



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Original Research Article

Optimal decision theory for diagnostic testing: Minimizing indeterminate classes with applications to saliva-based SARS-CoV-2 antibody assays

Paul N. Patrone^{a,*}, Prajakta Bedekar^{a,b}, Nora Pisanic^c, Yukari C. Manabe^d, David L. Thomas^d, Christopher D. Heaney^{c,1}, Anthony J. Kearsley^a^a National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA^b Johns Hopkins University, Department of Applied Mathematics and Statistics, USA^c Johns Hopkins University, Bloomberg School of Public Health, USA^d Johns Hopkins University, School of Medicine, USA

ARTICLE INFO

Keywords:

SARS-CoV-2

Classification

Antibody

Inconclusive

Saliva

Bathhtub-type principle

ABSTRACT

In diagnostic testing, establishing an indeterminate class is an effective way to identify samples that cannot be accurately classified. However, such approaches also make testing less efficient and must be balanced against overall assay performance. We address this problem by reformulating data classification in terms of a constrained optimization problem that (i) minimizes the probability of labeling samples as indeterminate while (ii) ensuring that the remaining ones are classified with an average target accuracy X . We show that the solution to this problem is expressed in terms of a *bathhtub-type principle* that holds out those samples with the lowest *local accuracy* up to an X -dependent threshold. To illustrate the usefulness of this analysis, we apply it to a multiplex, saliva-based SARS-CoV-2 antibody assay and demonstrate up to a 30 % reduction in the number of indeterminate samples relative to more traditional approaches.

1. Introduction

The SARS-CoV-2 pandemic has highlighted the importance of antibody testing as a means to monitor the spread of diseases such as COVID-19 [1,2]. But the widespread deployment of new assays has also revealed fundamental problems in the ability to reliably analyze the corresponding measurements. Early on, this shortcoming was attributed to low prevalence, which made it difficult to distinguish true and false positives [3]. However, it soon became clear that there were deeper issues related to statistical interpretation of raw data, suggesting the need to revisit the underlying theory of diagnostic classification [4–6].

In this context, a fundamental problem arises when many measurements fall near a cutoff used to distinguish positive and negative samples. The probability of correctly classifying these borderline cases hovers near 50%, so that even a small fraction thereof can significantly decrease overall accuracy. A common solution is to define a third, *indeterminate* class for which one cannot draw meaningful conclusions, although this is not always chosen to be near a cutoff [7–13]. While this approach increases the average accuracy for those samples that are classified, it also decreases testing efficiency. Thus, there is a need to develop strategies that balance the construction of indeterminate classes against overall assay performance.

The present manuscript addresses this problem by answering the question: what classification scheme (I) minimizes the fraction of indeterminate samples while (II) correctly identifying the remaining ones with a minimum average accuracy X ? When an indeterminate class is not permitted, common practice categorizes a sample as positive or negative if its measurement value r falls in a corresponding domain D_p^* or D_n^* ; see Fig. 1. Moreover, it was recently shown that these domains can be optimized by solving an unconstrained optimization problem that maximizes accuracy associated with “binary” classification [5]. In contrast, the present work views (I) and (II) as a *constrained optimization* problem, with the size of the indeterminate class being the objective and the desired accuracy recast as a constraint. We show that the solution to this problem extends the binary classification result by constructing the smallest indeterminate class via a bathhtub-type principle applied to D_p^* and D_n^* : one removes from them the measurements with the lowest probability of being correctly classified up to an X -dependent threshold. As a practical matter, this “waterline” bounding the indeterminate domain can be efficiently and accurately estimated via numerical techniques such as bisection, making our result computationally tractable. We provide examples and numerical validation using a saliva-based, multiplex SARS-CoV-2 antibody test; the Appendix provides mathematical proofs of our main results.

* Corresponding author.

E-mail address: paul.patrone@nist.gov (P.N. Patrone).¹ Questions about assay design should be directed to Christopher D. Heaney at cheaney1@jhu.edu.

At the outset and in contrast with traditional methods, it is important to note that concepts such as specificity and sensitivity *per se* are not fundamental quantities of interest in our analysis. As discussed in Section 6, they describe the accuracy of a fixed and subjective classification scheme in two degenerate cases: 0% and 100% prevalence. As such, it is trivial (but useless) to optimize either quantity by assigning all samples to a single class. *Rather, we treat accuracy – defined as a prevalence-weighted, convex combination of sensitivity and specificity – as fundamental, since this naturally interpolates between the aforementioned degenerate cases.* This choice also highlights an important (but often-ignored) fact: the numbers of false positives and false negatives change with prevalence. Thus, sensitivity and specificity may not be useful metrics of assay performance in a setting where a disease is actively spreading. The bathtub-type principle also reveals that these quantities are not mathematically fundamental, since they arise from more granular notions of conditional accuracy.

Ultimately, this analysis leads to the realization that classification accuracy has both a *local* and *global* interpretation, and the interplay between these interpretations is fundamental to both the problem considered herein the general theory of classification.² In particular, one can construct conditional probability density functions (PDFs) $P(\mathbf{r})$ and $N(\mathbf{r})$ of a measurement outcome \mathbf{r} – i.e. a local property – for (known) positive and negative samples. As shown in Ref. [5], these PDFs are necessary to maximize the global accuracy X , since the equation

$$pP(\mathbf{r}) = (1 - p)N(\mathbf{r}) \quad (1)$$

defines the boundary between D_p^* and D_N^* when p is the prevalence. In the present work, we show that $P(\mathbf{r})$ and $N(\mathbf{r})$ also directly define the local accuracy $Z(\mathbf{r})$, and that its global counterpart X is the average value of $Z(\mathbf{r})$. We next observe that the boundary given by Eq. (1) is the set for which $Z = 50\%$, its lowest possible value. The corresponding points are the first to be held out, since they contribute most to the average error.³ Moreover, one sees that systematically removing the least accurate \mathbf{r} yields the fastest increase in the global accuracy for the remaining points. Our bathtub-type principle formalizes this idea.

This intuition also distinguishes our approach from Ref. [5], which considered uncertainty in classification due to effects that are *external* to the assay. In that work, the authors demonstrated that the optimal binary domains become ambiguous when prevalence is only given to within confidence intervals. They showed that this problem can be addressed by holding out samples whose classes were likewise ambiguous. In contrast, our approach defines the indeterminate class as those measurements with the highest *inherent* uncertainty as quantified in terms of local accuracy. In practice, we find that such effects are often more important, especially as several references have addressed issues pertaining to accurate, unbiased prevalence estimation [5,6,14].

From a practical standpoint, the main inputs to our analysis are training data associated with positive and negative samples; thus our approach is compatible with virtually any antibody assay. These data are used to construct the conditional PDFs $P(\mathbf{r})$ and $N(\mathbf{r})$, so that the classification and holdout problems are reduced to mathematical modeling. This is also the key limitation of our approach insofar as such models are necessarily subjective. However, this problem is not unique to our method. Where possible, we incorporate objective information about the measurement process. See Section 4 and Ref. [5] for a deeper discussion of such issues and other limitations.

The remainder of this manuscript is organized as follows. Section 2 reviews key notation and terminology. Section 3 presents the general

² The testing community has largely restricted its attention to global assay properties, since regulatory reporting focuses on assay performance for large populations [2].

³ An interesting corollary of the proofs in Ref. [5] is that $Z \geq 50\%$ for optimally defined classification domains without indeterminates. Thus, we never need consider relative errors less than 50%. See also Section 3 and the Appendix.

theory for defining optimal indeterminate domains. Section 4 illustrates this analysis in the context of a saliva-based, multiplex SARS-CoV-2 saliva assay. Section 5 considers numerical validation of our analysis, and Section 6 concludes with a discussion and comparison with past works. The Appendix provides a proof of our main result and other supporting information.

2. Notation and terminology

One of our primary goals is to provide practical, accessible tools for constructing indeterminate classes. In doing so, we must combine abstract ideas from measure theory with concepts in applied diagnostics; it is not reasonable to assume that any reader will have a background in both. In this section, we provide necessary information so that: (I) an expert in diagnostics can understand and implement our main results (i.e. construct holdout domains) without needing to derive the associated proofs; and (II) a mathematician can verify our work and understand the applied context. Readers with background in either the mathematical notation (Section 2.1) or diagnostic terminology (Section 2.2) may skip the corresponding sections. [We also advise mathematicians to see our definition of a bathtub-type principle in Section 2.1, as it differs from Theorem 1.14 in [15].] We also refer readers to Refs. [16,17] for deeper background on measure theory.

2.1. Mathematical notation and concepts

- By a set, we mean a collection of objects, e.g. measurements or measurement values. By a domain, we typically mean a set in some continuous measurement space; see, e.g., Fig. 1.
- The symbol \in indicates set inclusion. That is, $\mathbf{r} \in A$ means that \mathbf{r} is in set A .
- The symbol \emptyset denotes the empty set, which has no elements.
- The operator \cup denotes the union of two sets. That is, $C = A \cup B$ is the set containing all elements that appear in either A or B .
- The operator \cap denotes the intersection of two sets. That is, $C = A \cap B$ is the set of elements shared by both A and B .
- The operator $/$ denotes the set difference. We write $C = A/B$ to mean the set of all objects in A that are not also in B . Note that in general, $A/B \neq B/A$. Equivalently, A/B can be interpreted as the “subtraction” or removal from A of the elements it shares in common with B .
- The notation $A = \{\mathbf{r} : *\}$ defines the set A as the collection of \mathbf{r} satisfying condition $*$.
- By a *bathtub-type principle*, we mean the solution to a constrained optimization problem that determines an optimal set A^* via a nonlinear inequality of the form $A^* = \{\mathbf{r} : f(\mathbf{r}) \leq F^*\}$ for some constant F^* and function $f(\mathbf{r})$.

Unless otherwise specified, the “size” or measure of a set refers to the probability of a sample falling within that set, i.e. its probability mass. By the same token, we generally avoid using size to describe the actual dimensions (in measurement space) of a domain.

2.2. Notation and concepts from applied diagnostics

- Training data correspond to samples whose true classes are known. In general, training data is used to construct probability models and/or validate an analysis.
- Test data corresponds to samples whose true classes are unknown, or treated as such for purposes of validation. Test data are the objects to which a classification analysis is applied.
- Prevalence is the fraction of samples in a population that are positive. As such, it is the probability that a person picked at random is positive, given no other information.
- Sensitivity S_e (Specificity S_p) is the relative fraction of positive (negative) samples that are classified correctly. We take the common convention that these quantities refer to expectation values or averages.

2.3. Notation specific to the present work

- The non-caligraphic sets D_P and D_N denote positive and negative classification domains in the binary (no-holdout) problem.
- Caligraphic sets \mathcal{D}_P and \mathcal{D}_N are the corresponding domains in the classification problem with indeterminate samples. \mathcal{D}_h is an indeterminate set.
- The shorthand $\mathcal{D} = \mathcal{D}_P \cup \mathcal{D}_N$ is used throughout and denotes the set of all samples that are classified as positive or negative.
- The use of a superscript \star denotes an optimal quantity. For example, D_P^\star is an optimal positive classification domain.

3. Minimum probability indeterminate class

We begin with the mathematical setting underlying classification. Consider an antibody measurement \mathbf{r} , which can be a vector associated with multiple distinct antibody targets. We take the set of all admissible measurements to be Ω . Our goal is to define three domains, \mathcal{D}_P , \mathcal{D}_N , and \mathcal{D}_h associated with positive, negative, and indeterminate (or h for “hold-out”) samples. In particular, we say that a *test sample* \mathbf{r} is positive if it falls inside \mathcal{D}_P (i.e. $\mathbf{r} \in \mathcal{D}_P$), and likewise for the other domains.

