

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Sarah M. Sullivan

Date

Mixture Modeling to Determine Population-Specific Cutoffs for Quantitative Diagnostic Tests without Gold Standards

By

Sarah M. Sullivan
Master of Science in Public Health

Biostatistics

Howard Chang
Thesis Advisor

Lance Waller
Reader

Mixture Modeling to Determine Population-Specific Cutoffs for Quantitative Diagnostic Tests without Gold Standards

By

Sarah M. Sullivan
B.S, B.A., University of Pittsburgh, 2013

Thesis Committee Chair: Howard H. Chang, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2017

Abstract

Mixture Modeling to Determine Population-Specific Cutoffs for Quantitative Diagnostic Tests without Gold Standards

By Sarah M. Sullivan

As Neglected Tropical Disease (NTD) programs succeed and transmission intensity declines, the ability to discriminate between positive and negative antibody tests becomes increasingly challenging. Previous techniques for defining diagnostic test cutoffs are no longer sufficient as prevalence rates decline towards zero. With varying degrees of non-specific background reactivity across populations and the absence of a gold standard, an objective yet flexible approach for cutoff determination is needed.

Mixture modeling allows for the probabilistic representation of subpopulations within an overall population. By fitting a mixture model to continuous data, members of the overall population can be assigned to groups (e.g., positive and negative), and the certainty of classification can be calculated based on the associated conditional probabilities. Certainty of classification can also be used to calculate absolute cutoffs and pre-specified indeterminate ranges (e.g. greater than 80% certainty of classification), resulting in positive, negative and indeterminate groups. The number of subpopulations and types of distribution (including Gaussian and skew-normal) may be specified or optimized by an algorithm using model selection criteria, such as the Bayesian Information Criterion.

However, current implementations of mixture modeling can be difficult to use for the public health and lab-based practitioners who determine diagnostic cutoffs. Therefore, we created an analytic tool which performs mixture modeling using the normal and skew-normal distributions for a variable number of subpopulations. The tool then calculates cutoffs, picks optimal models, and calculates indeterminate ranges with minimal user input.

We utilized this analytic tool to perform mixture modeling on standardized ELISA results from two post-treatment NTD settings. Antibody responses to a lymphatic filariasis recombinant antigen (Wb123) were analyzed via the cutoff tool and a two component, skew-normal model was found to be optimal. This yielded a cutoff of 0.115 and an indeterminate range of (0.100, 0.128), corresponding to 89.18% negative, 7.16% indeterminate and 3.66% positive results. The cutoff tool was also used to analyze responses to a recombinant onchocerciasis antigen (Ov16), resulting in an optimal skew-normal, three component model, which yielded a cutoff of 0.609 (0.567, 0.646) with 96.31% negative, 1.5% indeterminate and 2.2% positive results.

These results demonstrate the utility of mixture modeling as a tool to provide population-specific diagnostic cutoffs with a corresponding indeterminate group that reflects our certainty regarding the cutoff. Such an approach may benefit other neglected and infectious disease programs driving towards elimination.

Mixture Modeling to Determine Population-Specific Cutoffs for Quantitative Diagnostic Tests without Gold Standards

By

Sarah M. Sullivan
B.S, B.A., University of Pittsburgh, 2013

Thesis Committee Chair: Howard H. Chang, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2017

Acknowledgments

First, I would like to thank Dr. Howard Chang for connecting me to this project, as well as for his teaching and support throughout my masters program. Thank you for challenging me in mixture modeling, spatial analysis, and advanced regression techniques and supporting me in doing some of the work that I am most proud of.

I am also grateful to Dr. Katie Gass at the Task Force for Global Health who initiated this project and brought me on to carry it forward to its current fruition. The time that I spent working at the Task Force was some of the most stimulating of my career and I relished the opportunity to do such impactful work.

Thanks to my fellow Biostatistics students for your camaraderie and frequent company, and to the SPENSER group for igniting my curiosity. Thanks to the department for fostering an environment that was both challenging and supportive. In particular thank you to Dr. John Hanfelt for emphasizing that there is space for many people in biostatistics, for extending the lessons we learn from public health to how we treat people, and for prioritizing lasting learning.

Finally, I would like to thank my family and friends for their love and support during the two years that I worked on this project, as well as during the rest of my life. To parents for encouraging my curiosity and love of knowledge: my father for teaching me early that math is the language of science and shouldn't be discounted, and my mother for making books so central to my life. Thanks especially to Rachel, Jenn, Ameenay and Danielle who were willing to listen as I talked through roadblocks on this work, despite being skilled in other fields, and to Quinn who was willing to listen and is skilled in this one.

Contents

Introduction	1
Methods	2
Mixture Modeling	2
Mixture Modeling Implementation	4
Script Development and Capabilities	4
Results	7
Example A	7
Example B	8
Discussion	9
Figures	12
Figure 1	12
Figure 2	12
Figure 3	13
Figure 4	14
Figure 5	14
Figure 6	15
References	16

Introduction

Serology has long been a useful tool for detecting and understanding the dynamics of neglected tropical diseases (NTDs). For lymphatic filariasis and onchocerciasis, two NTDs driving towards elimination, serologic markers may prove to be pivotal indicators for demonstrating when elimination has been achieved. While newly developed rapid diagnostic antibody tests show a lot of promise[1], the majority of antibody detection still utilizes laboratory-based assays including ELISA, LIPS and Luminex.[4] These assays return a quantitative value that is related to the level of antibody response, optical density (OD) in the case of enzyme-linked immunosorbent assays (ELISA). However, setting an appropriate cutoff to distinguish positive and negative results can present a real challenge.

Ideally, one would utilize a gold standard and set the cutoff to optimize sensitivity and specificity via Receiver Operating Characteristic (ROC) curves. However, an ROC analysis requires the presence of known positive and negative samples and for many NTDs we lack a gold standard diagnostic tool. Therefore, ROC curves are of limited utility for setting appropriate cutoffs for NTDs.

