Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has emphasized the importance and challenges of correctly interpreting antibody test results. Identification of positive and negative samples requires a classification strategy with low error rates, which is hard to achieve when the corresponding measurement values overlap. Additional uncertainty arises when classification schemes fail to account for complicated structure in data. We address these problems through a mathematical framework that combines high dimensional data modeling and optimal decision theory. Specifically, we show that appropriately increasing the dimension of data better separates positive and negative populations and reveals nuanced structure that can be described in terms of mathematical models. We combine these models with optimal decision theory to yield a classification scheme that better separates positive and negative samples relative to traditional methods such as confidence intervals (CIs) and receiver operating characteristics. We validate the usefulness of this approach in the context of a multiplex salivary SARS-CoV-2 immunoglobulin G assay dataset. This example illustrates how our analysis: (i) improves the assay accuracy, (e.g. lowers classification errors by up to 42% compared to CI methods); (ii) reduces the number of indeterminate samples when an inconclusive class is permissible, (e.g. by 40% compared to the original analysis of the example multiplex dataset) and (iii) decreases the number of antigens needed to classify samples. Our work showcases the power of mathematical modeling in diagnostic classification and highlights a method that can be adopted broadly in public health and clinical settings.
and Infectious Diseases (NIAID) grant R21AI139784 (CDH, NP), National Institute of Environmental Health Sciences (NIEMS) grant R01ES026973 (CDH, NP), NIAID grant R01AI130066 and NIH grant U24OD023382 (CDH), NIAID grant 3R01AI148049 (DLT), the Johns Hopkins University School of Medicine COVID-19 Research Fund, the Sherrilyn and Ken Fisher Center for Environmental Infectious Diseases Discovery Program, and NIH grants U54EB007958-12, US41090366 (YCM). The aforementioned funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. RAL was also funded through the NIST PREP grant 70NANB18H162.

Competing interests: I have read the journal’s policy and the authors of this manuscript have the following competing interests: RAL, AJK, NP, CDH, and PNP have no conflicts of interest to declare. YCM has received consulting fees from Abbott Diagnostics; served leadership or fiduciary roles in the Infectious Diseases Institute, Infectious Diseases Society of America, and WHOTDR; and her institution received reagents from Hologic, Chembio, Roche, Cepheid, and Becton-Dickinson for studies. DLT has received funds for Merck DSMB board membership, consulting fees and stock or stock option for Excision Bio, fees for expert testimony, honors for CME programs only, and royalties from UpToDate.

Abbreviations: 2D, two-dimensional; 3D, three-dimensional; CI, confidence interval; COVID-19, coronavirus disease of 2019; ELISA, enzyme-linked immunosorbent assay; IgG, immunoglobulin G; MFI, median fluorescence intensity; N, nucleocapsid; RBD, receptor binding domain; ROC, receiver operating characteristics; S, spike; SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2.

**Introduction**

Antibody testing has become a crucial public health tool during the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. Interpretation of serology results typically uses mathematical analyses that classify samples as positive, negative, or indeterminate. Conventional classification strategies include confidence interval (CI)-based schemes or receiver operating characteristics (ROC) [1, 2]. The accuracies of both methods suffer from assumptions that fail to account for the structure of data.

Despite having a fundamental impact on classification, this concept of data structure is rarely acknowledged and requires further consideration. For example, CI and ROC-based methods do not address questions such as: how are positive and negative populations clustered? In what direction does each population’s data spread out? What shapes outline the measurement values of each population? Throughout we refer to such ideas as the *structure* of data.

Equally important, neither CIs nor ROC thoroughly characterize the structure of positive and negative populations relative to each other. CIs for negative populations are decoupled from information about positive samples. Specifically, they label measurements outside a fixed number of standard deviations (3σ) from the negative sample mean as positive [3–5]. However, low probability of being negative does not imply high probability of being positive. Moreover, this choice implicitly assumes that 99.7% of negatives fall within the CI, or that a Gaussian model fits the data, which may be unreasonable. ROC is a graphical method with a straightforward interpretation as minimizing error, but it fails to account for how the degree of overlap between positive and negative measurements changes with prevalence. These observations suggest that more accurate classification methods can be realized by explicitly quantifying the structure of data.