We require that these domains have several basic properties to ensure that they define a valid classification scheme. Recalling that $P(\mathbf{r})$ and $N(\mathbf{r})$ are conditional probabilities associated with positive and negative samples, define the measures of a set $S \subset \Omega$ with respect to P and N to be

$$\mu_P(S) = \int_S d\mathbf{r} P(\mathbf{r}) \quad (2a)$$

$$\mu_N(S) = \int_S d\mathbf{r} N(\mathbf{r}). \quad (2b)$$

That is, $\mu_P(S)$ is the probability of a positive sample falling in S , etc. We then require that

$$\mu_P(\mathcal{D}_P \cup \mathcal{D}_N \cup \mathcal{D}_h) = \mu_N(\mathcal{D}_P \cup \mathcal{D}_N \cup \mathcal{D}_h) = 1 \quad (3)$$

and

$$\mu_P(S \cap S') = \mu_N(S \cap S') = 0 \quad (4)$$

when $S \neq S'$, for S, S' chosen from \mathcal{D}_P , \mathcal{D}_N , or \mathcal{D}_h . Eq. (3) states that the probability of any event falling in the positive, negative, or indeterminate domains is one; i.e. any sample can be classified. Eq. (4) states that the probability of a sample falling in more than one domain is zero, i.e. a sample has a single classification.

Given $P(\mathbf{r})$ and $N(\mathbf{r})$, the law of total probability [18] implies that

$$Q(\mathbf{r}) = pP(\mathbf{r}) + (1-p)N(\mathbf{r}) \quad (5)$$

is the PDF that a test sample yields measurement \mathbf{r} , where p is the prevalence.⁴ The quantity $pP(\mathbf{r})$ is the probability that a sample is both positive and yields \mathbf{r} , with a corresponding interpretation for $(1-p)N(\mathbf{r})$. This motivates us to define the total error rate

$$\mathcal{E}[\mathcal{D}_P, \mathcal{D}_N] = \int_{\mathcal{D}_P} d\mathbf{r} (1-p)N(\mathbf{r}) + \int_{\mathcal{D}_N} d\mathbf{r} pP(\mathbf{r}). \quad (6)$$

The terms on the right-hand side (RHS) are the rates of false positives and false negatives (normalized by the number of tests). Eq. (6) treats any misclassification as equally undesirable, but importantly, indeterminates are not considered errors in Eq. (6). Thus, \mathcal{E} so defined is *not* the error rate of the assay restricted to samples that fall only within \mathcal{D}_P and \mathcal{D}_N . The latter is defined as

$$\mathcal{E}_r[\mathcal{D}_P, \mathcal{D}_N] = \frac{1}{p\mu_P(\mathcal{D}) + (1-p)\mu_N(\mathcal{D})} \mathcal{E}[\mathcal{D}_P, \mathcal{D}_N] \quad (7)$$

where $\mathcal{D} = \mathcal{D}_P \cup \mathcal{D}_N$ is the set of all samples not in the indeterminate region. Eq. (7) is a conditional expectation; i.e. it is the average error conditioned on the set of samples that can be classified.

⁴ See Refs. [5,14] for an unbiased method to estimate p without needing to classify.

In Ref. [5] we showed that when the set $\mathcal{X}_{1/2} = \{\mathbf{r} : pP(\mathbf{r}) = (1-p)N(\mathbf{r})\}$ has measure zero and \mathcal{D}_h is the empty set,⁵ \mathcal{E}_r is minimized by the binary classification scheme

$$D_P^\star = \{\mathbf{r} : pP(\mathbf{r}) > (1-p)N(\mathbf{r})\} \quad (8a)$$

$$D_N^\star = \{\mathbf{r} : (1-p)N(\mathbf{r}) > pP(\mathbf{r})\} \quad (8b)$$

for prevalence p . In light of the definition of $Q(\mathbf{r})$, interpretation of Eqs. (8a) and (8b) is straightforward: classify a sample as positive (negative) if the probability of being both positive (negative) and having value \mathbf{r} is greater than the corresponding probability of being negative (positive) and having value \mathbf{r} . [See Chapter 3 of [19] for related ideas.] While D_P^\star and D_N^\star are not the optimal sets for the problem at hand, they play a fundamental role in the analysis that follows. We also note an important corollary that when the $\mathcal{X}_{1/2}$ has non-zero measure, Eqs. (8a) and (8b) are generalized to

$$D_P^\star = \{\mathbf{r} : pP(\mathbf{r}) > (1-p)N(\mathbf{r})\} \cup \hat{\mathcal{Z}}_P \quad (9a)$$

$$D_N^\star = \{\mathbf{r} : (1-p)N(\mathbf{r}) > pP(\mathbf{r})\} \cup \hat{\mathcal{Z}}_N \quad (9b)$$

where $\hat{\mathcal{Z}}_P$ and $\hat{\mathcal{Z}}_N$ are an arbitrary partition of $\mathcal{X}_{1/2}$. The physical interpretation of this generalization is that any point having equal probability of being negative or positive can be assigned to either class without changing the error. In practice, however, classification often reverts to Eqs. (8a) and (8b) as $\mathcal{X}_{1/2}$ has zero measure for many practical PDFs.

In the present work, we assume that there is a desired average accuracy X and that $\mathcal{L} = 1 - \mathcal{E}_r[D_P^\star, D_N^\star] < X$ when all samples are classified. Our goal is to define a minimum probability indeterminate class \mathcal{D}_h^\star and domains \mathcal{D}_P^\star and \mathcal{D}_N^\star for which $\mathcal{L}[\mathcal{D}_P^\star, \mathcal{D}_N^\star] = X$; that is, we wish to hold out the fewest samples so that those remaining are classified with the desired accuracy. Mathematically, we seek to minimize

$$\mathcal{H}[\mathcal{D}_h] = \int_{\mathcal{D}_h} d\mathbf{r} Q(\mathbf{r}), \quad (10)$$

subject to the constraint that

$$p \int_{\mathcal{D}_P} d\mathbf{r} P(\mathbf{r}) + (1-p) \int_{\mathcal{D}_N} d\mathbf{r} N(\mathbf{r}) = X \int_{\mathcal{D}} Q(\mathbf{r}) \quad (11)$$

for $\mathcal{D} = \mathcal{D}_P \cup \mathcal{D}_N$. In light of Eq. (7), this constraint fixes the conditional expectation [20] of the assay accuracy restricted to \mathcal{D} ; i.e. the accuracy of the assay excluding the holdout domain must be X .

To solve this problem, it is useful to introduce several auxiliary concepts. In particular, define the local accuracy of the unconstrained (i.e. no indeterminate), binary classification to be

$$Z(\mathbf{r}, D_P, D_N) = \begin{cases} pP(\mathbf{r})/Q(\mathbf{r}) & \mathbf{r} \in D_P \\ (1-p)N(\mathbf{r})/Q(\mathbf{r}) & \mathbf{r} \in D_N \end{cases} \quad (12)$$

where D_P and D_N cover the whole set Ω up to sets of measure zero; moreover, let $Z^\star(\mathbf{r}) = Z(\mathbf{r}, D_P^\star, D_N^\star)$ be the local accuracy of the optimal solution to the binary problem. Then the solution to the constrained problem given by Eqs. (10) and (11) is

$$\mathcal{D}_h^\star = \{\mathbf{r} : Z^\star(\mathbf{r}) < Z_0(X)\} \cup \mathcal{C}(X) \quad (13a)$$

$$\mathcal{D}_P^\star = D_P^\star / \mathcal{D}_h^\star \quad (13b)$$

$$\mathcal{D}_N^\star = D_N^\star / \mathcal{D}_h^\star \quad (13c)$$

where $Z_0(X)$ is the solution to the equation

$$\int_{\Omega / \{\mathbf{r} : Z^\star(\mathbf{r}) < Z_0\} \cup \mathcal{C}} d\mathbf{r} [Z^\star(\mathbf{r}) - X] Q(\mathbf{r}) = 0, \quad (14)$$

for any set $\mathcal{C} \subset \{\mathbf{r} : Z^\star(\mathbf{r}) = Z_0\}$ satisfying Eq. (14). Proof of this result, as well as the strict interpretation of \mathcal{C} requires significant analysis of

⁵ \mathcal{E} and \mathcal{E}_r are equal when \mathcal{D}_h is the empty set. Note also that one can measure $\mathcal{X}_{1/2}$ with respect to either P or N . This is because the set $\mathcal{X}_{1/2}$ by definition entails that $p\mu_P(\mathcal{X}_{1/2}) = (1-p)\mu_N(\mathcal{X}_{1/2})$.

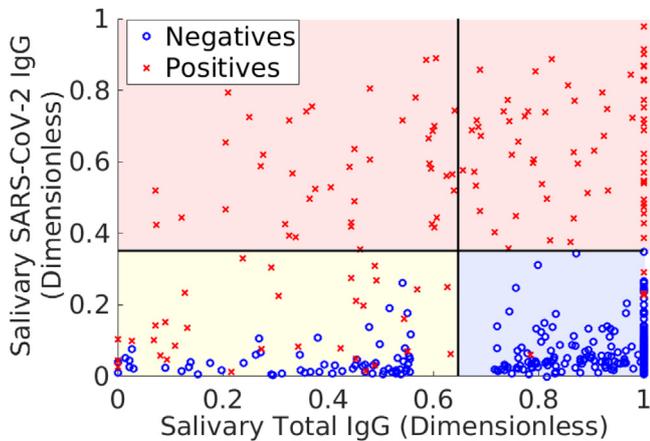


Fig. 1. Training data associated with the Saliva assay described in Refs. [11,12]. Red x denote known positives (confirmed via polymerase chain-reaction measurements), and blue o denote pre-pandemic samples, which are assumed to be negative for SARS-CoV-2 antibodies. The bold, horizontal and vertical black lines are cutoffs used to classify samples. Data falling above the horizontal line (red shaded domain) are classified positive; data in the lower right box (shaded blue) are negative, and data in the lower left box (shaded yellow) are indeterminates. The SARS-CoV-2 IgG measurements (vertical axis) are a sum of seven antibody levels measured by the assay, whereas the total IgG measurement (horizontal axis) is the total immunoglobulin-G (IgG) measurement as determined by an enzyme-linked immunosorbent assay (ELISA).

Eq. (11) and is reserved for Appendix. Here we provide an intuitive interpretation and describe a straightforward algorithm for computing Eqs. (13a)–(13c).

Eq. (13a) informs that the points to label indeterminate are those with the lowest local accuracy up to some threshold value Z_0 , which depends on X . Eqs. (13b) and (13c) then amount to the observations that the positive and negative domains are the same as in the unconstrained binary problem, except that we remove the corresponding points with low enough local accuracy. Eq. (14) requires that the average local accuracy for the classification sets \mathcal{D}_P^* and \mathcal{D}_N^* be X . By virtue of the fact that $\mathcal{D}_h = \Omega/\mathcal{D}$, this fixes the boundary of the indeterminate set. That is, the upper bound $Z_0(X)$ on the indeterminate local accuracy is the lower bound on the accuracy for sets that can be classified. The $\mathcal{E}(X)$ is a bookkeeping artifact accounting for the situation in which the set of points with local accuracy $Z_0(X)$ has non-zero probability mass. In this case, not all of these points need to be held out if doing so would make \mathcal{L} greater than X . The choice of which points to make indeterminate is subjective as they all have the same local accuracy. In practice (e.g. for smooth PDFs), $\mathcal{E}(X)$ is a set of measure zero with respect to Q , so that we can ignore it in Eq. (13a).