In the past, cutoffs were sometimes determined by setting a threshold at 3 standard deviations above the mean. While effective in a setting with many true positives, this method begins to break down as we head towards elimination and see fewer true positives and more skewed data. In the case of normally distributed data, setting this cutoff for positivity will, because of the distribution of the normal curve, lead to at least 0.3% positives, and in the case of a heavily skewed distribution will yield even more positives, even if there are no true positive results. Therefore, it is necessary to explore other approaches to setting a cutoff. Sigmoidal standard curves have also been used determine concentrations of antibody present and therefore positivity. However, these sigmoidal curves are most robust in the nearly linear portion of the curve and are much less reliable at the tails since the gradient of change is smaller, making it more difficult to distinguish between non-response and weak response. This presents problems in low intensity settings that we expect to encounter as background prevalence decreases.

New approaches are therefore required in order to identify the optimal cutoffs for antibody tests. Mixture modeling is a statistical tool that allows for the probabilistic representation of subpopulations within an overall population. By fitting a mixture model to observed data, members of the overall population can be assigned to groups (e.g., positive and negative) using the expectation maximization (EM) algorithm. [8]

Previous researchers have utilized mixture modeling to determine data-based cutoffs, using two components of Gaussian (normal) distribution.[9] In this current paper we describe an enhanced approach for determining optimal cut-offs for serologic data that expands upon this previous work in several key ways. Firstly, because the results of diagnostic tests are often positively skewed, we have incorporated the skew-normal distribution into the mixture modeling. Furthermore, given the diversity of treatment histories, exposure duration and co-infections that may exist within a population, it is plausible for there to be more than two serological phenotypes represented in the data (e.g. both a weak positive and a strong positive distribution). To account for this possibility, we have incorporated the ability to perform mixture modeling with more than two components. Additionally, building on the work of prior researchers,[10] we utilize the conditional probabilities from mixture modeling to create indeterminate ranges, which introduce nuance into the deterministic cutoff discussion since we know that the results near the cutoff are the ones most likely to be incorrectly classified.

Based on these concepts, we have created a publicly available tool for laboratory users: an analytic tool (contained in the supplementary R script) which (1) performs mixture modeling using the normal and skew-normal distributions and considering a variable number of components, (2) selects the optimal model based on a model selection criteria, the Bayesian Information Criterion (BIC), and other possible criteria from the user, and (3) provides the user a cutoff, indeterminate ranges, and classifications based on individual-level probabilistic assignments.

Methods

Mixture Modeling

Mixture modeling is a framework that allows us to model data that arise from a finite number of unobserved subpopulations.[8] It is a popular statistical approach that has been applied to a wide variety of fields including medicine, genetics, public health, psychology, and biology.[11] and has been increasingly used to determine the results of serological tests.[9],[12],[13],[14]

Mixture modeling estimates different, often overlapping, components (which represent the subpopulations) within data by assuming the overall distribution is a combination of simple and well-characterized probability distributions. For example, the observed quantitative diagnostic assay results

may contain a subpopulation that should be considered “negative” and another subpopulation that should be considered “positive”, with each subpopulation of responses following its own distribution of reactivity that is characterized by its mean and variance (assuming they follow a normal distribution). Mixture modeling aims to examine the overall observed distribution of a dataset with heterogeneous subpopulations to find the underlying distributions (and associated proportions) of each component distribution that together comprise the overall distribution.

Let X denote the vector of observations; the likelihood function of Θ for a mixture distribution with N observations and K populations is given by:

$$p(X|\Theta) = L(\Theta|X) = \prod_{i=1}^N \sum_{k=1}^K \alpha_k * p_k(x_i|\theta_k)$$

where i indexes the individual observation and k indexes the subpopulation. Parameter α_k represents the proportion the population falling in subpopulation k , and p_k denotes the distribution of observations within subpopulation k defined by parameters θ_k . Note $\sum_{k=1}^K \alpha_k = 1$. The parametric family of each component in a given model is typically treated as identical.

We estimate unknown distribution parameters of these subpopulations via maximum likelihood using the Expectation Maximization (EM) algorithm. The EM algorithm is a two-step iterative procedure which finds the maximum likelihood parameters of the subpopulation by first calculating the expected likelihood given a set of initial parameter estimates, then updating the parameter estimates based on the previous expectation. This procedure is repeated until convergence and the set of component parameters that are most likely to have given rise to the observed data is reached. Utilizing these estimated parameters, it is then possible to classify to which subpopulation any given observation belongs.

In practice, the EM algorithm estimates can converge to a local maximum, rather than a global maximum, so it is important to identify good initial values for model parameters and run the EM algorithm with different initial values, especially when populations are not normal. For a more detailed explanation of the EM algorithm, see McLachlan & Peel 2000.[15]

Mixture Modeling Implementation

Mixture modeling based on the EM algorithm and related algorithms has been implemented in several R packages, and for this application we focused on the *mclust*,[16] and *mixsmsn*[17] packages. The *mclust* package was one of the earlier mixture modeling packages created and provides robust, consistent parameter estimates for mixtures of normal data and contains the capabilities to either specify the number of components present or to choose the optimal number of components via the model selection criterion BIC. *Mclust* only enables mixture modeling with the normal distribution, and utilizes hierarchical (density) based clustering to find initial parameter values. Due to its implementation, *mclust* results in the exact same final models (and therefore BICs) each time the procedure is run on the same data.