One strategy to overcome certain limitations of CIs and ROC is to build mathematical modeling-based classification schemes, which can leverage structural properties of data to improve accuracy. Probability models can be formulated to quantify phenomena, such as: (i) the degree to which positive samples have higher antibody levels than negatives; (ii) statistical correlation of data; and (iii) the outline of data described in terms of shapes like spheres or cones. Several previous works have used mathematical modeling in diagnostic classification [6, 7]. A recent approach applied a combination of modeling and optimal decision theory to antibody testing and proved optimality for binary classification [6]; see Chapter 3 of [8]. Another study built a statistical model for binary classification that accounted for sample bias and used either antibody or viral-load tests [7].

In contrast to previous work, we use high-dimensional mathematical modeling to address the shortcomings of CIs and ROC. We first observe that appropriately increasing the dimension of data—i.e., measuring IgG binding to relevant additional antigens, or, in a multiplex assay, taking advantage of previously unused information—can reveal nuanced structure and better separate positive and negative populations. Next, we construct conditional probability models for the corresponding measurements. We use an optimal classification scheme [6] to improve accuracy and decrease the number of indeterminate test results when such a class is permissible. We validate these ideas by constructing three-dimensional (3D) models for a SARS-CoV-2 immunoglobulin G (IgG) assay [9, 10].

By construction, our 3D modeling classification is more accurate than CI methods, because we both adapt to the structure of the data and use an optimal method. ROC does not have an analog to our results because the technique is unusable beyond one dimension (see [6]). Further, antibody assays can have a narrow linear detection range that interferes with the
interpretation of statistical CIs; in contrast, our method makes no assumption about linearity of the detector, and we can account for such (and related) effects as needed.

A key result of our work is that increased data separation in higher dimensions decreases the number of indeterminate test results while improving classification accuracy. As an unexpected bonus, our higher dimensional work allows us to improve upon the performance in a related work [11] while reducing the number of measurement targets in the analysis from eight to three. We attribute this improvement to the increased data separation of working in 3D and the fidelity of our models to the structure revealed in higher dimensions.

A key goal of this manuscript is to be accessible to the broader community. As such, we provide intuitive examples and highlight core ideas, leaving technical details and most equations for S1 Appendix. In the Materials and methods section, we introduce the multiplex oral fluid (hereafter, salivary) SARS-CoV-2 IgG assay data and provide a straightforward understanding of our 3D models with figures. We display the optimal classification domains and compare our results to CI methods and the original analysis [9] in the Results section; this is followed by further analysis, limitations, and extensions in the Discussion.

**Materials and methods**

**Data introduction**

We consider a multiplex salivary SARS-CoV-2 IgG assay that measures anti-IgG specific to three domains of SARS-CoV-2: nucleocapsid (N) protein, receptor binding domain (RBD), and full spike (S) protein [9, 10]. Participants in the study [9, 10] provided verbal and/or written informed consent. Use of this data has been approved by: (1) the NIST Research Protections Office (IRB no. ITL-2020–0257); and (2) the Johns Hopkins School of Medicine Institutional Review Board (IRB no. IRB00247886).

Measurements are reported as median fluorescence intensities (MFIs). The dataset is separated into training and test populations following the same delineation as the original analysis [9, 10] for which the true classes are treated as known and unknown, respectively. Notably, the original assay used seven SARS-CoV-2 antigen-specific IgG targets; we will show that our modeling achieves better accuracy using only one N and one RBD antigen. We empirically select these antigens as the pair with the greatest separation as defined by the silhouette coefficient [12]. In this context, the silhouette coefficient measures how close a negative sample is to the mean of the negative population relative to how close it is to the mean of the positive population. (Questions of optimal down-selection are reserved for future work.)

Positive samples were confirmed via quantitative polymerase chain reaction tests and were collected more than fourteen days post-coronavirus disease of 2019 (COVID-19) symptom onset. Negative samples were collected prior to the pandemic and have small signals, potentially arising from cross-reactivity with other coronaviruses such as 229E, NL63, OC43, and HKU1. Data in the original analysis [9, 10] were labeled indeterminate if samples classified as SARS-CoV-2 IgG negative contained salivary total IgG less than or equal to 15 μg/mL as measured by an enzyme-linked immunosorbent assay (ELISA). In this way, salivary total IgG was used as a measure of sample adequacy. Additionally, some samples were labeled indeterminate in the original analysis due to low sample volume and instrument error. In contrast, we establish our indeterminate class by a variation of a method that holds out samples with the lowest probability of being correctly classified [11]. This is discussed in the Indeterminate Class subsection.

