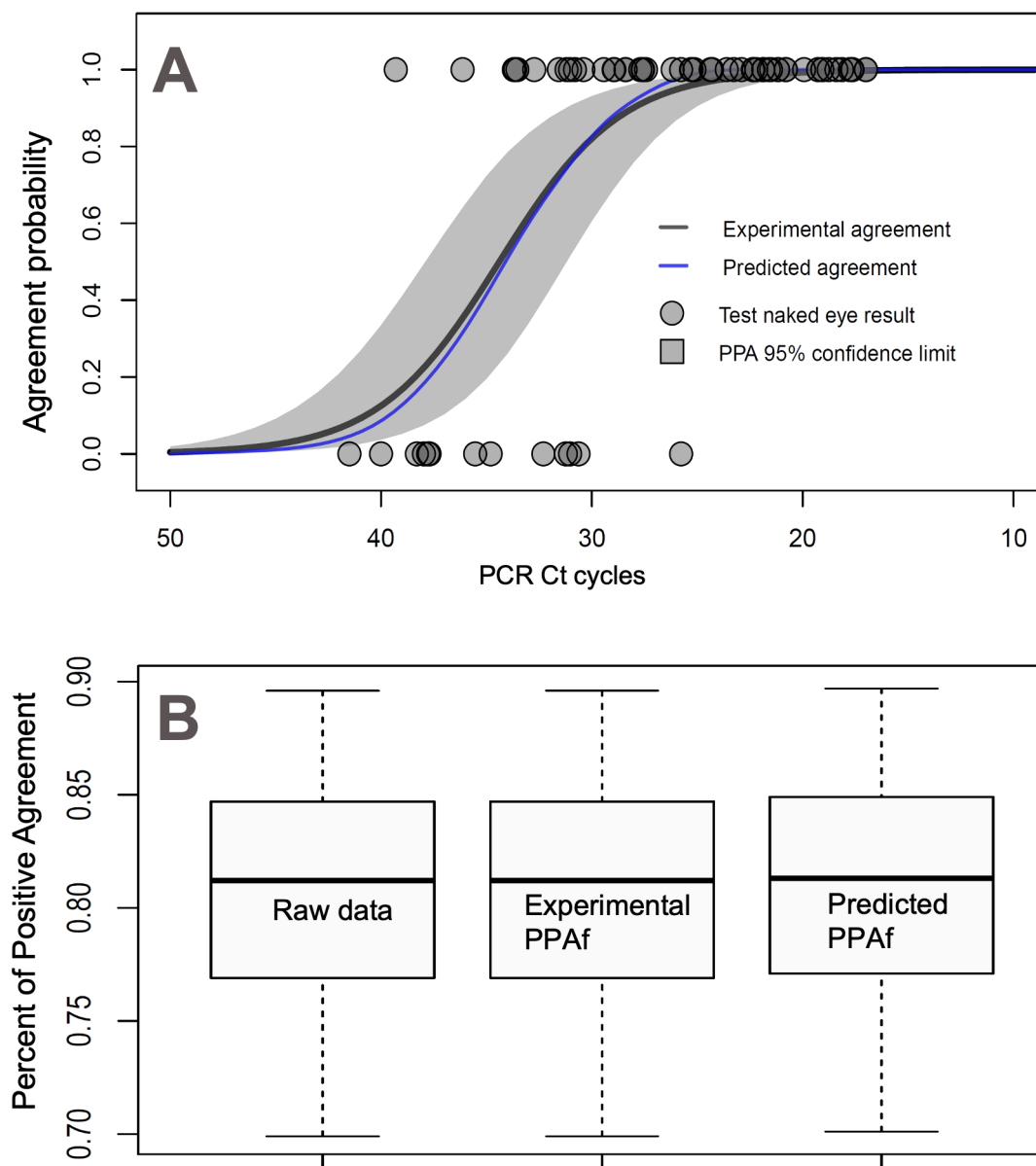


With the analysis shown in Figures 3-5, modeling the input components of the predictive model, we performed our calculations of the PPA function of the qRT-PCR cycles and compared it with the *observed* PPA function (ie, derived from real-world data).