However, *mclust* does not allow for non-normal subpopulation data, and since quantitative diagnostic tests are often very heavily skewed even in a single population, we also utilized other packages that incorporated distributions that may better fit the observed distribution of our serology data. *Mixsmsn* allows for mixture modeling of subpopulations as either normal or skew-normal distributions. The skew-normal distribution is an extension of the normal distribution that adds an additional parameter that controls skewness, (the normal distribution represents a special case of the skew-normal distribution where skewness=0). This characteristic of the skew-normal distribution is attractive given the that continuous immunologic assays often exhibit large skewness. *Mixsmsn* sets the initial parameter values via k-means, a robust, though variable procedure. Moreover, this method of selecting the initial values for the EM algorithm can differ slightly each time it is run and can lead to slight variation in the final model estimates and BICs. Though this variability is not desirable, when we used model selection criteria like BIC to compare the normal models from *mclust* and *mixsmsn* (on the same data), we found that the two fit similar models, but on average *mixsmsn* estimates were superior based on BIC, and that if run several times, the best (lowest BIC) estimates from *mixsmsn* almost always outperformed *mclust* in the normal case.

Script Development and Capabilities

Building off of the capacities of mixture modeling, we developed an R script which takes the observed data and finds the optimal distribution (either normal or skew-normal) and number of components to fit these data, considering that the observed data may have been generated from one or more subpopulations.

Models with 1 to 5 components (corresponding to subpopulations in the data) are considered, yielding 10 possible models. Each model—that is the unique combination of distributions and number of components—is iterated 10 times with different initial values from k-means. The iteration with the lowest BIC is chosen for each model. Because higher-order models (i.e. a model with 4 or 5 components) require the estimation of more parameters, they are more likely to fail to converge. Therefore, we have enabled the function to run, provided 8 of the 10 iterations for each model converge, and any iterations that do not converge are excluded from further analysis. The user is also allowed to control the maximum number of components considered; the default maximum number of components is 5.

Once the parameters have been estimated for each iteration of each model (i.e. each combination of distribution and number of components), our script utilizes BIC to determine which iteration is optimal for each model, as iterations can have different results based on different initial starting values. Then, based on the optimal iteration for each model, the script determines which model is optimal overall. Comparison of BICs for the possible models is output by the function and visually represented in graphical form. If the one-group model is optimal, this indicates that there are not multiple subpopulations in the sample (an example of what we would see in a post-elimination sample from a single population, or a population that had never been exposed). If a model with two or more components is optimal by BIC, this indicates that there are multiple subpopulations in the sample and a cutoff to differentiate between positives and negatives is necessary. Though the program by default chooses the optimal model by BIC, it is also possible for a researcher to choose a different model to base a cutoff on. The results of all ten potential models are available to researchers and may be utilized if the number and form of the subpopulation components are known a priori.

Once the (typically BIC-optimal) model has been chosen, the analytic tool calculates the certainty of classification, which is useful for determining subpopulation assignment for observations (the observed quantitative diagnostic results) falling in the overlap between the subpopulation distributions. Certainty of classification is the conditional probability of belonging to a particular group given the value of the observation and the parameters of the components, i.e. the conditional probability $P(A|X = x; \theta)$, where A is the classification. The observation is then classified into the component to which it has the highest conditional probability of belonging. The cutoff is also derived from this conditional probability; it is the point at which the conditional probability of belonging to either component is 50:50 or the place where the certainty of classification is minimized.

In the case of an optimal two-component model, the cutoff is set deterministically, since it will always fall between the two components. When a model with three or more components is selected as optimal, the user must choose which components to set the cutoff between (ie in the case of a 3 component model, setting the cutoff between the first and second component or the second and third), based on their scientific knowledge and context such as historical prevalence in the population and plausible test values. In order to assist in this process, the script provides a summary and graphical tools, summarized below, to inform the user in this decision.

Given the optimal model, a result summary is produced that identifies what proportion of the results would be deterministically defined as positive and negative for different possible cutoffs. On the graphical side, the first of these tools is a plot of classification uncertainty. Classification uncertainty is defined as one minus the certainty of classification into the most likely component plotted against the possible values. For example, in a two-component model, the maximum uncertainty is 0.5, which represents the value at which both components are equally likely; however, the uncertainty can be higher in models with multiple components. The maximum uncertainty also corresponds to the cutoff between components, and can graphically show the indeterminate ranges. The second plot displays the distributions overlaid on a histogram of the data and allows the user to visualize the estimated mixture distribution and model fit. The third plot is a histogram of all of the responses classified into components.

Certainty of classification can also be used to create indeterminate ranges. In addition to positive and negative groups, these ranges define a third classification, a group of samples whose probabilistic certainty of classification is below a certain threshold, for example 80% or 90% certainty, though these intervals should not be mistaken as prediction intervals.

Once the user has chosen between which components the cutoff should lie, another function performs the classification and provides the final boundaries for positives, indeterminates and negatives. By default, the function finds indeterminate groups with 80% and 90% certainty, but a separate function allows the user to set other certainty levels. Once the classifications, boundaries and cutoffs have been established, graphics and summary tables can be generated. Additionally, data with identifiers and classifications can be written out to a flat file.

Results

To illustrate the capabilities of this script, two case studies using the results from ELISA assays are presented: one that results in the straightforward 2 component optimal model and a second that results in a 3 component optimal model which requires more user input.

Example A

The first example utilizes an all ages dataset of Wb123 antibody ELISA results from a formerly lymphatic filariasis (LF) endemic country which has completed several rounds of mass drug administration and is preparing to end treatment. This optical density data has been normalized against a pooled positive serum control to adjust for variability across plates. The data are significantly right skewed as shown in Figure 1. After running the model fitting function, we find that for these data, the model with two components and the skew-normal distribution is optimal (Figure 2). After running the model selection function, we can examine the plots and ascertain that there is one component which is mildly right skewed with a large peak that comprises the majority of the observations, and another component which has few observations, and is very right skewed (Figure 3). In this case since there are only two components, we conclude that observations in the lower component pertain to a negative subpopulation and observations in the upper component pertain to a positive subpopulation.