The two antibody targets are represented as a measurement double \( \hat{r} = (\hat{x}, \hat{y}) \). The measurements are transformed to a logarithmic scale, via the equation

\[
x = \log_2(\hat{x} + 2) - 1,
\]

(1)
with a similar transformation defining the variable $y$. This transformation is a modeling choice that puts the data on the scale of bits and separates the data well. We denote the corresponding variables by $r = (x, y)$. Note that $r$ can be interpreted as an ordered pair in 2D. We plot $N$ vs. RBD in Fig 1a. The positive population is indicated by red Xs and the negatives by blue ♦s. The green box indicates the $3\sigma$ CIs for the negative population, which do not model the correlation structure of the data. Fig 1a shows that in two dimensions some positive and negative samples overlap.

In Fig 1b, adding the ELISA-based salivary total IgG values as a third dimension separates the data by lifting it into 3D. This additional measurement dimension must be related to the assay in a meaningful way to be useful. The three antibody measurements are transformed as above and represented as a measurement triple $r = (x, y, z)$, which can be thought of as a point in 3D space. In doing so, many positively and negatively classified samples are pulled away from each other, significantly reducing overlap. The N and RBD levels of the negative data are lower than the positive values. The positive data are distributed roughly along a diagonal from the origin to the upper right corner, where all three of N, RBD, and the ELISA-based salivary total IgG are large. By eye, the different classes are much better separated in 3D than 2D, although there remains some overlap. Even in 3D, the CIs do not capture the structure of the data. The next section shows how modeling overcomes this problem.

**Probability models**

Our models predict the probability that a known positive or known negative sample yields a triple $r$ of RBD, N, and ELISA-based salivary total IgG measurements. The training data is plotted with the models in Fig 2. Regions of constant color are equal probability contours; the color corresponds to the positive (yellow) and negative (purple) models. The inner, darker volumes are regions of high probability that a given population yields a specific measurement value. Fig 2 shows the probability model contours for positive and negative populations along with the training data. The models quantify the structure in the two populations, which will later allow us to better classify samples. Fig 2 and its corresponding animation (S2 Fig) are powerful illustrations of the model fits to the data and a useful way to understand its structure. We denote the positive model by $P(r)$ and the negative model by $N(r)$. Detailed mathematical descriptions of models are provided in S1 Appendix.
Classification

Denote the prevalence of previously-infected individuals in the population by \( q \); the fraction of uninfected individuals is \( 1 - q \). The probability that a random sample is both positive and has measurement \( r \) is \( qP(r) \); the probability that a random sample is both negative and has measurement \( r \) is \( (1 - q)N(r) \). A measurement equally likely to be positive or negative satisfies the equation:

\[
qP(r) = (1 - q)N(r).
\]

This equation defines a boundary in 3D that is analogous to cutoff values routinely used for single-antigen assays. If the probability of being positive is greater, the measurement is classified as positive. If the probability of being negative is greater, the measurement is classified as negative. This classification scheme maximizes the accuracy, which we define as the prevalence-weighted combination of sensitivity and specificity; see [6] for the objective function (Eq. 5) and justification of these statements.

Indeterminate class

We construct our indeterminate class by holding out samples with the lowest probability of being correctly classified. This concept of local accuracy [11] gives the probability conditioned on a measurement value that the corresponding sample is correctly classified by optimal classification domains. Equations are given in S1 Appendix. The method [11] solved an optimization problem to determine the minimal number of samples to hold out to achieve a desired classification accuracy. We take a simpler approach by identifying a local accuracy threshold up to which we hold out samples so that the empirical classification accuracy reaches a desirable level. In higher dimensions, the modeling allows us to hold out fewer samples than both the original analysis [9] and the local accuracy work [11] while improving classification accuracies. This capitalizes on the data separation abilities of additional measurement dimensions. In the Results section, we describe our optimal classification domains for our example dataset and compare our classification sensitivities, specificities, and accuracies to those of CI methods and the two aforementioned papers.
Results

The models built on the training data are used to create the optimal classification domains; the classification boundary is given by Eq 2. Fig 3 shows the optimal positive (yellow) and negative (purple) classification domains for the SARS-CoV-2 IgG data. See S3 and S4 Figs for animations of Fig 3. The curved boundary between the negative and positive optimal domains reflects the structure and separation of the sample classes.