From Eqs. (13a)–(14) it is clear that determining $Z_0(X)$ is the key step in defining the optimal classification domains. Fortunately, the interpretation of Eq. (14) leads to a straightforward bisection method. First note that $1/2 \leq Z^*(\mathbf{r}) \leq 1$. Let $\zeta_0 = 3/4$ be an initial guess for the value of $Z_0(X)$, and let ζ_j be the j th update computed iteratively as follows. For each ζ_j compute $\mathcal{D}_P(\zeta_j)$, $\mathcal{D}_N(\zeta_j)$, as well as the left-hand side (LHS) of Eq. (14), which we denote by \mathcal{F}_j . If $\mathcal{F}_j > 0$, then set $\zeta_{j+1} = \zeta_j - 2^{-(j+3)}$; if $\mathcal{F}_j < 0$, set $\zeta_{j+1} = \zeta_j + 2^{-(j+3)}$. If $|\mathcal{F}_j| \leq \epsilon_X$ for some user-defined tolerance ϵ_X , or if j reaches some maximum iteration number M , stop the algorithm. In the former case, the classified samples will have an average accuracy \mathcal{L} in the range $X - \mathcal{O}(\epsilon_X) \leq \mathcal{L} \leq X + \mathcal{O}(\epsilon_X)$. In the latter case, $\zeta_j - \mathcal{O}(\epsilon_Z) \leq Z_0(X) = \zeta_j + \mathcal{O}(\epsilon_Z)$, where $\epsilon_Z = 2^{-M+3}$ is the error in the estimate of $Z_0(X)$. For context, 20 iterations of this algorithm yields errors ϵ_Z on the order of 1 in 10^7 . In the second case, the existence of a non-trivial set $\mathcal{E}(X)$ can be deduced from the observation that \mathcal{F}_j does not converge, but rather cycles between two well-separated values, depending on whether ζ is greater than or less than $Z_0(X)$. In this case, the set $\mathcal{E}(X)$ can be defined arbitrarily but consistent with Eq. (14) once $Z_0(X)$ is identified to sufficient accuracy.

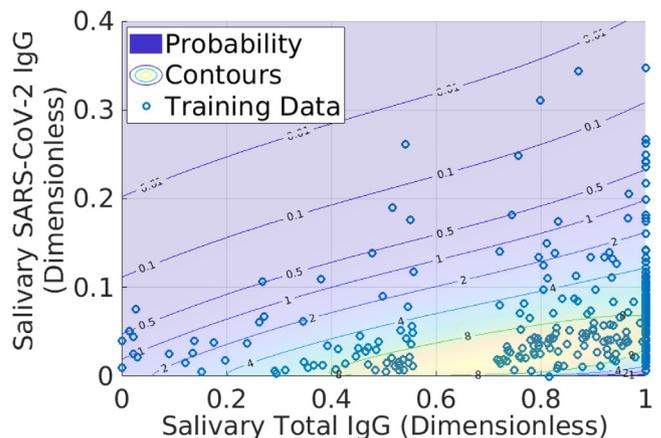
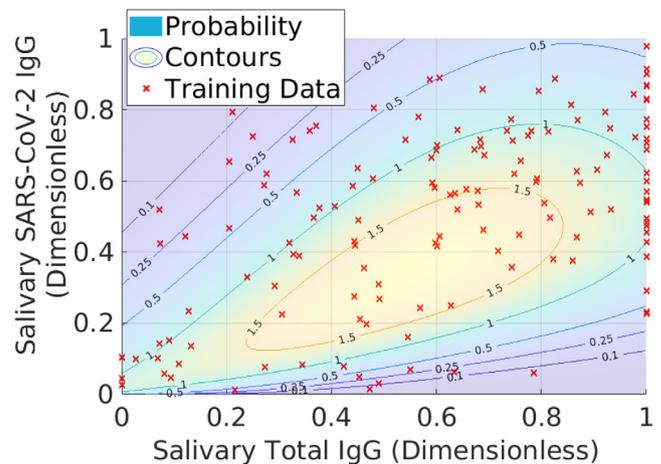


Fig. 2. Probability models associated with the training data. See main text for a description of the probability density functions and the considerations behind their construction. Top: Raw data and probability model for positive training samples. Bottom: Negative training data and probability model.

(In practice and given the speed of convergence, we find that there is little value in considering starting points other than $\zeta_0 = 3/4$.)

4. Example applied to a salivary SARS-CoV-2 IgG Assay

To illustrate the analysis of Section 3, we consider a saliva-based assay described in Refs. [11,12]. We refer the reader to those manuscripts for details of assay design, sample preparation, and measurement processes. For each sample, two measurement values are output: a total immunoglobulin G (IgG) enzyme linked immunosorbent assay (ELISA); and a sum of seven SARS-CoV-2 IgG measurements associated with distinct antigen targets. As a preliminary remark, we observe that the numerical range of the data spans several decades of median fluorescence intensity (MFI), which is difficult to model directly. We also note that the measurements are bounded from below by zero and have a finite upper bound. This motivates us to transform each numerical value d via $\log_2[d+2]-1$, which puts the data on the scale of bits. Empirically we also find that this transformation better separates positive and negative populations. Total IgG values are then rescaled to the domain $[0, 1]$ by dividing each measurement by the maximum. SARS-CoV-2 measurements are similarly rescaled to the domain $[0, 1]$, although we divide the log-transformed data by 7, since there were no samples with saturated values. After transformation, each sample is represented by a two-dimensional vector $\mathbf{r} = (x, y)$, where x is the normalized total IgG value, and y is the normalized SARS-CoV-2 counterpart.

The results of this transformation are shown in Fig. 1, along with classification domains currently used with this assay.⁶ The goal of the analysis is to maintain accuracy while decreasing the number of indeterminate samples by finding the domain \mathcal{D}_h with the smallest probability mass. We remind the reader that size does not refer to the (generalized) volume in measurement space. Rather it refers to the fraction of samples expected to fall within the domain, since this is what controls the number of indeterminate samples. Thus, it is possible that \mathcal{D}_h can be quite large when expressed in terms of antibody levels and yet contain very few samples.

To motivate our probability models, we consider the phenomena that could affect measurements. In particular, we anticipate that for positive samples, there should be a degree of correlation between total IgG and SARS-CoV-2 specific antibodies. However, at extreme total IgG values, the SARS-CoV-2 levels may become independent as (i) all measurements will revert to noise when $x \rightarrow -\infty$ or (ii) SARS-CoV-2 antibody levels will decouple from total antibody levels when the latter is excessively high, e.g. if an individual has been exposed to a large number of different pathogens. We also recognize that the ELISA instrument only reports numerical values on the domain $[x_{\min}, x_{\max}]$. Thus, fluorescence levels above x_{\max} are rounded down to the upper bound, and levels below x_{\min} are rounded up to the lower bound. As shown in Fig. 1, this has the effect of accumulating data (and thus probability mass) on the lines $x = x_{\min}$ and $x = x_{\max}$, which is a manifestation of data censoring [21,22]. While details are reserved for the Appendix, this observation leads us to model positive and negative samples via a PDF of the form

$$P(x, y) = \mathcal{P}_0(x, y) + \mathcal{P}_l(y)\delta(x) + \mathcal{P}_r(y)\delta(x - 1), \quad (15)$$

where $0 \leq x \leq 1$, $0 \leq y < 1$, $\delta(x)$ is the Dirac delta function, and $\mathcal{P}_0(x, y)$ is assumed to be bounded and continuous on the whole domain. The functions $\mathcal{P}_l(y)$ and $\mathcal{P}_r(y)$ characterize the probability of SARS-CoV-2 antibody levels for measurement values saturated at the left (l) and right (r) bounds. We emphasize that the use of delta functions in Eq. (15) is formal and should be treated with care. A more rigorous interpretation of what is meant by Eq. (15) is discussed in Appendix.

To model the function $\mathcal{P}_0(x, y)$, we treat the total IgG measurements as independent normal random variables with an unknown mean and variance. Within the domain $0 < x < 1$ (note the strict inequalities) and $0 \leq y \leq \infty$, we assume that the SARS-CoV-2 measurements are well described by a Gamma distribution with a fixed (but unknown) scale factor and shape parameter with a sigmoidal dependence on x . This dependence is motivated by the correlation described previously. Taken together, this yields the PDF

$$\mathcal{P}_0(x, y) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma}} y^{k(x)-1} \frac{e^{-y/\theta}}{\Gamma(k(x))\theta^{k(x)}} \quad (16a)$$

$$k(x) = \alpha_1^2 [\tanh(\alpha_2(x - \alpha_3)) + 1] + \alpha_4^2 \quad (16b)$$

where μ , σ , θ , and the α_j are to-be-determined. The boundary functions are defined to be

$$\mathcal{P}_l(y) = \frac{y^{k(0)-1} e^{-y/\theta}}{\Gamma(k(0))\theta^{k(0)}} \int_{-\infty}^0 dx \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma}} \quad (17)$$

$$\mathcal{P}_r(y) = \frac{y^{k(1)-1} e^{-y/\theta}}{\Gamma(k(1))\theta^{k(1)}} \int_1^{\infty} dx \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma}} \quad (18)$$

which describes the probability that a total IgG value below (above) $x = 0$ ($x = 1$) will be mapped back to the lower (upper) instrument

⁶ All data correspond to samples for which more than 14 days have elapsed since symptoms onset. Also, the original training total IgG data included samples that were diluted to achieve measurement values above a saturation value. All such data were rounded down to the undiluted upper threshold to be consistent with the validation data. This amounts to data censoring, for which we can still define the relevant likelihood functions used in parameter estimation. See also Appendix.

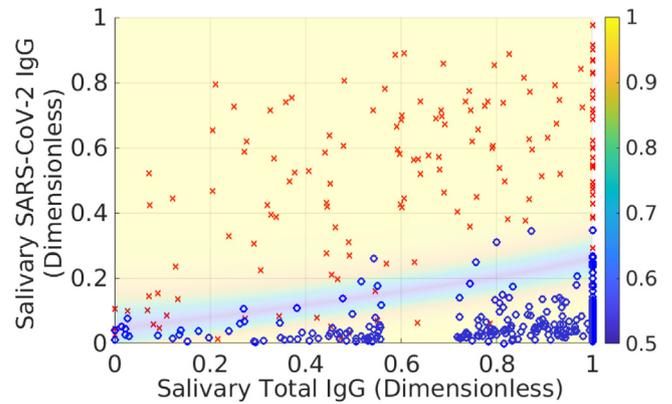


Fig. 3. Local accuracy $Z^*(r)$ of the assay according to the probability models shown in Fig. 2. Note that $Z^*(r)$ approaches 100% in regions where $P(r)$ and $N(r)$ do not overlap. Conversely, in regions where the PDFs overlap, it is more challenging to correctly identify samples. Thus $Z^*(r)$ decreases towards its minimal value of 1/2 in such regions. Note that $Z^*(r)$ is never less than 1/2 (50/50 odds of correct classification).