By running the cutoff function, we can ascertain the cutoff between components, the associated indeterminate ranges, and the classifications for each observed data point. The cutoff for this dataset falls at 0.115 (standardized optical density units), with an 80% indeterminate range of (0.100, 0.128). Utilizing the 80% indeterminate range, the observed data are characterized as having 89.18% negative, 7.16% indeterminate and 3.66% positive results. If one were only to use the raw cutoff, and discard the use of an indeterminate range, the corresponding negative and positive percents would be 94.50% and 5.50%, respectively.

The second dataset utilizes all ages Ov16 ELISA antibody data from an onchocerciasis endemic country which has been under many years of ongoing treatment, with the data normalized against a recombinant human anti-Ov16 IgG4 antibody-based positive control.[] As in example A, the data are significantly right skewed, though in this case the skewness is more extreme, as shown in Figure 4. The

model fitting function finds that the optimal model is a three-component skew-normal mixture (Figure 5). Therefore, we must run the model selection function and examine the associated plots in order to ascertain which two components our cutoff should be between, keeping in mind the scientific context of the problem and any a priori assumptions. The two possible cutoffs for this dataset are 0.61 and 3.18 (standardized optical density units) which would correspond to 2.75% and 0.12% percent positive, respectively.

Example B

The second dataset utilizes all ages Ov16 ELISA antibody data from an onchocerciasis endemic country which has been under many years of ongoing treatment, with the data normalized against a recombinant human anti-Ov16 IgG4 antibody-based positive control.[18] As in example A, the data are significantly right skewed, though in this case the skewness is more extreme, as shown in Figure 4. The model fitting function finds that the optimal model is a three-component skew-normal mixture (Figure 5). Therefore, we must run the model selection function and examine the associated plots in order to ascertain which two components our cutoff should be between, keeping in mind the scientific context of the problem and any a priori assumptions. The two possible cutoffs for this dataset are 0.61 and 3.18 (standardized optical density units) which would correspond to 2.75% and 0.12% percent positive, respectively.

In this case, the knowledge that onchocerciasis has historically been controlled but endemic in the country is important. Since we are aware that Ov16 test is an antibody test which measures historical exposure to L3 larvae—a sign of circulating onchocerciasis in the community.[19] Though the length of Ov16 antibody response is not precisely know, it is a long-lasting antibody, and so in a community where onchocerciasis has previously been circulating, one would expect so antibody positivity among this all ages sample to reflect the historical circulation of the parasite. Therefore, we chose the lower cutoff since it was empirically valid and supported by our a priori knowledge.

After running the cutoff function, setting the cutoff between the first and second components, we can examine the plots and ascertain the negative component is mildly right skewed with a large peak that comprises the majority of the observations, and the positive component which has few observations, and is very right skewed (Figure 6). The cutoff for this dataset is 0.609, with an 80% indeterminate range of (0.567, 0.646). Utilizing the 80% indeterminate range, the sample is comprised of 96.31% negative, 1.5% indeterminate and 2.2% positive results. If one were only to use the raw cutoff, the percent negative and positive would be 97.25% and 2.75% respectively. The plot of the positive and negative distributions

with indeterminate ranges echoes this conclusion, with a narrow indeterminate range fairly high in the data (Figure 6).

Discussion

Serological assays are common tools for disease monitoring and surveillance, but without a gold standard, identifying a cutoff is challenging. In the case of NTDs, serological tools are essential for monitoring population prevalence as we approach elimination, but without reliable cutoffs their utility for decision-making is hampered. Therefore, we present a data-based mixture model approach that builds on the work of others by incorporating the ability to detect multiple phenotypes, utilize the skew-normal distribution to better fit the data, and acknowledging the uncertainty inherent in setting a cutoff.

One novel aspect of this analytic cutoff tool, and an important contribution to the interpretation of serologic results, is the inclusion of an indeterminate range. Though this may not appear to be an advantage to users who are accustomed to receiving a definitive positive or negative result, it reflects the reality of our understanding, and provides a quantification of the certainty of classification. Given that the distributions of positives and negatives overlap, the indeterminate range recognizes and quantifies the degree to which any given value may be misclassified. Though misclassification is not a desired characteristic, it is unavoidable in all diagnostic tests and indeterminate ranges provide a more honest interpretation of the data, reflecting the inherent uncertainty of classification around cutoff designation.

The incorporation of an indeterminate range also helps to more accurately reflect the limitations of the antibody assays. Though highly specific, neither the Wb123 nor the Ov16 assays attains 100% specificity. Additionally, differing overall worm burden can lead to immune up- and down-regulation and consequently differing OD values over an entire population, regardless of the LF or onchocerciasis prevalence. Given these diverse drivers, the inclusion of an indeterminate range is particularly important, because even with perfect knowledge of the optical density results, some misclassification would likely still occur.

Though we have presented an empirically based method for finding cutoffs for quantitative diagnostic tests, the method does have its limitations. When the optimal model fit involves more than two components, a subjective decision is required of the end-user to determine between which two components to set the cutoff. This can be avoided by restricting the analytic tool to one or two component solutions,

yielding results that are easier to interpret; however, this approach risks over simplifying reality and may result in poorer model fit. Regardless of the optimal number of components, the cutoffs derived are data-based, which has the advantage of naturally adjusting for differences in epidemiology and laboratory procedures. It is important to keep in mind that the analytic tool does not provide a universal cutoff for an assay and must be repeated for any new population that is being tested, although utilizing the derived cutoff for the same endemic area is appropriate. Pooling data across different sites could provide a path toward more generalizable cutoffs, however potential heterogeneity of cutoffs across and within populations requires careful consideration.

An additional limitation is that the mixture model relies on specifying a single family of distributions for all subpopulations. Hence the analytic tool can be sensitive to outliers; in the presence of extreme outliers the analytic tool may fail to run. However, because the classification of extreme outliers is generally obvious (e.g., an OD value >4 times a serum control clearly represents a positive) they can be removed from the data prior to running the cutoff analytic tool and the resulting classifications will remain operational. Additionally, if the underlying distribution of the data cannot be represented by a normal or skew-normal distribution, the analytic tool will not produce valid classifications. In some cases, this will cause the analytic tool to fail to converge or attain very poor fit, which can be visually checked via the distributional plots; therefore, it is important to check that the descriptive plots that include the distribution appear to fit the histogram of the data, especially at the tails. Another sign of poor fit, which is particularly likely to be seen models with a high number of components (4+), is a fit which assigns equal proportions of the population to each component where the distribution of the data does not indicate this shape.