Figure 6A shows the real-world data of binary results obtained with the AT (ie, self-tested and logged data from the Chelsea study) and the corresponding observed PPA as a function of the qRT-PCR cycles. The calculation involved a description of the model uncertainties, illustrated in the

plot with confidence intervals. Superimposed, we represented the predicted PPA function in the domain of the qRT-PCR, calculated using our method (Equation 8). The plot clearly shows that the predicted and observed PPA functions are very close and within the figure's confidence limits. Figure 6B shows a box plot of the overall data PPA for samples collected in the real-world setting for the AT and a comparison of the corresponding predicted real-world PPA according to expression 10. We can verify that both box plots provided close results, within statistical significance.

Figure 6. (A) Probability of agreement functions based on experimental real-world data (black curve) and resulting from our predictive model (blue curve). The experimental PPA function was obtained from a logistic regression analysis of the Chelsea project data for the test users; the gray area shows 95% confidence limits on the experimental PPA function. (B) Box plots showing the PPA calculated over the real-world raw data, the observed PPA function, and the predicted PPA function. Box plot bounds are Clopper-Pearson confidence limits for percentiles 50% (box) and 95% (segment). PPA: positive percent agreement; PCR: polymerase chain reaction.



Discussion

Principal Findings

Our predictive framework and the COVID-19 case study indicate that the real-world performance of ATs can be forecast from rapid laboratory measurements together with a statistical characterization of the population's LoD. First, analysis of test signal intensity and calibration to qRT-PCR can be completed entirely under controlled laboratory conditions—a rapid process that yields precise estimates of baseline AT performance. Second, we show that empirically characterizing the observer population's LoD distribution for visual interpretation of ATs is feasible.

Across two distinct observer groups and testing settings, the LoD PDFs and the resulting PPA-normalized signal

intensity relationships were remarkably similar, suggesting a relatively stable visual decision threshold inherent to this device design. We also demonstrated the practicality of paired AT/qRT-PCR sampling with image uploads, with minimal discrepancy between self-reported results and those recorded by trained staff. Nonetheless, environmental variability (lighting, optics and image compression, and differences among phone cameras), user motivation, and other external factors can broaden the effective LoD distribution.

Because we estimated the PPA as a continuous function of qRT-PCR Ct via logistic regression and then computed the expected sensitivity over a predefined reference Ct distribution, performance can be standardized across sites, time periods, and brands despite heterogeneous sampling. Bayesian fitting used in this study provides a practical path to hierarchical (brand-level) extensions we have recently

reported in a complementary publication of the Chelsea clinical study.

Laboratory analyses and our predictive modeling framework provide a fast, standardized path to anticipate clinical performance and guide smarter evaluations of ATs, which is very useful information for the real-world validation of ATs. For regulatory decisions, quantitative summaries should be reported—PPA(Ct) with uncertainty bandwidth and expected sensitivity under a preregistered reference Ct distribution—alongside observer LoD distributions (trained vs lay) and robustness to environmental and site effects, all with traceable calibrations and bridging analyses across lots, sites, and readers. After clearance, life cycle quality control should be maintained via routine lot verification, stability tracking, app-enabled field analytics for normalized signal intensity and invalid rates, anomaly detection, and timely label updates when variant-linked changes arise. At minimum, each evaluation should include the PPA(Ct) function with 95% CIs, expected sensitivity under a stated reference Ct distribution, observer LoD PDFs, calibration, and fit diagnostics.

The accurate calibration of the relational components (Figure 2) is fundamental for the fitness of the model. We performed triplicate dilution curves during the process of sample preparation for analysis. Additionally, we verified that thermal inactivation of the virus, as carried out for generating the serial virus dilutions, resulted in marginal loss of capsid protein detection (approximately a 2-fold difference), likely due to heat stability. We also used chemically inactivated virus alongside heat-inactivated virus stocks for signal intensity calibration.