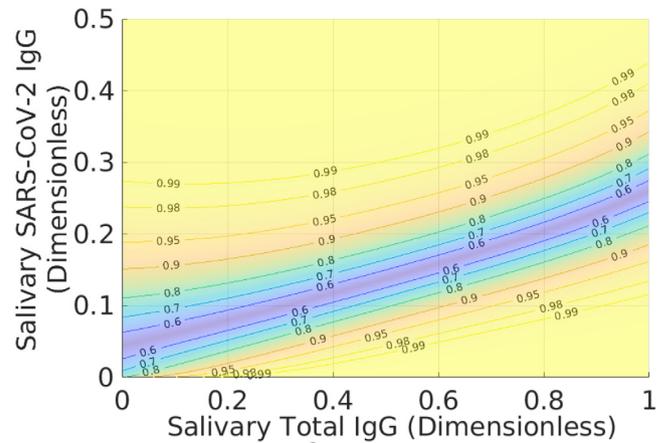


Fig. 4. Illustration of the bathtub-type principle used to compute the minimal probability indeterminate domain. The contour lines are different “waterlines” up to which we can hold out samples. The label on each contour is the local accuracy of the assay. In order to define the indeterminate region, we use the target global accuracy X to define a maximum local accuracy up to which we hold out samples. Increasing the global accuracy of the restricted classification increases the waterline, thereby holding out more samples.

bound. The free parameters are determined via maximum likelihood estimation using a censoring-based technique; see the Appendix and Refs. [21,22]. As an approximation, we truncate the y -domain to be $0 \leq y \leq 1$ and renormalize the resulting PDF on this domain.

For the negative PDF $N(x, y)$, we anticipate that non-specific binding of the total IgG antibodies to the SARS-CoV-2 antigens will lead to a degree of correlation, albeit to a less extent than for positives. Thus, we use the same form of $P(x, y)$, but refit the parameters using the negative training data. Fig. 2 shows the outcome of this exercise for the two training sets. Because $\mathcal{P}_l(y)$, $\mathcal{P}_r(y)$, and corresponding terms for $N(x, y)$ are continuous with respect to the Gamma portion of $P(x, y)$ and $N(x, y)$, the former can be inferred from the contour lines in the figure (up to a normalization factor) and are thus not shown.

Figs. 3 and 4 show $Z^*(r)$ and waterlines necessary to achieve different average accuracies. The bathtub-type principle is shown in the latter; see also Ref. [15] for related ideas. To ensure that $\mathcal{L} = X$, we only hold out samples up to the corresponding value of $Z_0(X)$. Note that indeterminates are concentrated in regions where there is significant overlap between positive and negative samples. Fig. 5 shows the corresponding classification domains computed according to the

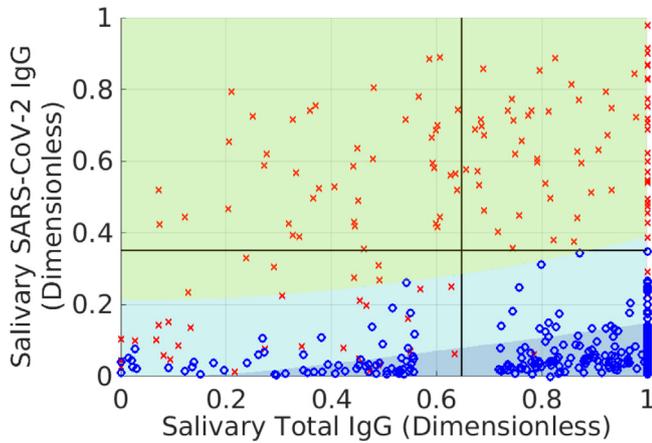


Fig. 5. Positive (yellow-green), negative (dark blue), and indeterminate (light-blue) classification domains defined for a theoretical target accuracy of 99.6% for the training data in the previous figures. Symbols have the same meaning as in previous figures. The empirical accuracy is 98.8%, with a specificity of 100% and sensitivity of 96.7%. The total accuracy is the prevalence-weighted combination of these latter quantities. Note the prevalence is associated with the restricted set of samples that are actually classified; see Section 6. Discrepancy between the theoretical and empirical accuracies is due to idealization of the modeling and stochasticity in the data. For comparison, the horizontal and vertical black lines are the same as in Fig. 1 and denote the corresponding classification domains originally used for this assay. The indeterminate region based on the bathtub-type principle reduces the number of unclassified samples by more than 12% relative to the original domains while maintaining specificity and improving sensitivity for the training data. See also Table 1 and Section 6 for other examples of indeterminate domains.

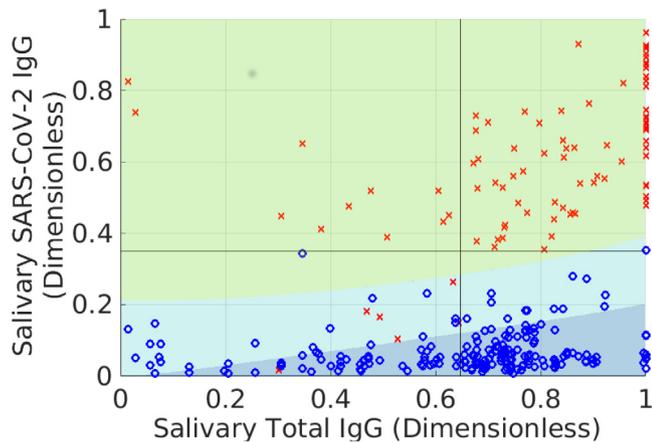


Fig. 6. Positive (yellow-green), negative (dark blue), and indeterminate (light-blue) classification domains for validation data and defined for a theoretical target accuracy of 99.6%. The validation data was not used for training the probability models. Symbols have the same meaning as in previous figures. The empirical accuracy is 99.2%, with a specificity of 99.4% and sensitivity of 98.8%. The indeterminate region based on the bathtub-type principle reduces the number of unclassified samples by almost 40%. See Section 6 for other examples of indeterminate domains.

bathhtub-type principle for a target accuracy of 99.6%; see also Table 1. Relative to the original classification domains, the analysis reduces the empirical rate of indeterminate samples by more than 12% while increasing both accuracy and sensitivity of the assay (with empirical specificity remaining constant). See also Fig. 6 and Section 6 for additional examples of holdout domains.

5. Numerical validation

To validate that the sets \mathcal{D}_p^* , \mathcal{D}_N^* , and \mathcal{D}_h^* obtained in Section 3 are optimal, we consider a numerical experiment wherein we perturb \mathcal{H} as

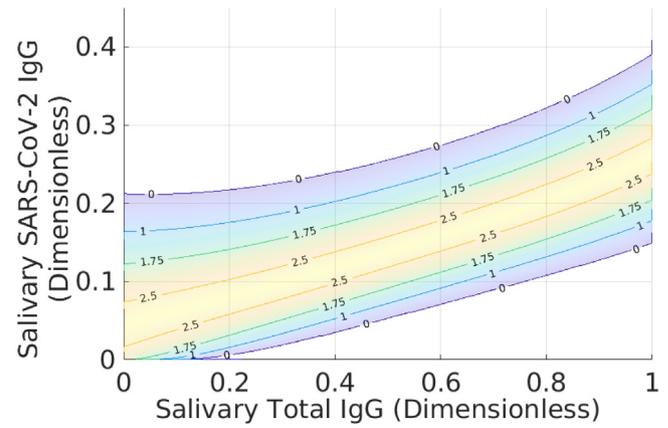


Fig. 7. The logarithm of the swap derivative given by Eq. (19) computed for the optimal domains \mathcal{D}_p^* , \mathcal{D}_N^* , and \mathcal{D}_h^* . The 0-level line is the boundary of the indeterminate region. Note that the logarithm is everywhere positive. Thus, swapping any infinitesimal regions between \mathcal{D}_h^* and \mathcal{D}^* will increase the probability mass in the indeterminate, provided constraint Eq. (11) is satisfied.

a function of these domains. For point $\mathbf{r} \in \mathcal{D}_h$ and $\mathbf{r}' \in \mathcal{D}$, we formally define a “point-swap derivative” to be

$$\frac{\delta \mathcal{H}[\mathcal{D}_h]}{\delta \mathbf{r} \delta \mathbf{r}'} = \frac{Z(\mathbf{r}) - X}{Z(\mathbf{r}') - X}. \quad (19)$$

In principle $Z(\mathbf{r})$ can be an arbitrary definition of local accuracy, although in practice we take $Z(\mathbf{r}) = Z^*(\mathbf{r})$ in this section. The interpretation of Eq. (19) is as follows. In taking point \mathbf{r}' from \mathcal{D} and adding it to \mathcal{D}_h and vice-versa for \mathbf{r} , we must ensure that the constraint Eq. (11) remains satisfied. The ratio $\frac{Z(\mathbf{r}) - X}{Z(\mathbf{r}') - X}$ provides the “rate-of-exchange” of probability. For example, if $Z(\mathbf{r}) - X < Z(\mathbf{r}') - X < 0$, then adding \mathbf{r} to \mathcal{D} will infinitesimally decrease the global accuracy, so that we must hold out a larger yet still infinitesimal fraction of Q in the vicinity of \mathbf{r}' . It is clear that Eq. (19) goes through a singularity when $Z(\mathbf{r}') \rightarrow X$ and becomes negative for $Z(\mathbf{r}') > X$ and $Z(\mathbf{r}) < X$. The interpretation of this is straightforward: we should always reverse any swap for which a point with local accuracy greater than the average is put in the indeterminate class. Such points are not considered in the analysis below. More rigorous interpretations of Eq. (19) are considered in the Appendix, especially in the context of the singular PDF given by Eq. (15).

The benefit of Eq. (19) is that it allows us to estimate a “set-partial derivative” by computing the relative probability exchange for any point in the indeterminate domain. In particular, we compute

$$\frac{\delta \mathcal{H}[\mathcal{D}_h]}{\delta \mathbf{r}} = \inf_{\substack{\mathbf{r}' \in \mathcal{D} \\ Z(\mathbf{r}') < X}} \left[\frac{Z(\mathbf{r}) - X}{Z(\mathbf{r}') - X} \right] \quad (20)$$

for the optimal domains \mathcal{D}_h^* and \mathcal{D}^* . Fig. 7 shows the logarithm of Eq. (20) for a mesh of points in the indeterminate region, taking $Z(\mathbf{r}) = Z^*(\mathbf{r})$. Note that swapping any point in the indeterminate region with one in the positive and negative classification domains increases the size of the indeterminate, as expected.

To validate that swapping points between \mathcal{D}_p^* and \mathcal{D}_N^* does not increase the accuracy of the assay or decrease the size of the indeterminate domain, we examine the quantity $Z(\mathbf{r})$ directly. In particular, the Appendix shows that $Z^*(\mathbf{r}) \geq 1/2$ for all $\mathbf{r} \in \mathcal{D}^*$ guarantees that $\mathcal{D}_p^* = D_p^*/\mathcal{D}_h^*$ and $\mathcal{D}_N^* = D_N^*/\mathcal{D}_h^*$ are optimal for the indeterminate region \mathcal{D}_h^* . Fig. 3 demonstrates that this inequality holds for the solution given by Eqs. (13a)–(14). Thus, no rearrangement of points decreases the size of the indeterminate domain.