Although the examples discussed in this paper were ELISA WB123 and Ov16 tests, the cutoff analytic tool presented here has much broader applicability that reaches beyond NTDs. As its basis is data-driven, rather than grounded in a particular disease pathology, any diagnostic test platform that produces a continuous result (e.g. LIPS, Luminex) could utilize this analytic tool to determine cutoffs. Future directions for research include exploring additional subpopulation distributions and head-to-head comparisons of the described analytic tool with existing cutoff determination techniques. Though Gaussian mixture modeling has been frequently used to determine serological cutoffs and the skew-normal distribution provides a more flexible extension, underlying distributions in data may be neither normal nor skew-normal, or may be a mixture of the two. By incorporating additional distributions, and allowing

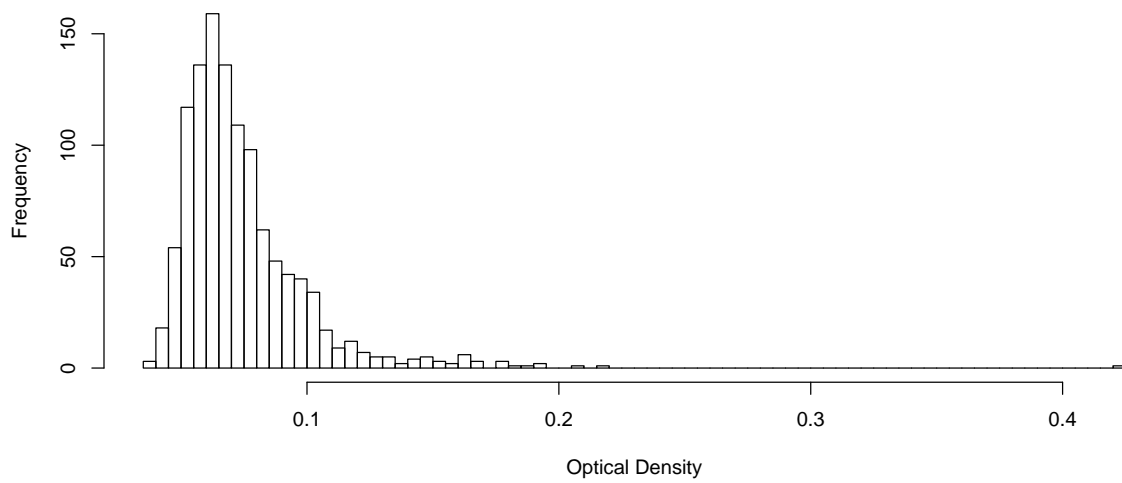
for different distributions for different subpopulations within the same model, the applicability of the tool to a broader range of applications could be achieved.

Comparing the analytic cutoff tool against existing methods of determining diagnostic cutoffs would also provide guidance for best practices in the field as we head toward elimination of NTDs. Though various well-characterized positive and negative serum samples do exist, setting a universal cutoff based on these specimens is not advisable. Inter-lab variability in methods, materials, and equipment makes a single universal cutoff inappropriate. The 3 standard deviations method, dose-response sigmoidal curves and ROC analyses all have their advantages, and a head-to head-comparison would provide insight into the degree to which these results differ.

Developing cutoffs for diagnostic tests in the absence of a gold standard presents a hurdle for researchers in a variety of fields. Our analytic cutoff tool provides a data-based and objective method of determining these cutoffs that takes into account the reality of more than two possible subpopulations in the data, as well as different underlying distributions of data. It also provides conditional probabilities and indeterminate ranges, which allow one to quantify certainty of result classification and make programmatic decisions based upon a richer understanding of the diagnostics that we rely upon to achieve control, elimination and eradication of disease.