The purpose of this study was to describe a method for the quantification and prediction of AT performance. Further work is underway, as we seek to expand our data for the

comparative analysis of performance prediction across several test brands, comparative calibration of the qRT-PCR cycles across service providers, and further characterization of the signal intensity LoD for common ATs.

Conclusions

An accurate description of the AT signal intensity response conditioned by variables related to viral load, such as concentration of recombinant protein and concentration of inactivated virus, was established under laboratory conditions. These evaluations involved image processing of photographs and human naked-eye assessments using dilution series of the nucleoprotein of the SARS-CoV-2 virus. Modeling the performance on real testing conditions involved integrating the mentioned laboratory evaluation with information on the ability of the observer population to recognize the device's visual response, which can be described by the LoD PDF of the observer population in the domain of the test signal intensity. We described the overall test performance with the PPA *function* of the qRT-PCR Cts instead of using the common PPA (ie, sensitivity) for segments of clinical data. Our framework predicts PPA versus Ct by linking laboratory-normalized signal intensity–concentration models, observer LoD distributions, and a Ct–viral-load calibration. Conclusions are mathematical and are specific to the disease, assay, and setup. External validity depends on specific Ag-LFTs, user populations (for visual acuity and test interpretation), and qRT-PCR calibration across laboratories. Although the Chelsea dataset supports internal validity, broader clinical evaluation under intended-use conditions is required before generalized clinical claims, particularly for disease targets other than COVID-19. Therefore, the presented methodology has promising applications for the evaluation of ATs, as it involves a quick appraisal of real-world test performance.

Acknowledgments

We gratefully acknowledge the collaboration of Dr Alfred Harding and Dr Lee Gehrke for their virology expertise and for providing SARS-CoV-2 virus isolates to the study; Dr Karen Weeks and the personnel of Eco Laboratory, Acton, MA, for the processing and reporting of the nasal swab qRT-PCR data; Michael Fannon and BioIT Solutions personnel for their support on data management and reporting; Mateo Bonnet for front-end digital platform design and computer science expertise; the Center for Complex Intervention in Cambridge, MA, for community and scientific advice; the Chelsea Housing Authority; the Chelsea City Health Department; and the participants in this research for their utmost enthusiasm and support. We also extend our sincere thanks to the Reagan-Udall Foundation for the Food and Drug Administration personnel and their generous financial support through grant number 02282022 RUF, which enabled the development of this study. This research was supported by the Reagan-Udall Foundation for the Food and Drug Administration through grant number 02282022 RUF, awarded to IDX20 Inc.

Conflicts of Interest

IB is a founder of IDX20 Inc, a private company affiliated with this study. MB is a founder of Info Analytics Innovations, a private company affiliated with this study. The authors declare no additional conflicts of interest.

References

1. Wagenhäuser I, Knies K, Rauschenberger V, et al. Clinical performance evaluation of SARS-CoV-2 rapid antigen testing in point of care usage in comparison to RT-qPCR. *EBioMedicine*. Jul 2021;69:103455. [doi: [10.1016/j.ebiom.2021.103455](https://doi.org/10.1016/j.ebiom.2021.103455)] [Medline: [34186490](https://pubmed.ncbi.nlm.nih.gov/34186490/)]