Table 1

Summary of fraction of holdouts, sensitivity, and specificity for the data in Figs. 5 and 6. The rectilinear classification method is described in Fig. 1, while the optimal method is given by Eqs. (13a)–(14). For sensitivity, specificity, and accuracy calculations, the numbers in brackets are empirical 95% confidence intervals.

| Data & Method | COVID samples | Pre-COVID samples | All samples |
|------------------------------|----------------------------------|---------------------------------|----------------------------------|
| Training samples | 147 | 283 | 430 |
| Holdouts (Rectilinear) | 32/147, 21.8% | 64/283, 22.6% | 96/430, 22.3% |
| Holdouts (Optimal) | 28/147, 19.1% | 56/283, 19.8% | 84/430, 19.5% |
| | Sensitivity | Specificity | Accuracy |
| Classification (Rectilinear) | 111/115, 96.5% [92.0%, 98.9%] | 219/219, 100% [98.6%, 100%] | 330/334, 98.8% [97.2%, 99.6%] |
| Classification (Optimal) | 115/119, 96.6% [92.3%, 99.0%] | 227/227, 100% [98.7%, 100%] | 342/346, 98.8% [97.3%, 99.6%] |
| Validation samples | 87 | 192 | 279 |
| Holdouts (Rectilinear) | 6/87, 6.9% | 66/192, 34.4% | 72/279, 25.8% |
| Holdouts (Optimal) | 5/87, 5.8% | 34/192, 17.7% | 39/279, 14.0% |
| | Sensitivity | Specificity | Accuracy |
| Classification (Rectilinear) | 81/81, 100% [96.3%, 100%] | 125/126, 99.2% [96.3%, 100%] | 206/207, 99.5% [97.7%, 100%] |
| Classification (Optimal) | 81/82, 98.8% [94.4%, 100.0%] | 157/158, 99.4% [97.0%, 100%] | 238/240, 99.2% [97.3%, 99.9%] |

6. Discussion: Historical context, open direction, and limitations

6.1. The role of prevalence

Examination of Eq. (11) reveals that the terms of the LHS are proportional to prevalence-weighted estimates of sensitivity and specificity. In particular, recognize that

$$S_e = \left[\int_{\mathcal{D}} P(\mathbf{r}) \, d\mathbf{r} \right]^{-1} \int_{\mathcal{D}_p} P(\mathbf{r}) \, d\mathbf{r}, \tag{21a}$$

$$S_p = \left[\int_{\mathcal{D}} N(\mathbf{r}) \, d\mathbf{r} \right]^{-1} \int_{\mathcal{D}_N} N(\mathbf{r}) \, d\mathbf{r} \tag{21b}$$

are the sensitivity and specificity restricted to the domain \mathcal{D} . When there is no indeterminate domain, the normalization factors $\int_{\mathcal{D}} P(\mathbf{r}) \, d\mathbf{r} = \int_{\mathcal{D}} N(\mathbf{r}) \, d\mathbf{r} = 1$, so that Eqs. (21a) and (21b) revert to the standard definitions of these quantities. In this case, we see that Eq. (11), which no longer acts as a constraint, amounts to the statement that the prevalence-weighted sum of sensitivity and specificity is equal to X ; that is

$$pS_e + (1-p)S_p = X. \tag{22}$$

When we permit an indeterminate class, however, the interpretation is not as straightforward. In particular, the presence of the term $n_Q = \int_{\mathcal{D}} Q(\mathbf{r}) \, d\mathbf{r}$ on the right-hand side (RHS) appears problematic, for note that it implies

$$n_Q^{-1} \left[p \int_{\mathcal{D}_p} P(\mathbf{r}) \, d\mathbf{r} + (1-p) \int_{\mathcal{D}_N} N(\mathbf{r}) \, d\mathbf{r} \right] = X. \tag{23}$$

The normalization factor n_Q differs from its counterparts in Eqs. (21a) and (21b). Thus, it is not obvious what our constraint enforces about the sensitivity and specificity of the assay restricted to \mathcal{D} .

The resolution to this conundrum is to recognize that the prevalence of the population also changes when we restrict classification to \mathcal{D} . This is not to say that the value of p itself (i.e. associated with the total population) changes, but rather that the relative fraction of positives and negatives differs on $\mathcal{D} \subset \Omega$. This is not unexpected, since the shape of the indeterminate region is a function of the local accuracy Z , which depends on the specifics of the probability models. Mathematically, we understand these observations by rewriting Eq. (23) in the form

$$\frac{pn_p}{n_Q} \int_{\mathcal{D}_p} \frac{P(\mathbf{r})}{n_p} \, d\mathbf{r} + \frac{(1-p)n_N}{n_Q} \int_{\mathcal{D}_N} \frac{N(\mathbf{r})}{n_N} \, d\mathbf{r} = X \tag{24}$$

where $n_p = \int_{\mathcal{D}_p} P(\mathbf{r}) \, d\mathbf{r}$ and $n_N = \int_{\mathcal{D}_N} N(\mathbf{r}) \, d\mathbf{r}$ are the required normalization constants. Eq. (24) becomes an analogue to Eq. (22) of the form

$$\frac{pn_p}{n_Q} S_e + \frac{(1-p)n_N}{n_Q} S_p = X \tag{25}$$

where $p_{\mathcal{D}} = pn_p/n_Q$ is the prevalence restricted to the domain \mathcal{D} . Note that $p_{\mathcal{D}}$ has the properties necessary to be a prevalence:

$$\frac{pn_p}{n_Q} + \frac{(1-p)n_N}{n_Q} = 1 \implies 1 - p_{\mathcal{D}} = \frac{(1-p)n_N}{n_Q} \tag{26}$$

which is a consequence of the definition of n_Q . Thus, we see that the constraint corresponds to a domain-restricted-prevalence weighted sum of sensitivity and specificity.

From a theoretical standpoint, Eq. (26) is extremely serendipitous. The constraint as defined by Eq. (11) only refers to the prevalence of the full population. It is not obvious that this equation will remain a prevalence-weighted sum when holding out samples, especially as the restricted-prevalence does not in general equal p . Further implications of this observation are explored in the next section.

However, an immediate practical consequence of Eq. (26) is that *the relative fraction of positives from an assay using indeterminates is not a reliable estimator of total prevalence*. In order for the restricted prevalence $p_{\mathcal{D}}$ to equal p , one requires

$$n_p - n_Q = 0 = \int_{\mathcal{D}} P(\mathbf{r}) - pP(\mathbf{r}) - (1-p)N(\mathbf{r}) \, d\mathbf{r},$$

which implies

$$0 = \int_{\mathcal{D}} P(\mathbf{r}) - N(\mathbf{r}) = n_p - n_N. \tag{27}$$

That is, $p = p_{\mathcal{D}}$ only occurs when the holdout domain removes equal mass from the probability models, which is extremely restrictive.

To overcome this problem, we recall that Ref. [5], demonstrated how *an unbiased estimate of the total prevalence can be constructed without classifying samples* using a simple counting exercise on subdomains of Ω . The validity of that method is independent of the assay accuracy, so that it can be used to estimate p in the present work. Indeed, such techniques are necessary to construct the optimal classification domains, given the fundamental role of p in their definitions. We refer the reader to Ref. [5] for a deeper discussion of such issues.

6.2. Other notions of optimality

A common practice in the testing community is to preferentially optimize an assay so that either the specificity or sensitivity reaches a desired target, but not explicitly a linear combination of the two. Eq. (25) and the bathtub-type principle suggest a route by which our method can solve an analogue of this problem. However, a deeper investigation of sensitivity and specificity is first necessary to motivate this generalization and understand how such methods differ from Eqs. (13a)–(14). [See also Ref. [23] for additional notions of optimality, as

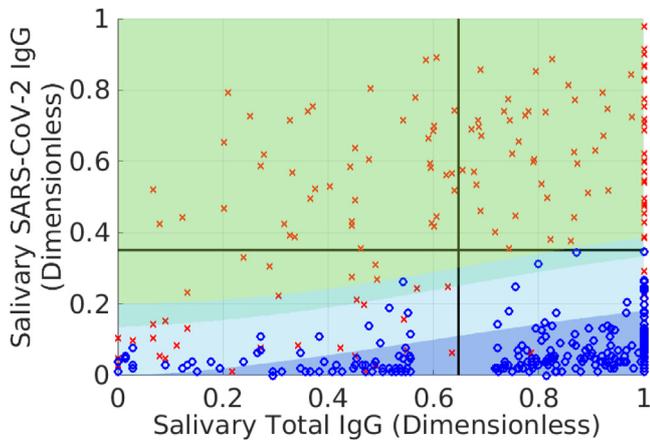


Fig. 8. Holdout domain computed with a target accuracy of 99% and according to the constraints given by inequalities (29) and (30). For the latter, we set X_- indirectly by holding out samples up to $Z_p = 0.972$ in the positive classification domain. This yields an empirical specificity of the training data was 100% while keeping the empirical sensitivity above 94%. Note that the indeterminate domain (light-blue) is increased only into the positive classification domain (yellow-green) in attempting to satisfy inequality (30). The teal strip adjacent to the light blue and yellow-green is the modified indeterminate domain. After increasing the empirical specificity to 100%, the optimized domains holds out 15.1% of samples, as opposed to 22.3% for the rectilinear method; see Table 1.

well as Refs. [24–26] for other approaches to defining classification domains.]

Examination of the binary problem reveals that when $p = 1/2$, the domains D_p^* and D_N^* equally weight sensitivity and specificity; that is, errors in either are treated as equally undesirable. It is straightforward to show that increasing p will increase sensitivity at the expensive of specificity, and vice versa. The interpretation of this observation is that as the number of positive samples increases, we should increase the size of the positive classification domain so as to capture the their increasing share of the population. It is therefore possible and even likely that when the prevalence approaches 0 or 100%, either sensitivity or specificity may be unacceptably low, since the corresponding contribution to the total accuracy becomes negligible.

A possible solution to this problem is to recast Eq. (11) as an inequality constraint of the form

$$p \int_{\mathcal{D}_p} dr P(\mathbf{r}) + (1 - p) \int_{\mathcal{D}_N} dr N(\mathbf{r}) \geq X \int_{\mathcal{D}} Q(\mathbf{r}) \quad (28)$$

together with the additional constraints

$$S_e \geq X_+ \quad (29)$$

$$S_p \geq X_- \quad (30)$$

where X_+ and X_- are user-defined lower bounds. While an optimal solution to this problem is beyond the scope of the current manuscript, the bathtub-type principle suggests a construction akin to active-set methods [27]. First, solve the optimization problem associated with Eqs. (10)–(11) and check the resulting values of sensitivity and specificity. If these quantities are deemed to small, remove samples up to user-defined waterlines $Z_n \geq Z_0$ and $Z_p \geq Z_0$ (which may be different), where Z_n and Z_p apply only to samples in the negative and positive classification domains. Fig. 8 shows an example of this approach applied to the data in previous figures. We originally set $X = 0.99$ but required that the empirical specificity be 100% for the training set. To accomplish this, we set $Z_p = 0.972$, which augments the size of the indeterminate domain (teal strip added to the light blue domain) without decreasing the number of true negatives.

6.3. Relationship between prevalence, sensitivity, and specificity

Eq. (25) and the examples of Secs. 6.1 and 6.2 beg the question: to what extent is prevalence-weighted accuracy a preferred or natural framework for diagnostic classification, as opposed to methods based on explicit reference to sensitivity and specificity? To unravel this, consider that the latter two are purely theoretical properties of a specific choice of classification domain and are only loosely connected to the reality of testing. This is evident from the definitions given by Eqs. (21a) and (21b). The concept of prevalence, *i.e. implying existence of a population*, does not enter; rather all that is needed is a choice of the classification domains. Thus, an assay can have exceptional sensitivity and yet still be wrong half the time if the prevalence is low. In a related vein, it is clear that specificity and sensitivity only characterize assay accuracy in the limits $p \rightarrow 0$ and $p \rightarrow 1$, respectively.