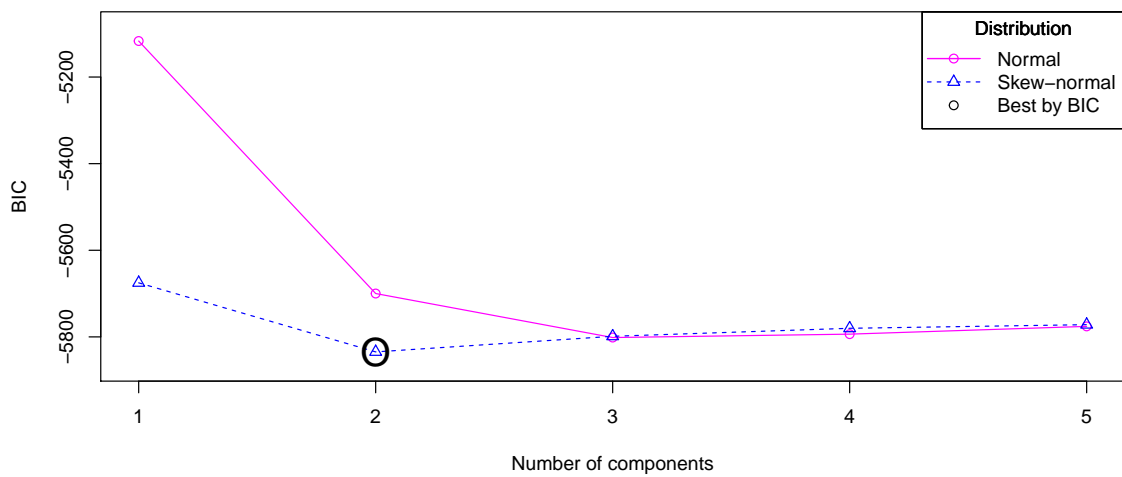
Figures

Figure 1



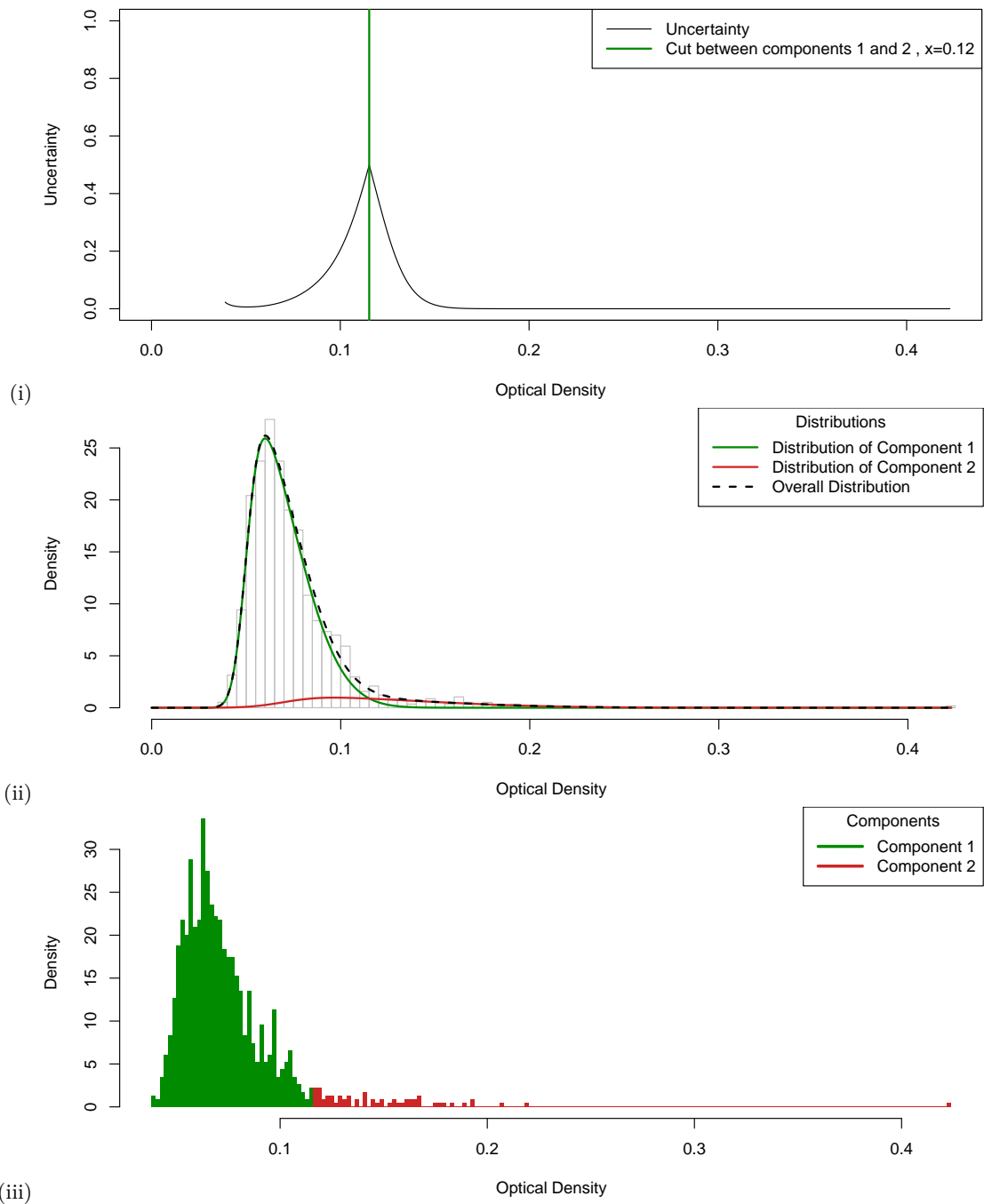
Histogram of standardized optical density readings for example A.

Figure 2



Bayesian Information Criterion (BIC) for 1 to 5 components with normal or skewnormal distributions for example A.

Figure 3

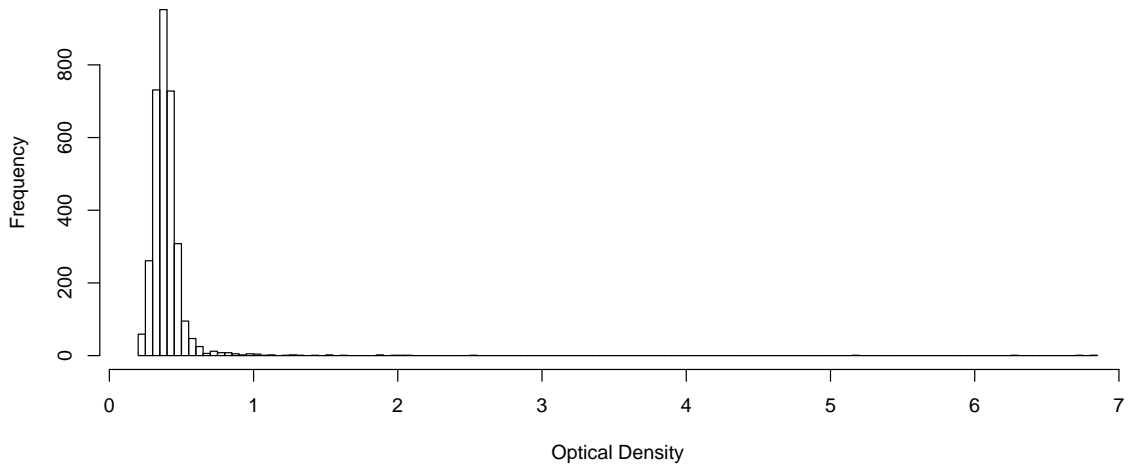


(i) Uncertainty of classification for a two-component skew-normal model with the cutoff set at peak uncertainty for example A.