Our classification accuracy directly depends on this fidelity of our models to the data. Fig 3a shows results for the training data with known prevalence $q$. Our model-based classification distinguishes between high and low salivary total IgG concentration for the same anti-N and anti-RBD IgG level. For example, in the middle of the plot we correctly classify negative samples with high total IgG concentrations and positive samples with low total IgG concentrations. Further, the optimal positive domain captures many positive samples even though it is nearly impossible to separate some negatives and positives that cluster along the diagonal. In contrast, for this dataset a CI categorizes samples with a fixed same N and RBD measurements and a variable ELISA-based salivary total IgG as members of the same class, leading to higher error rates. Our method admits four false positives and 16 false negatives; the CI yields six false positives and 26 false negatives. Despite inaccuracies, the modeling reduces false classifications by 37.5%.

Fig 3b shows the classification scheme applied to the test data using an estimated prevalence $\hat{q}$. This estimated prevalence can be computed using the models without classification (see S1 Appendix and [6]). Our method yields seven false classifications; the CI admits ten. The optimal classification domains are similar to those in Fig 3a due to slight randomness in the prevalence estimate.

Fig 4 shows the optimal classification domains for the models with indeterminate samples excluded in the white domain; S5 and S6 Figs are animations of Fig 4. The indeterminate domain removes many samples in regions of high overlap between populations. Specifically, many positive samples with low N and RBD values are held out as these previously overlapped with many negative samples. The indeterminate domain describes the characteristic minimum separation between positive and negative samples that is needed to classify them with confidence; holding out samples that cannot be classified with sufficient accuracy increases sensitivity and specificity of the remaining data. The CI-based classification is also applied to the samples without indeterminates for comparison.
Fig 4a shows the training data. Our method yields four false negatives and no false positives. In contrast, the CI yields one false positive and four false negatives. Fig 4b shows the test data; the model and CI both admit one false negative but the CI allows five false positives, whereas our method correctly classifies all negative samples.

Our classification results highlight the usefulness of working in higher dimensions while allowing flexible domains/shapes to separate data instead of CIs. Table 1 shows our method.

![Fig 4. Optimal classification domains for the training (a) and test (b) data with a holdout region (white; magenta and cyan markers are indeterminate samples). Positive samples are indicated by red Xs and negatives with blue ♦s. The green boxes drawn in (a) and (b) are the negative sample mean plus 3σ confidence intervals. S5 and S6 Figs show animations.](https://doi.org/10.1371/journal.pone.0280823.g004)

**Table 1. Summary information about the SARS-CoV-2 datasets with sensitivities, specificities, and classification accuracies for training and test data with and without allowing an indeterminate class.** Model and 3σ (relative to negative sample mean) confidence interval results are shown for all samples; the original analysis [9] was conducted on all seven antibody targets and the ELISA-based total IgG without indeterminate samples.

| Data and method | Positive | Negative | Total |
|-----------------|----------|----------|-------|
| **Training samples** | | | |
| All data | Sensitivity (%) | Specificity (%) | Accuracy (%) |
| Model | 131/147, 89.1 | 279/283, 98.6 | 410/430, 95.3 |
| Confidence interval | 121/147, 82.3 | 277/283, 97.9 | 398/430, 92.6 |
| **Inconclusive data excluded** | | | |
| Model | 111/115, 96.5 | 256/256, 100 | 367/371, 98.9 |
| Confidence interval | 111/115, 96.5 | 255/256, 99.6 | 366/371, 98.7 |
| Pisanic et al. (2020) | 111/115, 96.5 | 219/219, 100 | 330/334, 98.8 |
| Patrone et al. (2022) | 115/119, 96.6 | 227/227, 100 | 342/346, 98.8 |
| **Test samples** | | | |
| All data | Sensitivity (%) | Specificity (%) | Accuracy (%) |
| Model | 83/87, 95.4 | 189/192, 98.4 | 272/279, 97.5 |
| Confidence interval | 81/87, 93.1 | 188/192, 97.9 | 269/279, 96.4 |
| **Inconclusive data excluded** | | | |
| Model | 80/81, 98.8 | 163/163, 100 | 243/244, 99.6 |
| Confidence interval | 80/81, 98.8 | 158/163, 96.9 | 238/244, 97.5 |
| Pisanic et al. (2020) | 81/81, 100 | 125/126, 99.2 | 206/207, 99.5 |
| Patrone et al. (2022) | 81/82, 98.8 | 157/158, 99.4 | 238/240, 99.2 |

https://doi.org/10.1371/journal.pone.0280823.t001
applied to multiplex salivary SARS-CoV-2 IgG data for training and test populations [9, 10].
The table compares sensitivity, specificity, and classification accuracy results from our models
to $3\sigma$ CIs from the negative sample means, the original analysis [9], and holdout work [11].