Here we encourage a new perspective. As a baseline strategy, the most important task is to correctly classify samples; at least this is of the utmost importance to patients. Moreover, computing accurate prevalence estimates is critical for epidemiologists (although we have shown previously that this problem is solved accurately without recourse to classification). With this goal in mind, the sensitivity and specificity are subservient to accuracy via Eq. (11), and it is not unreasonable to let them change with prevalence if doing so increases overall testing accuracy. We highlight this because under such a paradigm, S_e and S_p lose their status as the key performance metrics that define the “quality” of an assay, and they cannot be viewed as static properties. Such observations are not to say that S_e and S_p are useless, however. Clearly there are times when it is more important to correctly identify samples from one class, and this motivates the generalization of Section 6.2.

But these observations clarify our perspective of why the prevalence sets a natural scale for classification. In particular, Eq. (11) has two equivalent interpretations: (i) the accuracy of the assay must be X ; and (ii) the prevalence-weighted sensitivity and specificity must be X . The equivalence of these interpretations arises from the fact that *notions of accuracy assume the existence of a population to which the test is applied*. Thus, Eq. (25) is perhaps unsurprising in light of Eq. (11) because both are self-consistent statements about the properties of a population.

The benefit of treating prevalence-weighting as a natural framework for diagnostic classification is that one can easily identify when subjective elements (*i.e. not intrinsic to the population*) have been added to the analysis. For example, the indeterminate domain in Fig. 8 associated with the inequalities (28)–(30) is not optimal insofar as there is a smaller counterpart that yields the same average accuracy for the classified data. However, it is clear by construction how we have modified the latter, *i.e. by adding a user-defined constraint on the specificity*. Likewise, even Eq. (11) should be viewed as a subjective modification of the unconstrained, prevalence-weighted classification problem.

Ultimately the choice of classification method is best determined by assay developers, and there may be situations in which prevalence weighting is inappropriate. Nonetheless, we feel that the analysis herein highlights the assumptions behind our work and attempts to ground it in objective elements inherent to the population of interest.

6.4. Implications of an indeterminate class

The use of an indeterminate class in diagnostics can have consequences, especially for individual patients. We briefly consider such issues here.

We first note what does not change: it is still possible to estimate prevalence of an entire population, even though some samples are held out. This is a consequence of the methods in Ref. [5,14], which yield unbiased predictions of p that converge in mean-square without ever classifying data. These results are understood physically by recognizing that prevalence estimation entails determining the number of positive samples, not identifying their classes. The latter task is more specific

and amounts to a choice about how to interpret the data, which has no bearing on its underlying statistical properties (such as prevalence).

This observation highlights the subjective nature of classification insofar as Eqs. (6)–(11) are choices of the types of errors we wish to minimize. These choices are informed by the measurement setting and dictate what role an indeterminate class plays in reporting results to individuals. In mass surveillance studies, for example, the primary goal may be to deduce prevalence of various populations. In such cases, a large fraction of inconclusive results could reduce individual confidence in testing, although it should not affect the overall aims of the study. In other settings, such as antibody testing to assess immunity, the specific test results matter more. In some cases an indeterminate class as constructed herein (in terms of local accuracy) could in fact increase confidence and/or usefulness of the diagnostic. For example, in testing an immunocompromised individual, the potential loss associated with an incorrect result suggests the need for a more stringent criterion (e.g. using local accuracy) to determine if an individual has seroconverted.

6.5. Limitations and open directions

A fundamental limitation of our analysis is the assumption that the probabilistic models describing positive and negative samples can be used outside the scope of training data. This problem is common to virtually any classification scheme and is primarily an issue of modeling. Such issues have been explored in a previous manuscript, to which we refer the reader [5]. We note here, however, that model-form errors may introduce uncertainty on the order of a few percent in the conditional probability densities. Thus, it is likely that modeled estimates of accuracy will be incorrect by a proportional amount. This is seen, for example, in the holdout domain computed in Fig. 5. However, Section 6.2 provides means of ensuring that the indeterminate domains are recomputed to satisfy any constraints on empirical estimates of accuracy. We also note that approaches that do not explicitly account for prevalence and/or conditional probabilities are likely to have significantly more model-form errors than estimates based on our approach.

Regarding the indeterminate analysis, Eqs. (13a)–(14) and the generalization considered in Section 6.2 may be a challenging optimization problem to solve, although the solution could be extremely useful for satisfying regulatory and/or public health requirements. Moreover, formalizing the algorithm described in that section and studying its properties relative to the optimal solution may be useful.

A practical limitation of our analysis is the definition of assay performance, provided we allow for variable, prevalence-dependent classification domains. Current standards advocate using sensitivity and specificity estimated for a single validation population having a fixed prevalence. To realize the full potential of our analysis, it is necessary to (i) estimate assay accuracy and uncertainty therein, (ii) characterize the admissible classification domains, and (iii) compute sensitivities and specificities, all as a function of the variable prevalence. While such issues have been partly considered in [5], and deeper investigation of this uncertainty quantification is necessary for widespread adoption of these techniques.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is a contribution of the National Institute of Standards and Technology and is not subject to copyright in the United States. Funding for Johns Hopkins University authors was provided by the Johns Hopkins COVID-19 Research and Response Program, USA, the FIA Foundation, USA, a gift from the GRACE Communications Foundation (C.D.H., N.P.), National Cancer Institute (NCI), USA SeroNet grant U01CA260469 (C.D.H.), National Institute of Allergy and Infectious Diseases (NIAID), USA grant R21AI139784 (C.D.H. and N.P), National Institute of Environmental Health Sciences (NIEHS), USA grant R01ES026973 (C.D.H., N.P.), NIAID, USA grant R01AI130066 and NIH, USA grant U24OD023382 (C.D.H.), NIAID, USA grant 3R01AI148049 (D.L.T.), the Johns Hopkins University School of Medicine COVID-19 Research Fund, USA, the Sherrilyn and Ken Fisher Center for Environmental Infectious Diseases Discovery Program, USA, and NIH, USA grants U54EB007958-12, U5411090366 (Y.C.M.). The aforementioned funders had no role in study design, data analysis, decision to publish, or preparation of the manuscript. P.B. was also funded through the NIST PREP, USA grant 70NANB18H162. The authors wish to thank Daniel Anderson and Arun Moorthy for useful feedback during preparation of this manuscript.

Research involving human participants and/or animals

Use of data provided in this manuscript has been approved by: (1) the NIST Research Protections Office; and (2) the Johns Hopkins School of Medicine Internal Review Board.

Appendix A. Proof of main result

Lemma 1. Assume that $P(\mathbf{r})$ and $N(\mathbf{r})$ are summable functions on Ω and that the measure of any point \mathbf{r} is zero with respect to all distributions. Also assume that $\mathcal{L}[D_p^*, D_N^*] < X$ and that there exists a set of non-zero measure for which $Z(\mathbf{r}) > X$. Then the sets defined by Eqs. (13a)–(14) minimize Eq. (10) subject to Eq. (11).

We first show that Eq. (14) defines $Z_0(X)$ and $\mathcal{C}(X)$. Let $1/2 \leq \zeta \leq 1$ and define

$$\mathcal{D}_h(\zeta) = \{\mathbf{r} : Z^*(\mathbf{r}) < \zeta\} \tag{31a}$$

$$\mathcal{D}_P(\zeta) = D_p^*/\mathcal{D}_h(\zeta) \tag{31b}$$

$$\mathcal{D}_N(\zeta) = D_N^*/\mathcal{D}_h(\zeta). \tag{31c}$$

Eq. (14) motivates the function

$$I(\zeta) = \left[\int_{\mathcal{D}_P(\zeta) \cup \mathcal{D}_N(\zeta)} d\mathbf{r} Q(\mathbf{r}) \right]^{-1} \left[\int_{\mathcal{D}_P(\zeta) \cup \mathcal{D}_N(\zeta)} d\mathbf{r} [Z^*(\mathbf{r}) - X] Q(\mathbf{r}) \right], \tag{32}$$

which is a monotone increasing function of ζ satisfying the inequalities $I(1/2) < 0$ and $I(\zeta) > 0$ for some $\zeta > 1/2$. Thus, there exists a unique value of $Z_0(X)$ for which one of two situations holds: either (I) the function $I(\zeta)$ is continuous at $Z_0(X)$ and $I(Z_0(X)) = 0$, which directly implies Eq. (14); or (II) $I(\zeta)$ suffers a discontinuity, so that $I(Z_0(X)) < 0$ and $I(Z_0(X) + \epsilon) > 0$ for any positive ϵ . The latter case occurs when $\mathcal{S} = \{\mathbf{r} : Z^*(\mathbf{r}) = Z_0(X)\}$ has non-zero measure, and we may set \mathcal{C} to be any subset $\mathcal{C} \subset \mathcal{S}$ provided Eq. (14) is satisfied. The existence of such a \mathcal{C} is guaranteed by the linearity of integration, which implies that

$$\hat{I}(\mathcal{C}) = \left[\int_{\mathcal{D}_P(Z_0) \cup \mathcal{D}_N(Z_0) \cup \mathcal{C}} d\mathbf{r} Q(\mathbf{r}) \right]^{-1} \left[\int_{\mathcal{D}_P(Z_0) \cup \mathcal{D}_N(Z_0) \cup \mathcal{C}} d\mathbf{r} [Z^*(\mathbf{r}) - X] Q(\mathbf{r}) \right]$$

is a continuous, monotone increasing function of the measure of $\mathcal{C} \subset \mathcal{S}$ that passes through zero. Any zero of $\hat{I}(\mathcal{C})$ implies Eq. (14) and defines an appropriate \mathcal{C} .

The proof that Eqs. (13a)–(14) minimize Eq. (10) relies on the observation that any $Z^*(\mathbf{r}) < Z_0(X)$ is farther from the mean value

X than any $Z^*(\mathbf{r}) > Z_0(X)$. Thus, it ‘‘costs additional probability’’ to swap points between the indeterminate region and $\mathcal{D}^* = \mathcal{D}_p^* \cup \mathcal{D}_N^*$ while satisfying the constraint. To see this mathematically, let \mathcal{D} be any other union of positive and negative classification domains satisfying Eq. (11). We do not consider any domains \mathcal{D} that consist only of choosing a different subset $\mathcal{C} \subset \mathcal{S}$ while maintaining Eq. (14). By Eq. (11) one find

$$\int_{\mathcal{D}/\mathcal{D}^*} d\mathbf{r} Q(\mathbf{r})[X - Z^*(\mathbf{r})] - \int_{\mathcal{D}^*/\mathcal{D}} d\mathbf{r} Q(\mathbf{r})[X - Z^*(\mathbf{r})] = 0 \quad (33)$$