2. Pickering S, Batra R, Merrick B, et al. Comparative performance of SARS-CoV-2 lateral flow antigen tests and association with detection of infectious virus in clinical specimens: a single-centre laboratory evaluation study. *Lancet Microbe*. Sep 2021;2(9):e461-e471. [doi: [10.1016/S2666-5247\(21\)00143-9](https://doi.org/10.1016/S2666-5247(21)00143-9)] [Medline: [34226893](https://pubmed.ncbi.nlm.nih.gov/34226893/)]
3. Pollock NR, Savage TJ, Wardell H, et al. Correlation of SARS-CoV-2 nucleocapsid antigen and RNA concentrations in nasopharyngeal samples from children and adults using an ultrasensitive and quantitative antigen assay. *J Clin Microbiol*. Mar 19, 2021;59(4):e03077-20. [doi: [10.1128/JCM.03077-20](https://doi.org/10.1128/JCM.03077-20)] [Medline: [33441395](https://pubmed.ncbi.nlm.nih.gov/33441395/)]
4. Karlafti E, Tsavdaris D, Kotzakioulafi E, et al. The diagnostic accuracy of SARS-CoV-2 nasal rapid antigen self-test: a systematic review and meta-analysis. *Life (Basel)*. Jan 19, 2023;13(2):281. [doi: [10.3390/life13020281](https://doi.org/10.3390/life13020281)] [Medline: [36836639](https://pubmed.ncbi.nlm.nih.gov/36836639/)]
5. Wagenhäuser I, Knies K, Hofmann D, et al. Virus variant-specific clinical performance of SARS coronavirus two rapid antigen tests in point-of-care use, from November 2020 to January 2022. *Clin Microbiol Infect*. Feb 2023;29(2):225-232. [doi: [10.1016/j.cmi.2022.08.006](https://doi.org/10.1016/j.cmi.2022.08.006)] [Medline: [36028089](https://pubmed.ncbi.nlm.nih.gov/36028089/)]
6. Vaeth MJE, Abdullah O, Cheema M, et al. Accuracy of expired BinaxNOW rapid antigen tests. *Microbiol Spectr*. Aug 17, 2023;11(4):e0208823. [doi: [10.1128/spectrum.02088-23](https://doi.org/10.1128/spectrum.02088-23)] [Medline: [37428037](https://pubmed.ncbi.nlm.nih.gov/37428037/)]
7. Soni A, Herbert C, Lin H, et al. Performance of rapid antigen tests to detect symptomatic and asymptomatic SARS-CoV-2 infection: a prospective cohort study. *Ann Intern Med*. Jul 2023;176(7):975-982. [doi: [10.7326/M23-0385](https://doi.org/10.7326/M23-0385)] [Medline: [37399548](https://pubmed.ncbi.nlm.nih.gov/37399548/)]
8. Corman VM, Haage VC, Bleicker T, et al. Comparison of seven commercial SARS-CoV-2 rapid point-of-care antigen tests: a single-centre laboratory evaluation study. *Lancet Microbe*. Jul 2021;2(7):e311-e319. [doi: [10.1016/S2666-5247\(21\)00056-2](https://doi.org/10.1016/S2666-5247(21)00056-2)] [Medline: [33846704](https://pubmed.ncbi.nlm.nih.gov/33846704/)]
9. Bornemann L, Kaup O, Kleideiter J, et al. Virus variant-specific clinical performance of a SARS-CoV-2 rapid antigen test with focus on Omicron variants of concern. *Clin Microbiol Infect*. Aug 2023;29(8):1085. [doi: [10.1016/j.cmi.2023.05.009](https://doi.org/10.1016/j.cmi.2023.05.009)] [Medline: [37182639](https://pubmed.ncbi.nlm.nih.gov/37182639/)]
10. NIH study informs antigen testing for the SARS-cov-2 virus: repeat testing reduces false-negatives; FDA updates recommendations. National Institute of Biomedical Imaging and Bioengineering. 2025. URL: <https://www.nibib.nih.gov/news-events/newsroom/nih-study-informs-antigen-testing-sars-cov-2-virus> [Accessed 2025-09-30]
11. Yoon CS, Park HY, Park HK, et al. The influence of pneumococcal positivity on clinical outcomes among patients hospitalized with COVID-19: a retrospective cohort study. *PLoS One*. 2025;20(8):e0329474. [doi: [10.1371/journal.pone.0329474](https://doi.org/10.1371/journal.pone.0329474)] [Medline: [40839694](https://pubmed.ncbi.nlm.nih.gov/40839694/)]
12. Testing for COVID-19. CDC. 2025. URL: <https://www.cdc.gov/covid/testing/index.html> [Accessed 2025-09-30]
13. Kost GJ. The impact of increasing disease prevalence, false omissions, and diagnostic uncertainty on coronavirus disease 2019 (COVID-19) test performance. *Arch Pathol Lab Med*. Jul 1, 2021;145(7):797-813. [doi: [10.5858/arpa.2020-0716-SA](https://doi.org/10.5858/arpa.2020-0716-SA)] [Medline: [33684204](https://pubmed.ncbi.nlm.nih.gov/33684204/)]
14. Stohr J, Wennekes M, van der Ent M, et al. Clinical performance and sample freeze-thaw stability of the cobas®6800 SARS-CoV-2 assay for the detection of SARS-CoV-2 in oro-/nasopharyngeal swabs and lower respiratory specimens. *J Clin Virol*. Dec 2020;133:104686. [doi: [10.1016/j.jcv.2020.104686](https://doi.org/10.1016/j.jcv.2020.104686)] [Medline: [33221622](https://pubmed.ncbi.nlm.nih.gov/33221622/)]
15. Stohr J, Zwart VF, Goderski G, et al. Self-testing for the detection of SARS-CoV-2 infection with rapid antigen tests for people with suspected COVID-19 in the community. *Clin Microbiol Infect*. May 2022;28(5):695-700. [doi: [10.1016/j.cmi.2021.07.039](https://doi.org/10.1016/j.cmi.2021.07.039)] [Medline: [34363945](https://pubmed.ncbi.nlm.nih.gov/34363945/)]
16. Vaeth MJE, Cheema M, Omer S, et al. Self-administered versus clinician-performed BinaxNOW COVID rapid test: a comparison of accuracy. *Microbiol Spectr*. Mar 5, 2024;12(3):e0252523. [doi: [10.1128/spectrum.02525-23](https://doi.org/10.1128/spectrum.02525-23)] [Medline: [38349164](https://pubmed.ncbi.nlm.nih.gov/38349164/)]
17. Wagenhäuser I, Knies K, Pscheidt T, et al. SARS-CoV-2 antigen rapid detection tests: test performance during the COVID-19 pandemic and the impact of COVID-19 vaccination. *EBioMedicine*. Nov 2024;109:105394. [doi: [10.1016/j.ebiom.2024.105394](https://doi.org/10.1016/j.ebiom.2024.105394)] [Medline: [39388783](https://pubmed.ncbi.nlm.nih.gov/39388783/)]
18. Herbert C, Wang B, Lin H, et al. Performance of and severe acute respiratory syndrome coronavirus 2 diagnostics based on symptom onset and close contact exposure: an analysis from the Test Us at Home prospective cohort study. *Open Forum Infect Dis*. Jun 2024;11(6):ofae304. [doi: [10.1093/ofid/ofae304](https://doi.org/10.1093/ofid/ofae304)] [Medline: [38911947](https://pubmed.ncbi.nlm.nih.gov/38911947/)]
19. Harmon A, Chang C, Salcedo N, et al. Validation of an at-home direct antigen rapid test for COVID-19. *JAMA Netw Open*. Aug 2, 2021;4(8):e2126931. [doi: [10.1001/jamanetworkopen.2021.26931](https://doi.org/10.1001/jamanetworkopen.2021.26931)] [Medline: [34448871](https://pubmed.ncbi.nlm.nih.gov/34448871/)]
20. Bosch I, de Puig H, Hiley M, et al. Rapid antigen tests for dengue virus serotypes and Zika virus in patient serum. *Sci Transl Med*. Sep 27, 2017;9(409):eaan1589. [doi: [10.1126/scitranslmed.aan1589](https://doi.org/10.1126/scitranslmed.aan1589)] [Medline: [28954927](https://pubmed.ncbi.nlm.nih.gov/28954927/)]
21. Safenkova I, Zherdev A, Dzantiev B. Factors influencing the detection limit of the lateral-flow sandwich immunoassay: a case study with potato virus X. *Anal Bioanal Chem*. Jun 2012;403(6):1595-1605. [doi: [10.1007/s00216-012-5985-8](https://doi.org/10.1007/s00216-012-5985-8)] [Medline: [22526658](https://pubmed.ncbi.nlm.nih.gov/22526658/)]
22. IDX20. 2025. URL: <https://idx20.us/> [Accessed 2025-09-12]