Our model-based classification accuracy (prevalence-weighted sum of sensitivity and specificity) always improves upon that given by CIs. We achieve classification error rates under 1.1\% in two cases, which are not attained by the CIs for any subpopulation. On average, we reduce the classification error by 41.6\%, with an 84.0\% reduction for the test data with an indeterminate class. Overall errors are larger without an indeterminate class; this shows the challenges of working with overlapping positive and negative data.

Table 1 also compares our model-based accuracy to the original analysis [9], which created their own indeterminate class. Using a local accuracy threshold of 99\%, we reduce the number of indeterminate samples by 28.7\% for the training data and by 51.6\% for the test data. The original analysis [9] used the ELISA-based salivary total IgG and all seven antibody targets, although the latter were summed to form a single number. Thus, while the original work [9] required eight dimensions (i.e., measurements), the analysis was projected onto two. In contrast, we only use two antibody targets and the ELISA-based salivary total IgG, treating these as a measurement triple. We improve classification accuracy from 98.8\% to 98.9\% for the training data and from 99.5\% to 99.6\% for the test data. While the improvements are small, we achieve them while using 13.8\% more of the available data. This example highlights the ability of modeling to simultaneously minimize classification errors, use fewer antibody targets, and hold out fewer samples.

In Table 1 we also compare sensitivity and specificity results from our models against corresponding results from $3\sigma$ CIs. Our model sensitivities and specificities match or best those given by CIs. We increase sensitivities from 92.7\% to 95.0\% and specificities from 98.1\% to 99.3\% on average, even including the data in our indeterminate region. Excluding indeterminate data, we achieve 100\% specificity for both the training and test populations, which is not realized by CIs. The number of false positives and false negatives increases when indeterminate samples are considered. However, we still have fewer false classifications as compared to CIs.

Table 1 includes sensitivities and specificities from the original analysis [9]. Our approach matches training data sensitivities and specificities of 96.5\% and 100\%. Test data sensitivity is lowered from 100\% to 98.8\% but specificity is increased from 99.2\% to 100\%. This discrepancy arises from the different objectives in our work versus the original analysis [9]: our optimal classification maximizes the prevalence-weighted sensitivity and specificity, whereas the prior method maximized specificity while maintaining an acceptable sensitivity. As such, our results do not outperform the original analysis for each case, but we improve overall classification accuracy. Further, our method yields acceptable sensitivities and specificities using the indeterminate data in our holdout region; the original analysis excluded those samples.

We also report sensitivity, specificity, and classification accuracy results from holdout work [11] in Table 1 for comparison. We reduce the number of holdout samples by 14.6\% relative to their method. In doing so, we use 25 more training samples while holding fixed the number of misclassified points at four. Further, we use four additional test samples and have one fewer misclassification, thereby reducing test classification error by 50\%. This improvement is surprising given that their original work already significantly decreased the number of holdouts. Like the original analysis of the data [9], that study [11] summed seven of the eight available antibody measurements to yield a single number; our results suggest that considering measurements separately improves performance.
Discussion

High dimensional mathematical modeling is a powerful tool for classification. Adding dimensions can improve data separation and allow models to better leverage the underlying structure, thereby increasing classification accuracy. We illustrate high dimensional modeling with a binary classification of 3D SARS-CoV-2 antibody response measurements. Our 3D modeling yields significant improvements over CI methods, even when indeterminate data are considered. For the examples considered herein, we decrease average classification errors by 41.6%. A combined specificity + sensitivity of at least 150% is desirable [13]; our average (194 ± 4.79)% is close to the perfect 200% and higher than that of CIs (191 ± 7.40)%.