We can further expand the second term as

$$\begin{aligned} & \int_{\mathcal{D}^*/\mathcal{D}} d\mathbf{r} Q(\mathbf{r})[X - Z^*(\mathbf{r})] \\ &= \int_{\mathcal{X}^+/\mathcal{D}} d\mathbf{r} Q(\mathbf{r})[X - Z^*(\mathbf{r})] + \int_{\mathcal{X}^-/\mathcal{D}} d\mathbf{r} Q(\mathbf{r})[X - Z^*(\mathbf{r})] \end{aligned} \quad (34)$$

where $\mathcal{X}^+ = \{\mathbf{r} : Z^*(\mathbf{r}) > X\}$ and $\mathcal{X}^- = \{\mathbf{r} : Z_0(X) < Z^*(\mathbf{r}) < X\}$. Clearly the first term on the RHS of Eq. (34) is negative, whereas the second term is positive. Noting that $Z^*(\mathbf{r} \in \mathcal{D}^*) > Z^*(\mathbf{r} \in \mathcal{D}_h^*)$, one finds by inserting Eq. (34) into Eq. (33) that the latter can be expressed in the form

$$\int_{\mathcal{D}/\mathcal{D}^*} d\mathbf{r} Q(\mathbf{r})A(\mathbf{r}) = \int_{\mathcal{X}^+/\mathcal{D}} d\mathbf{r} Q(\mathbf{r})B(\mathbf{r}) + \int_{\mathcal{X}^-/\mathcal{D}} d\mathbf{r} Q(\mathbf{r})C(\mathbf{r}) \quad (35)$$

where $A(\mathbf{r}) > 0$, $B(\mathbf{r}) < 0$, and $0 < C(\mathbf{r}) < A(\mathbf{r})$. This implies that

$$\int_{\mathcal{D}/\mathcal{D}^*} d\mathbf{r} Q(\mathbf{r}) < \int_{\mathcal{D}^*/\mathcal{D}} d\mathbf{r} Q(\mathbf{r}) \implies \int_{\mathcal{D}_h/\mathcal{D}_h^*} d\mathbf{r} Q(\mathbf{r}) > \int_{\mathcal{D}_h^*/\mathcal{D}_h} d\mathbf{r} Q(\mathbf{r}). \quad (36)$$

Consider now the difference of objective functions

$$\begin{aligned} \Delta\mathcal{H} &= \mathcal{H}[\mathcal{D}_h] - \mathcal{H}[\mathcal{D}_h^*] = \int_{\mathcal{D}_h} d\mathbf{r} Q(\mathbf{r}) - \int_{\mathcal{D}_h^*} d\mathbf{r} Q(\mathbf{r}) \\ &= \int_{\mathcal{D}_h/\mathcal{D}_h^*} d\mathbf{r} Q(\mathbf{r}) - \int_{\mathcal{D}_h^*/\mathcal{D}_h} d\mathbf{r} Q(\mathbf{r}). \end{aligned} \quad (37)$$

By inequality (36), we see that $\Delta\mathcal{H} > 0$. Moreover, note that $Z(\mathbf{r}, D_p, D_N) \leq Z^*(\mathbf{r})$ for any classification domains associated with the binary problem. Clearly any choice besides D_N^* and D_p^* entails increasing the measure of \mathcal{D}_h to ensure that the constraint is satisfied. \square

Remark. Lemma 1 is distinct from Theorem 1.14 of Ref. [15] in several subtle ways. The latter minimized a functional $\mathcal{L}[g(z)f(z)]$ of a product of two functions, where $f(z)$ is arbitrary and $g(z)$ satisfies the inequality $0 \leq g(z) \leq 1$ for $z \in \Omega$ (z plays the role of our \mathbf{r} and Ω has the same meaning as in our work). This objective is supplemented with the constraint that the expectation value of $g(z)$ be a constant. The corresponding bathtub principle identifies an optimal $g(z)$ as either an indicator function or a sum of two indicator functions. In the former case, $g(z)$ defines an analog of our optimal domain. However, the structure of the constraint in Theorem 1.14 fixes the measure of the domain and only allows its shape to vary. In contrast, we minimize the measure of the holdout domain subject to an auxiliary constraint, which is a *conditional* expectation value. Doing so allows both the shape and measure of the holdout domain to vary. We refer the reader to Theorem 1.14 of Ref. [15] for more in-depth comparison. The reader may also verify that $g(z)$ being a sum of two indicator functions does not alter the interpretation above.

Appendix B. On PDFs with Dirac masses

Fig. 1 illustrates that biological phenomena may generate a signal so strong that the instrument saturates, i.e. it reaches a limit x_{\max} above which it cannot distinguish different measurement values. This saturation effectively rounds the ‘‘true’’ measurement down to the x_{\max} . The only conclusion we can draw about a reported value x_{\max} is that the true value χ satisfies the inequality $\chi \geq x_{\max}$. Similarly there exists a lower limit x_{\min} up to which smaller measurements values are

rounded. The goal of this section is to incorporate such information into probability modeling.

For concreteness, we restrict ourselves to the one dimensional measurements x associated with the total IgG assay. We assume that were the optical photodetector not restricted to the range $[x_{\min}, x_{\max}]$, the recorded measurement would have been χ returned on the domain $-\infty < \chi < \infty$. Because the measurements have been transformed to a logarithmic coordinate system, $\chi \rightarrow -\infty$ is meaningful. Without additional information about probability of total IgG antibody levels, we make a minimal assumption that χ is described by a Gaussian distribution with an unknown mean μ and variance σ^2 . Thus, on the *open* domain (x_{\min}, x_{\max}) , assume that $x = \chi$, so that the probability of measuring x is

$$\hat{P}_0(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x_{\min} < x < x_{\max}. \quad (38)$$

However, on the boundaries x_{\min} and x_{\max} , we only know that the true values are below and above the respective thresholds. Thus, the probabilities of measuring x_{\min} and x_{\max} are given by

$$\hat{P}_l = \int_{-\infty}^{x_{\min}} \hat{P}_0(\chi|\mu, \sigma^2) d\chi \quad (39)$$

$$\hat{P}_r = \int_{x_{\max}}^{\infty} \hat{P}_0(\chi|\mu, \sigma^2) d\chi \quad (40)$$

where $\hat{P}_0(\chi|\mu, \sigma^2)$ is the same as Eq. (38), but with x replaced by χ . We may then write the full probability model for x as

$$\begin{aligned} \hat{P}(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathbb{I}(x, x_{\min}, x_{\max}) \\ &\quad + \delta(x - x_{\min})\hat{P}_l + \delta(x - x_{\max})\hat{P}_r. \end{aligned} \quad (41)$$

where $\mathbb{I}(x, a, b)$ is the indicator function that x is in the open set (a, b) .

Eq. (41) motivates a generalization of the MLE. Define the likelihood function

$$L(x) = \begin{cases} \hat{P}_0(x) & x_{\min} < x < x_{\max} \\ \hat{P}_l & x = x_{\min} \\ \hat{P}_r & x = x_{\max}. \end{cases} \quad (42)$$

To determine the values of σ and μ , we maximize with respect to these parameters the product of N likelihoods given by

$$\mathcal{L}_{\text{like}}(\mathbf{x}) = \prod_{j=1}^N L(x_j). \quad (43)$$

or alternatively, we minimize the negative log of $\mathcal{L}_{\text{like}}(\mathbf{x})$. To construct the two-dimensional PDF associated with Eq. (15), we assume a corresponding probability model for the SARS-CoV-2 IgG measurements and use standard MLE to identify the distribution parameters. The full PDF for training data is then given by the product of the corresponding PDFs for total IgG and SARS-CoV-2 measurements and has the form given by Eq. (15).

Note that Eq. (15) does not require modification of the proof in the previous section, since any point (x, y) is a set of measure zero, provided that $\mathcal{P}_l(y)$ and $\mathcal{P}_r(y)$ (and their negative counterparts) are bounded functions of y . However, we do require care in defining the local accuracy and classification domains D_N^* and D_p^* . Let

$$A_+ = \{\mathbf{r} : p\mathcal{P}_0(\mathbf{r}) > (1-p)\mathcal{N}_0(\mathbf{r}), x_{\min} < x < x_{\max}\} \quad (44a)$$

$$B_+ = \{\mathbf{r} : p\mathcal{P}_l(y) > (1-p)\mathcal{N}_l(y), x = x_{\min}\} \quad (44b)$$

$$C_+ = \{\mathbf{r} : p\mathcal{P}_r(y) > (1-p)\mathcal{N}_r(y), x = x_{\max}\} \quad (44c)$$

from which we construct $D_p^* = A_+ \cup B_+ \cup C_+$ and the analogous definition for D_N^* . Note that any point for which the prevalence-weighted probabilities of being negative and positive are identical can

be assigned to either class. The corresponding definition of $Z^*(\mathbf{r})$ is given by

$$Z^*(\mathbf{r}) = \begin{cases} \frac{\max[p\mathcal{P}_0(\mathbf{r}), (1-p)\mathcal{N}_0(\mathbf{r})]}{p\mathcal{P}_0(\mathbf{r}) + (1-p)\mathcal{N}_0(\mathbf{r})} & x_{\min} < x < x_{\max} \\ \frac{\max[p\mathcal{P}_r(y), (1-p)\mathcal{N}_r(y)]}{p\mathcal{P}_r(y) + (1-p)\mathcal{N}_r(y)} & x = x_{\min} \\ \frac{\max[p\mathcal{P}_r(y), (1-p)\mathcal{N}_r(y)]}{p\mathcal{P}_r(y) + (1-p)\mathcal{N}_r(y)} & x = x_{\max} \end{cases} \quad (45)$$

where $\mathcal{N}_l(y)$ and $\mathcal{N}_r(y)$ are the analogous of $\mathcal{P}_l(y)$ and $\mathcal{P}_r(y)$ for the negative PDF.

Appendix C. On the point-swap derivatives

To justify the use of Eq. (19), return to Eq. (11) and consider a set \mathcal{D} and its complement \mathcal{D}_h . Consider balls $\mathcal{B} = B(\mathbf{r}, \epsilon)$ and $\mathcal{B}' = B(\mathbf{r}', \epsilon')$ having radii ϵ, ϵ' and centered about \mathbf{r} and \mathbf{r}' . Let these balls be entirely contained in \mathcal{D}_h and \mathcal{D} , respectively. Momentarily assume that the PDFs do not contain Dirac masses. Define \mathcal{D}'_h and \mathcal{D}' to be the sets where \mathcal{B} and \mathcal{B}' have been interchanged without violating Eq. (11). Taking the difference of Eq. (11) defined relative to \mathcal{D} and \mathcal{D}' yields

$$\int_{\mathcal{B}} d\hat{\mathbf{r}} [Z(\hat{\mathbf{r}}) - X] Q(\hat{\mathbf{r}}) - \int_{\mathcal{B}'} d\hat{\mathbf{r}} [Z(\hat{\mathbf{r}}) - X] Q(\hat{\mathbf{r}}) = 0. \quad (46)$$

Assuming that $Z(\hat{\mathbf{r}})$ and $Q(\hat{\mathbf{r}})$ are sufficiently smooth, to leading order in ϵ, ϵ' one finds

$$\epsilon^d [Z(\mathbf{r}) - X] Q(\mathbf{r}) = (\epsilon')^d [Z(\mathbf{r}') - X] Q(\mathbf{r}'). \quad (47)$$

where d is the dimensionality of \mathbf{r} . Rearranging this last equation yields

$$\frac{(\epsilon')^d Q(\mathbf{r}')}{\epsilon^d Q(\mathbf{r})} = \frac{Z(\mathbf{r}) - X}{Z(\mathbf{r}') - X}. \quad (48)$$

Note that ϵ^d and $(\epsilon')^d$ are proportional to the volumes of the respective balls about the points \mathbf{r} and \mathbf{r}' , so that the quantity $(\epsilon')^d Q(\mathbf{r}')$ is, for example, proportional to the infinitesimal probability mass contained in the corresponding ball. Thus, the given by Eq. (48) is the relative change probability mass exchanged between \mathcal{D} and \mathcal{D}_h in swapping \mathbf{r} and \mathbf{r}' .