23. Douven S, Paez CA, Gommers CJ. The range of validity of sorption kinetic models. J Colloid Interface Sci. Jun 15, 2015;448:437-450. [doi: [10.1016/j.jcis.2015.02.053](https://doi.org/10.1016/j.jcis.2015.02.053)] [Medline: [25765735](https://pubmed.ncbi.nlm.nih.gov/25765735/)]
24. Syafiuddin A, Salmiati S, Jonbi J, Fulazzaky MA. Application of the kinetic and isotherm models for better understanding of the behaviors of silver nanoparticles adsorption onto different adsorbents. J Environ Manage. Jul 15, 2018;218:59-70. [doi: [10.1016/j.jenvman.2018.03.066](https://doi.org/10.1016/j.jenvman.2018.03.066)] [Medline: [29665487](https://pubmed.ncbi.nlm.nih.gov/29665487/)]
25. Altın O, Özbelge HÖ, Doğu T. Use of general purpose adsorption isotherms for heavy metal–clay mineral interactions. J Colloid Interface Sci. Feb 1998;198(1):130-140. [doi: [10.1006/jcis.1997.5246](https://doi.org/10.1006/jcis.1997.5246)]
26. Saadi R, Saadi Z, Fazaeli R, Fard NE. Monolayer and multilayer adsorption isotherm models for sorption from aqueous media. Korean J Chem Eng. May 2015;32(5):787-799. [doi: [10.1007/s11814-015-0053-7](https://doi.org/10.1007/s11814-015-0053-7)]
27. De Nardo P, Tebon M, Savoldi A, et al. Diagnostic accuracy of a rapid SARS-CoV-2 antigen test among people experiencing homelessness: a prospective cohort and implementation study. Infect Dis Ther. Apr 2023;12(4):1073-1082. [doi: [10.1007/s40121-023-00787-0](https://doi.org/10.1007/s40121-023-00787-0)] [Medline: [36907951](https://pubmed.ncbi.nlm.nih.gov/36907951/)]