Our work is limited by the inherent subjectivity of selecting mathematical models. In high dimensions, new structures of the data become apparent, which can necessitate more nuanced modeling. However, this does not suggest how to determine when a model is optimal. One approach is to construct a family of models and identify the one with the smallest error rate [6]. Encouragingly, the subjectivity issue may be lessened by adding more data points [14]. Our models are not designed to account for waning antibody levels, which can drop below detection thresholds when measured several months after infection [15]. Further, our models do not differentiate between demographic factors like age or biological sex, although data could be stratified based on these factors. For example, children have better COVID-19 outcomes than adults [16], suggesting the usefulness of modeling their antibody test results as a distinct group.

A significant benefit of our modeling approach is its adaptability to any number of measurement targets and classes. While our work is demonstrated using N, RBD, and ELISA-based salivary total IgG, modeling four targets and beyond is possible. A more challenging extension is to identify which antibody targets minimize classification errors; we anticipate difficulties in visualization and comparison between dimensions. To this end, models could be constructed for all possible antigen and ELISA combinations to determine the result yielding the highest classification accuracy. More generally, our method does not address the problem of antigen down-selection, which is important for assay design. Given the competing needs to understand correlates of protection and estimate prevalence, it is unclear how to define an objective function that yields an optimal choice of antigens. We leave this for future work.

Finally, our classification scheme minimizes average rates of false negatives and false positives, which relate to sensitivity and specificity. The problem could be formulated as a constrained optimization to meet desired sensitivity and specificity targets to create a “rule-out” or “rule-in” test [2].

In conclusion, the usefulness of our procedure is due to the inherent separation of data in additional dimensions and the ability of models to fit the structure of the data, which molds the classification domains to the negative and positive populations. Our high dimensional models achieve superior classification accuracy and have the potential to replace traditional methods.

Supporting information

S1 Appendix. Details of models, prevalence estimation, and local accuracy.
(PDF)

S1 Fig. Animation of Fig 1b. Positive samples are indicated with red Xs and negatives with blue ♦s. The green box gives the negative sample mean plus 3σ confidence interval.
(GIF)
S2 Fig. Animation of Fig 2. 3D probability models plotted along with the training data. Positive samples are indicated with red Xs and negatives with blue ♦s. (GIF)

S3 Fig. Animation of Fig 3a. Optimal classification domains for the training data. Positive samples are indicated by red Xs and negatives with blue ♦s. The green box gives the negative sample mean plus $3\sigma$ confidence interval. (GIF)

S4 Fig. Animation of Fig 3b. Optimal classification domains for the test data. Positive samples are indicated by red Xs and negatives with blue ♦s. The green box gives the negative sample mean plus $3\sigma$ confidence interval. (GIF)

S5 Fig. Animation of Fig 4a. Optimal classification domains for the training data without the holdout region (white; magenta and cyan markers are indeterminate samples). Positive samples are indicated by red Xs and negatives with blue ♦s. The green box gives the negative sample mean plus $3\sigma$ confidence interval. (GIF)

S6 Fig. Animation of Fig 4b. Optimal classification domains for the test data without the holdout region (white; magenta and cyan markers are indeterminate samples). Positive samples are indicated by red Xs and negatives with blue ♦s. The green box gives the negative sample mean plus $3\sigma$ confidence interval. (GIF)

Acknowledgments
The authors wish to thank Dr. Daniel Anderson and Erica Ramsos for useful feedback during preparation of the manuscript.

This work is a contribution of the National Institutes of Standards and Technology and is therefore not subject to copyright in the United States.

Author Contributions
Conceptualization: Rayanne A. Luke, Anthony J. Kearsley, Christopher D. Heaney, Paul N. Patrone.
Data curation: Anthony J. Kearsley, Nora Pisanic, Paul N. Patrone.
Formal analysis: Rayanne A. Luke, Anthony J. Kearsley, Paul N. Patrone.
Funding acquisition: Anthony J. Kearsley, Nora Pisanic, Yukari C. Manabe, David L. Thomas, Christopher D. Heaney, Paul N. Patrone.
Investigation: Nora Pisanic, Yukari C. Manabe, David L. Thomas, Christopher D. Heaney.
Methodology: Rayanne A. Luke, Anthony J. Kearsley, Paul N. Patrone.
Project administration: Anthony J. Kearsley, Paul N. Patrone.
Resources: Anthony J. Kearsley, Christopher D. Heaney, Paul N. Patrone.
Software: Rayanne A. Luke.
Supervision: Anthony J. Kearsley, Paul N. Patrone.
Validation: Anthony J. Kearsley, Paul N. Patrone.