If we change the class of \mathbf{r} (either from \mathcal{D}_P to \mathcal{D}_N or vice versa), it may be necessary to hold out additional points \mathbf{r}' , or it may be possible to move points from the indeterminate into the classification domain. In either case, letting \mathcal{B} and \mathcal{B}' have the same definitions as before and assuming Eq. (11) holds, one finds

$$\int_{\mathcal{B}} d\hat{\mathbf{r}} [2Z(\hat{\mathbf{r}}) - 1] Q(\hat{\mathbf{r}}) \pm \int_{\mathcal{B}'} d\hat{\mathbf{r}} [Z(\hat{\mathbf{r}}) - X] Q(\hat{\mathbf{r}}) = 0, \quad (49)$$

where \mathcal{B}' is the ball moved to (+) or from (-) the indeterminate domain, depending on the sign of the first term; note that we also require $Z < X$ inside \mathcal{B}' . Again taking the limit that the respective ϵ are small, one finds

$$\epsilon^d [2Z(\mathbf{r}) - 1] Q(\mathbf{r}) = \pm (\epsilon')^d [X - Z(\mathbf{r}')] Q(\mathbf{r}') \quad (50)$$

Dividing by $\pm[X - Z(\mathbf{r}')] Q(\mathbf{r}')$ yields the infinitesimal probability mass moved to or from the indeterminate

$$(\epsilon')^d Q(\mathbf{r}') = \pm \frac{\epsilon^d [2Z(\mathbf{r}) - 1] Q(\mathbf{r})}{X - Z(\mathbf{r}')}. \quad (51)$$

The LHS must be positive, and the denominator on the RHS is positive. Thus, the + and - signs on the RHS occur when $Z(\mathbf{r}) > 1/2$ and $Z(\mathbf{r}) < 1/2$, corresponding to the situations in which probability moves to and from the indeterminate region. Thus, in assessing when \mathcal{D}_h grows, it is sufficient to test the inequality $Z(\mathbf{r}) > 1/2$.

The analysis of this section is easily generalized to the case of Eq. (15) by noting that for points on the lines $x = 0$ and $x = 1$, the balls of radius ϵ should be taken as intervals on the line with length 2ϵ . This yields the appropriate generalization of probability associated with those points.

References

- [1] A.M. Lerner, R.W. Eisinger, D.R. Lowy, L.R. Petersen, R. Humes, M. Hepburn, M.C. Cassetti, The COVID-19 serology studies workshop: Recommendations and challenges, *Immunity* (ISSN: 1074-7613) 53 (1) (2020) 1–5.
- [2] FDA, EUA authorized serology test performance, 2020, <https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance> (Accessed: 2020-09-16).
- [3] K. Bond, S. Nicholson, S.M. Lim, T. Karapanagiotidis, E. Williams, D. Johnson, T. Hoang, C. Sia, D. Purcell, F. Mordant, S.R. Lewin, M. Catton, K. Subbarao, B.P. Howden, D.A. Williamson, Evaluation of serological tests for SARS-CoV-2: Implications for serology testing in a low-prevalence setting, *J. Infect. Dis.* 222 (8) (2020) 1280–1288, <http://dx.doi.org/10.1093/infdis/jiaa467>.
- [4] W.H. Bermingham, T. Wilding, S. Beck, A. Huissoon, SARS-CoV-2 serology: Test, test, test, but interpret with caution!, *Clin. Med.* 20 (4) (2020) 365–368, <http://dx.doi.org/10.7861/clinmed.2020-0170>.
- [5] P.N. Patrone, A.J. Kearsley, Classification under uncertainty: data analysis for diagnostic antibody testing, *Math. Med. Biol.: J. IMA* (2021) dqab007, <http://dx.doi.org/10.1093/imammb/dqab007>.
- [6] L. Böttcher, M.R. D’Orsogna, T. Chou, A statistical model of COVID-19 testing in populations: effects of sampling bias and testing errors, *Phil. Trans. R. Soc. A* 380 (2214) (2022) 20210121, <http://dx.doi.org/10.1098/rsta.2021.0121>.
- [7] B. Meyer, G. Torriani, S. Yerly, L. Mazza, A. Calame, I. Arm-Venez, G. Zimmer, T. Agoritsas, J. Stirnemann, H. Spechbach, I. Gueussou, S. Stringhini, J. Pugin, P. Roux-Lombard, L. Fontao, C.-A. Siegrist, I. Eckerle, N. Vuilleumier, L. Kaiser, for Emerging Viral Diseases, Geneva Center, Validation of a commercially available SARS-CoV-2 serological immunoassay, 2020/06/27, *Clin. Microbiol. Infect. : Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* 26 (10) (2020) 1386–1394, <http://dx.doi.org/10.1016/j.cmi.2020.06.024>, 32603801 [pmid].
- [8] N. Lee, S. Jeong, M.-J. Park, W. Song, Comparison of three serological chemiluminescence immunoassays for SARS-CoV-2, and clinical significance of antibody index with disease severity, *PLoS One* 16 (6) (2021) 1–13, <http://dx.doi.org/10.1371/journal.pone.0253889>.
- [9] D.M. Manthei, J.F. Whalen, L.F. Schroeder, A.M. Sinay, S.-H. Li, R. Valdez, D.A. Giachero, C. Gherasim, Differences in performance characteristics among four high-throughput assays for the detection of antibodies against SARS-CoV-2 using a common set of patient samples, *Am. J. Clin. Path.* 155 (2) (2020) 267–279, <http://dx.doi.org/10.1093/ajcp/aqaa200>.
- [10] E.S. Theel, J. Harring, H. Hilgart, D. Granger, Performance characteristics of four high-throughput immunoassays for detection of IgG antibodies against SARS-CoV-2, *J. Clin. Microbiol.* 58 (8) (2020) e01243–20, <http://dx.doi.org/10.1128/JCM.01243-20>, 32513859 [pmid].
- [11] N. Pisanic, P.R. Randad, K. Kruczynski, Y.C. Manabe, D.L. Thomas, A. Pekosz, S.L. Klein, M.J. Betenbaugh, W.A. Clarke, O. Laeyendecker, P.P. Caturegli, H.B. Larman, B. Detrick, J.K. Fairley, A.C. Sherman, N. Roupheal, S. Edupuganti, D.A. Granger, S.W. Granger, M.H. Collins, C.D. Heaney, M.J. Loeffelholz, COVID-19 serology at population scale: SARS-CoV-2-specific antibody responses in saliva, *J. Clin. Microbiol.* 59 (1) (2020) e02204–20, <http://dx.doi.org/10.1128/JCM.02204-20>.
- [12] P.R. Randad, N. Pisanic, K. Kruczynski, T. Howard, M.G. Rivera, K. Spicer, A.A. Antar, T. Penson, D.L. Thomas, A. Pekosz, N. Ndahiro, L. Aliyu, M.J. Betenbaugh, H. Manley, B. Detrick, M. Katz, S. Cosgrove, C. Rock, I. Zyskind, J.I. Silverberg, A.Z. Rosenberg, P. Duggal, Y.C. Manabe, M.H. Collins, C.D. Heaney, Durability of SARS-CoV-2-specific IgG responses in saliva for up to 8 months after infection, 2021.
- [13] C.D. Heaney, N. Pisanic, P.R. Randad, K. Kruczynski, T. Howard, X. Zhu, K. Littlefield, E.U. Patel, R. Shrestha, O. Laeyendecker, S. Shoham, D. Sullivan, K. Gebo, D. Hanley, A.D. Redd, T.C. Quinn, A. Casadevall, J.M. Zenilman, A. Pekosz, E.M. Bloch, A.A. Tobian, Comparative performance of multiplex salivary and commercially available serologic assays to detect SARS-CoV-2 IgG and neutralization titers, *J. Clin. Virol.* 145 (2021) 104997.
- [14] P. Patrone, A. Kearsley, Minimizing uncertainty in prevalence estimates, 2022, <http://dx.doi.org/10.48550/ARXIV.2203.12792>.
- [15] E. Lieb, M. Loss, M. LOSS, American Mathematical Society, Analysis, Crm Proceedings & Lecture Notes, American Mathematical Society, 2001.
- [16] T. Tao, An Introduction to Measure Theory, in: Graduate studies in mathematics, American Mathematical Society, 2011.
- [17] F. Jones, Lebesgue Integration on Euclidean Space, in: Jones and Bartlett books in mathematics, Jones and Bartlett, 2001.
- [18] H. Pishro-Nik, Introduction to Probability, Statistics, and Random Processes, Kappa Research, LLC, 2014.
- [19] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, in: Adaptive computation and machine learning series, University Press Group Limited, 2006.

- [20] F. Klebaner, *Introduction to Stochastic Calculus with Applications*, in: *Introduction to Stochastic Calculus with Applications*, Imperial College Press, 2005.
- [21] A. Clifford Cohen, *Truncated and Censored Samples*, CRC Press, 2016.
- [22] N.H.K. Tony, *Censoring methodology*, in: M. Lovric (Ed.), *International Encyclopedia of Statistical Science*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 221–224.
- [23] C.M. Florkowski, Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests, *Clin. Biochem. Rev.* 29 Suppl 1 (Suppl 1) (2008) S83–S87.
- [24] A. Algaissi, M.A. Alfaleh, S. Hala, T.S. Abujamel, S.S. Alamri, S.A. Almahboub, K.A. Alluhaybi, H.I. Hobani, R.M. Alsulaiman, R.H. AlHarbi, M.-Z. ElAssouli, R.Y. Alhabbab, A.A. AlSaieedi, W.H. Abdulaal, A.A. Al-Somali, F.S. Alofi, A.A. Khogeer, A.A. Alkayyal, A.B. Mahmoud, N.A.M. Almontashiri, A. Pain, A.M. Hashem, SARS-CoV-2 S1 and N-based serological assays reveal rapid seroconversion and induction of specific antibody response in COVID-19 patients, *Sci. Rep.* 10 (1) (2020) 16561, <http://dx.doi.org/10.1038/s41598-020-73491-5>.
- [25] L. Grzelak, S. Temmam, C. Planchais, C. Demeret, L. Tondeur, C. Huon, F. Guivel-Benhassine, I. Staropoli, M. Chazal, J. Dufloo, D. Planas, J. Buchrieser, M.M. Rajah, R. Robinot, F. Porrot, M. Albert, K.-Y. Chen, B. Crescenzo-Chaigne, F. Donati, F. Anna, P. Souque, M. Gransagne, J. Bellalou, M. Nowakowski, M. Backovic, L. Bouadma, L. Le Fevre, Q. Le Hingrat, D. Descamps, A. Pourbaix, C. Laouénan, J. Ghosn, Y. Yazdanpanah, C. Besombes, N. Jolly, S. Pellerin-Fernandes, O. Cheny, M.-N. Ungeheuer, G. Mellon, P. Morel, S. Rolland, F.A. Rey, S. Behillil, V. Enouf, A. Lemaitre, M.-A. Créach, S. Petres, N. Escriou, P. Charneau, A. Fontanet, B. Hoen, T. Bruel, M. Eloit, H. Mouquet, O. Schwartz, S. van der Werf, A comparison of four serological assays for detecting anti-SARS-CoV-2 antibodies in human serum samples from different populations, 12 (559) 2020. <http://dx.doi.org/10.1126/scitranslmed.abc3103>.
- [26] A. Hachim, N. Kavian, C.A. Cohen, A.W.H. Chin, D.K.W. Chu, C.K.P. Mok, O.T.Y. Tsang, Y.C. Yeung, R.A.P.M. Perera, L.L.M. Poon, J.S.M. Peiris, S.A. Valkenburg, ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection, *Nat. Immunol.* 21 (10) (2020) 1293–1301, <http://dx.doi.org/10.1038/s41590-020-0773-7>.
- [27] J. Nocedal, S. Wright, *Numerical Optimization*, in: *Springer Series in Operations Research and Financial Engineering*, Springer New York, 2006.