Abbreviations

Ag-LFT: antigen lateral flow test
AT: antigen test
Ct: cycle threshold
IRB: institutional review board
LoD: limit of detection
PCR: polymerase chain reaction
PDF: probability density function
PPA: percent positive agreement
qRT-PCR: quantitative real-time polymerase chain reaction
ROI: region of interest

Edited by Fuqing Wu; peer-reviewed by Gerald Kost, Helena de Puig; submitted 08.11.2024; final revised version received 20.08.2025; accepted 29.08.2025; published 06.10.2025

Please cite as:

Bosch M, Garcia D, Rudtner L, Salcedo N, Colmenares R, Hoche S, Arocha J, Hall D, Moreno A, Bosch I
Real-World Performance of COVID-19 Antigen Tests: Predictive Modeling and Laboratory-Based Validation
JMIRx Med 2025;6:e68376
URL: <https://med.jmirx.org/2025/1/e68376>
doi: [10.2196/68376](https://doi.org/10.2196/68376)

© Miguel Bosch, Dawlyn Garcia, Lindsey Rudtner, Nol Salcedo, Raul Colmenares, Sina Hoche, Jose Arocha, Daniella Hall, Adriana Moreno, Irene Bosch. Originally published in JMIRx Med (<https://med.jmirx.org>), 06.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIRx Med, is properly cited. The complete bibliographic information, a link to the original publication on <https://med.jmirx.org/>, as well as this copyright and license information must be included.