Visualization: Rayanne A. Luke.

Writing – original draft: Rayanne A. Luke, Anthony J. Kearsley, Paul N. Patrone.

Writing – review & editing: Rayanne A. Luke, Anthony J. Kearsley, Nora Pisanic, Yukari C. Manabe, David L. Thomas, Christopher D. Heaney, Paul N. Patrone.

References

1. Jacobson R. Validation of serological assays for diagnosis of infectious diseases. Rev Sci Tech. 1998; 17:469–486. https://doi.org/10.20506/rst.17.2.1115 PMID: 9713892

2. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. Clin Biochem Rev. 2008; 29(Suppl 1):S83. PMID: 18852864

3. Algaissi A, Alfaleh MA, Hala S, Abujamel TS, Alamhoub SA, et al. SARS-CoV-2 S1 and N-based serological assays reveal rapid seroconversion and induction of specific antibody response in COVID-19 patients. Sci Rep. 2020; 10(1):1–10. https://doi.org/10.1038/s41598-020-73491-5 PMID: 33024213

4. Grzelak L, Temmam S, Planchais C, Demeret C, Tondeur L, Huon C, et al. A comparison of four serological assays for detecting anti–SARS-CoV-2 antibodies in human serum samples from different populations. Sci Transl Med. 2020; 12(559):eabc3103. https://doi.org/10.1126/scitranslmed.abc3103 PMID: 32817357

5. Hachim A, Kavian N, Cohen CA, Chin AW, Chu DK, Mok CK, et al. ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. Nat Immunol. 2020; 21(10):1293–1301. https://doi.org/10.1038/s41590-020-0788-0 PMID: 32807944

6. Patrone PN, Kearsley AJ. Classification under uncertainty: data analysis for diagnostic antibody testing. Math Med Biol. 2021; 38(3):396–416. https://doi.org/10.1093/imammb/dqab007 PMID: 34387345

7. Böttcher L, D’Orsogna MR, Chou T. A statistical model of COVID-19 testing in populations: effects of sampling bias and testing errors. Philos Trans R Soc A. 2022; 380(2214):20210121. https://doi.org/10.1098/rsta.2021.0121

8. Rasmussen CE, Williams CK. Gaussian processes for machine learning. vol. 2. MIT press Cambridge, MA; 2006.

9. Pisanic N, Randad PR, Kruczynski K, Manabe YC, Thomas DL, Pekosz A, et al. COVID-19 serology at population scale: SARS-CoV-2-specific antibody responses in saliva. J Clin Microbiol. 2020; 59(1):e02204–20. https://doi.org/10.1128/JCM.02204-20 PMID: 33067270

10. Randad PR, Pisanic N, Kruczynski K, Howard T, Rivera MG, Spicer K, et al. Durability of SARS-CoV-2-specific IgG responses in saliva for up to 8 months after infection. medRxiv. 2021.

11. Patrone PN, Bedekar P, Pisanic N, Manabe YC, Thomas DL, Heaney CD, et al. Optimal Decision Theory for Diagnostic Testing: Minimizing Indeterminate Classes with Applications to Saliva-Based SARS-CoV-2 Antibody Assays. Math Biosci. 2022; 351:108858. https://doi.org/10.1016/j.mbs.2022.108858 PMID: 35714754

12. Kogan J. Introduction to clustering large and high-dimensional data. Cambridge University Press; 2007.

13. Power M, Fell G, Wright M. Principles for high-quality, high-value testing. BMJ Evid Based Med. 2013; 18(1):5–10. https://doi.org/10.1136/eb-2012-100645 PMID: 22740357

14. Schwartz SC. Estimation of probability density by an orthogonal series. Ann Math Stat. 1967; p. 1261–1265. https://doi.org/10.1214/aoms/1177698795

15. Xia W, Li M, Wang Y, Kazis LE, Berlo K, Melikechi N, et al. Longitudinal analysis of antibody decay in convalescent COVID-19 patients. Sci Rep. 2021; 11(1):1–9. https://doi.org/10.1038/s41598-021-96171-4 PMID: 34408200

16. Yuki K, Fujigoi M, Koutsogiannaki S. COVID-19 pathophysiology: A review. Clin Immunol. 2020; 215:108427. https://doi.org/10.1016/j.clim.2020.108427 PMID: 32325252