(ii) Distribution of the components from the BIC optimal for example A two-component skew normal model, overlaid on top of the histogram of the observed data. Here the green denotes the negative component and the green denotes the positive component.

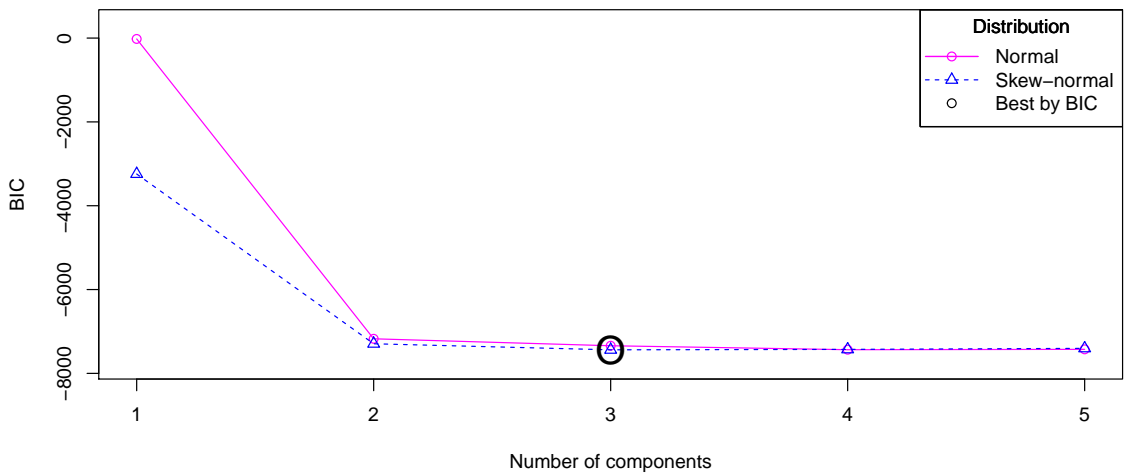
(iii) Histogram of the classification of the observed data for example A. Here the green results would be classified as negative and the red results would be classified as positive.

Figure 4



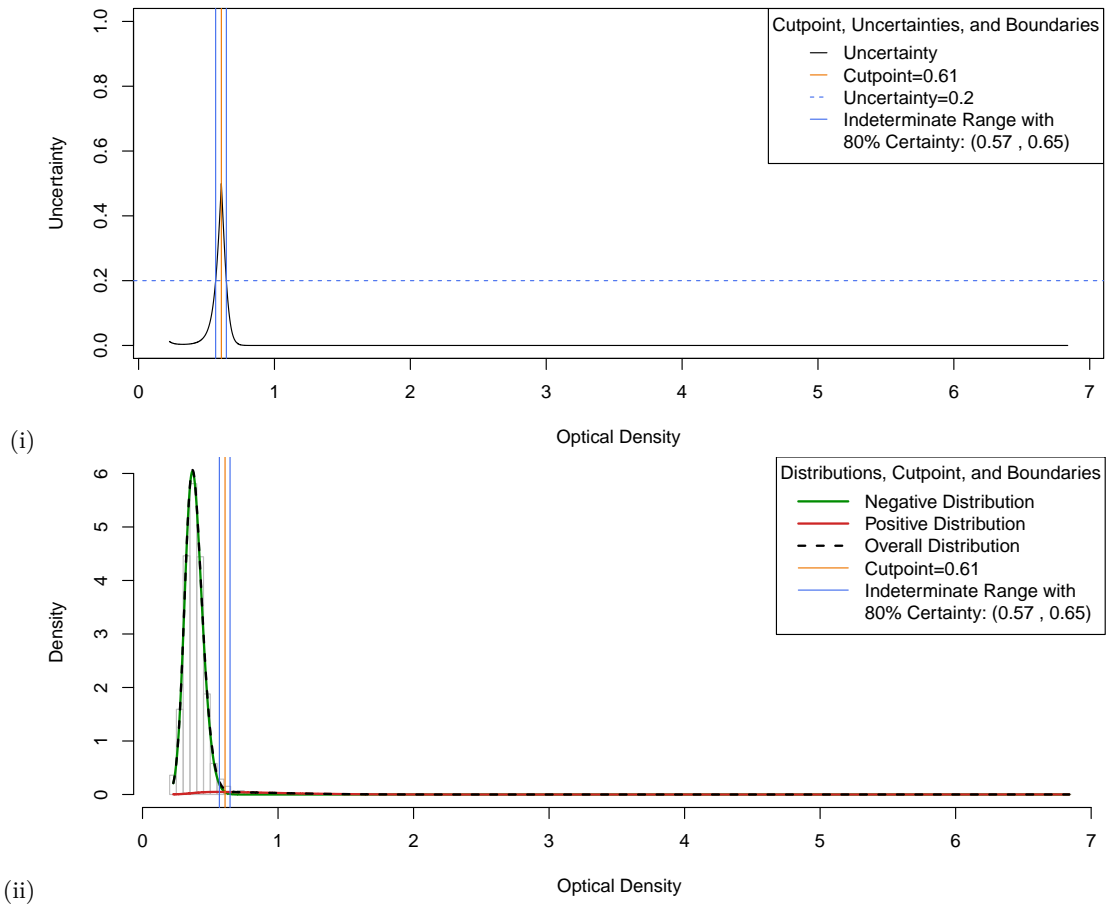
Histogram of standardized optical density readings for example B.

Figure 5



Bayesian Information Criterion (BIC) for 1 to 5 components with normal or skewnormal distributions for example B.

Figure 6



(i) Uncertainty of classification for a three-component skew-normal model with the cutoff set between distributions 1 and 2 for example B. Additionally, the cutoff and bounds of an 80% indeterminate range are displayed.

(ii) Distribution of the components from the BIC optimal for example B, a three-component skew-normal model with the cutoff set between components 1 and 2 for example B, overlaid on top of the histogram of the observed data. Here the green denotes the negative group (comprised of the first component) and the red denotes the positive group (made up of the second and third components).

References

1. Weil GJ, Steel C, Liftis F, Li BW, Mearns G, Lobos E, et al. A rapid-format antibody card test for diagnosis of onchocerciasis. *J Infect Dis.* Oxford University Press; 2000;182: 1796–1799. doi:10.1086/317629
2. Steel C, Golden A, Kubofcik J, LaRue N, De los Santos T, Domingo GJ, et al. Rapid *Wuchereria bancrofti*-specific antigen Wb123-based IgG4 immunoassays as tools for surveillance following mass drug administration programs on lymphatic filariasis. *Clinical and Vaccine Immunology.* 2013;20: 1155–1161. doi:10.1128/CVI.00252-13
3. Steel C, Golden A, Stevens E, Yokobe L, Domingo GJ, De los Santos T, et al. Rapid point-of-contact tool for mapping and integrated surveillance of *Wuchereria bancrofti* and *Onchocerca volvulus* infection. *Clinical and Vaccine Immunology.* 2015;22: 896–901. doi:10.1128/CVI.00227-15
4. Kubofcik J, Fink DL, Nutman TB. Identification of Wb123 as an Early and Specific Marker of *Wuchereria bancrofti* Infection. Mackenzie CD, editor. *PLoS Neglected Tropical Diseases.* Public Library of Science; 2012;6: e1930. doi:10.1371/journal.pntd.0001930
5. Priest JW, Jenks MH, Moss DM, Mao B, Buth S, Wannemuehler K, et al. Integration of Multiplex Bead Assays for Parasitic Diseases into a National, Population-Based Serosurvey of Women 15-39 Years of Age in Cambodia. Mutapi F, editor. *PLoS Neglected Tropical Diseases.* World Health Organization; 2016;10: e0004699. doi:10.1371/journal.pntd.0004699
6. Lammie PJ, Moss DM, Brook Goodhew E, Hamlin K, Krolewiecki A, West SK, et al. Development of a new platform for neglected tropical disease surveillance. *International Journal for Parasitology.* 2012;42: 797–800. doi:10.1016/j.ijpara.2012.07.002
7. Burbelo PD, Leahy HP, Iadarola MJ, Nutman TB. A four-antigen mixture for rapid assessment of *Onchocerca volvulus* infection. Bradley J, editor. *PLoS neglected tropical diseases.* World Health Organization; 2009;3: e438. doi:10.1371/journal.pntd.0000438
8. Everitt B. An introduction to finite mixture distributions. *Statistical Methods in Medical Research.* 1996;5: 107–127. doi:10.1177/096228029600500202
9. Migchelsen SJ, Martin DL, Southisombath K, Turyaguma P, Heggen A, Rubangakene PP, et al. Defining Seropositivity Thresholds for Use in Trachoma Elimination Studies. Johnson C, editor. *PLOS Neglected Tropical Diseases.* R Foundation for Statistical Computing; 2017;11: e0005230.

doi:10.1371/journal.pntd.0005230

10. Golden A, Faulx D, Kalnoky M, Stevens E, Yokobe L, Peck R, et al. Analysis of age-dependent trends in Ov16 IgG4 seroprevalence to onchocerciasis. *Parasites & Vectors*. *Parasites & Vectors*; 2016;9: 338. doi:10.1186/s13071-016-1623-1

11. Zhang H, Huang Y. *Finite Mixture Models and Their Applications : A Review*. *Austin Biometrics and Biostatistics*. 2015;2: 1–6.

12. Vyse a J, Gay NJ, Hesketh LM, Pebody R, Morgan-Capner P, Miller E. Interpreting serological surveys using mixture models: the seroepidemiology of measles, mumps and rubella in England and Wales at the beginning of the 21st century. *Epidemiology and infection*. 2006;134: 1303–1312. doi:10.1017/S0950268806006340

13. Parker RA, Erdman DD, Anderson LJ. Use of mixture models in determining laboratory criterion for identification of seropositive individuals: application to parvovirus B19 serology. *Journal of Virological Methods*. 1990;27: 135–144. doi:10.1016/0166-0934(90)90130-8

14. Fujii Y, Kaneko S, Nzou SM, Mwau M, Njenga SM, Tanigawa C, et al. Serological Surveillance Development for Tropical Infectious Diseases Using Simultaneous Microsphere-Based Multiplex Assays and Finite Mixture Models. *PLoS Neglected Tropical Diseases*. 2014;8. doi:10.1371/journal.pntd.0003040

15. McLachlan G, Peel D. *Finite Mixture Models [Internet]*. Wiley; 2000. p. 456. doi:10.1002/0471721182

16. Fraley C, Raftery, Adrian E. Model-based Methods of Classification : Using the mclust Software in Chemometrics. *Journal Of Statistical Software*. 2007;18: 1–13. doi:10.18637/jss.v018.i06

17. Prates MO, Cabral CRB. mixsmsn: Fitting Finite Mixture of Scale Mixture of Skew-Normal Distributions Marcos. *Journal of Statistical Software*. 2013;54. doi:http://dx.doi.org/10.18637/jss.v054.i12

18. Golden A, Stevens EJ, Yokobe L, Faulx D, Kalnoky M, Peck R, et al. A Recombinant Positive Control for Serology Diagnostic Tests Supporting Elimination of *Onchocerca volvulus*. *PLoS Neglected Tropical Diseases*. 2016;10: 1–16. doi:10.1371/journal.pntd.0004292

19. Lobos E, Weiss N, Karam M, Taylor H, Ottesen E, Nutman T. An immunogenic *Onchocerca volvulus* antigen: a specific and early marker of infection. *Science*. 1991;251: 1603–1605. doi:10.1126/science.2011741