Linking multiple serological assays to infer dengue virus infections from paired samples using mixture models

Marco Hamins-Puértolas[1], Darunee Buddhari[2], Henrik Salje[3,4], Angkana T. Huang[2,3], Taweewun Hunsawong[2], Derek A.T. Cummings[4,5], Stefan Fernandez[2], Aaron Farmer[6], Surachai Kaewhiran[7], Direk Khampaen[7], Anon Srikiatkhachorn[8,9], Sopon Iamsirithaworn[7], Adam Waickman[10], Stephen J. Thomas[10,11], Timothy Endy[10,12], Alan L. Rothman[8], Kathryn B. Anderson[10,11*], Isabel Rodriguez-Barraquer[1*]

* Indicates shared senior authorship

1. Department of Medicine, University of California, San Francisco, USA
2. Department of Virology, WRAIR-Armed Forces Research Institute of Medical Sciences, Thailand
3. Department of Genetics, University of Cambridge, UK
4. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, USA
5. Department of Biomedical Engineering, Whiting School of Engineering, Johns Hopkins University, USA
6. Center for Infectious Diseases Research, Walter Reed Army Institute of Research, Silver Spring, MD
7. Ministry of Public Health, Tiwanond, Nonthaburi, Thailand
8. Institute for Immunology and Informatics, Department of Cell and Molecular Biology, University of Rhode Island, Providence, RI
9. Faculty of Medicine, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.
10. Department of Microbiology and Immunology, SUNY Upstate Medical University, Syracuse, NY, USA.
11. Institute for Global Health and Translational Sciences, SUNY Upstate Medical University, Syracuse, NY, USA.
12. Coalition for Epidemic Preparedness Innovations (CEPI), Washington DC, USA

Correspondence to: marco.hamins-puertolas@ucsf.edu

Abstract

Dengue virus (DENV) is an increasingly important human pathogen, with already half of the globe's population living in environments with transmission potential. Since only a minority of cases are captured by direct detection methods (RT-PCR or antigen tests), serological assays play an important role in the diagnostic process. However, individual assays can suffer from low sensitivity and specificity and interpreting results from multiple assays remains challenging, particularly because interpretations from multiple assays may differ, creating uncertainty over how to generate finalized interpretations. We develop a Bayesian mixture model that can jointly model data from multiple paired serological assays, to infer infection events from paired serological data. We first test the performance of our model using simulated data. We then apply our model to 677 pairs of acute and convalescent serum collected as a part of illness and household investigations across two longitudinal cohort studies in Kamphaeng Phet, Thailand, including data from 232 RT-PCR confirmed infections (gold standard). We compare the classification of the new model to prior standard interpretations that independently utilize information from either the hemagglutination inhibition assay (HAI) or the enzyme-linked immunosorbent assay (EIA). We find that additional serological assays improve accuracy of infection detection for both simulated and real world data. Models incorporating paired IgG and IgM data as well as those incorporating IgG, IgM, and HAI data consistently have higher accuracy when using PCR confirmed infections as a gold standard (87-90% F1 scores, a combined metric of sensitivity and specificity) than currently implemented cut-point approaches (82-84% F1 scores). Our results provide a probabilistic framework through which multiple serological assays across different platforms can be leveraged across sequential serum samples to provide insight into whether individuals have recently experienced a DENV infection. These methods are applicable to other pathogen systems where multiple serological assays can be leveraged to quantify infection history.

Introduction

Dengue virus (DENV) is a flavivirus spread by *Aedes* mosquitoes, primarily in tropical and sub-tropical regions of the globe. It is considered to be one of the leading causes of morbidity across the world [1,2]. With shifting geographical distributions of its vector driven by changes in climate, more populations will likely be at risk of infection in the coming decades[3–5]. A large proportion of infections are subclinical [6–8] meaning that estimates of infections can have large uncertainties when derived from surveillance for symptomatic cases alone. Diagnosis of dengue in clinical settings also remains challenging, since only a minority of cases are captured by direct detection methods (RT-PCR or antigen tests), and interpretation of serological assays may be ambiguous if multiple assays disagree in their interpretations. Improved accuracy of diagnostic tools is important for multiple reasons; detecting missed infections, estimation of epidemiological parameters that provide insight into drivers of risk for both infection and disease, and when implementing and evaluating the efficacy of countermeasures like vector control or vaccine deployment [9–12].

Serological data is most commonly analyzed using defined cut offs for both seropositivity and seroconversion [13,14]. However, these approaches may not be robust to variation in underlying population immunity and relative sampling times across populations that may lead to more

misclassification [15]. In particular, differences in prior flavivirus exposure (e.g. vaccination programs, recent epidemics, original antigenic sin), sample timing[13], infecting serotype [16,17], underlying dengue strain utilized for antigen development in each assay, and whether the patient is undergoing a primary or post-primary infection[18] all could impact the sensitivity and specificity of this cutoff approach. Having low sensitivity due to false negatives may lead to underestimation of burden while low specificity due to false positives will lead to a misattribution of infection to dengue and could impact control efforts as well as vaccine deployment. In turn, optimizing diagnostic tools is vital in the management of dengue at both the population and individual patient scale. Mixture models have recently become more common in the field, allowing for data driven interpretations of serological data,[19–25] classification of DENV infections as primary or post-primary (secondary, tertiary, quaternary) [26,27] as well as to estimate the force of infection of DENV [26,28]. This is particularly advantageous as these methods can be applied to a novel dataset and provide estimates catered directly to the population of interest.

Similarly, paired sera (e.g. sera obtained at acute and convalescent time-points) are also usually evaluated using a cut off metric to identify seroconversions or boosting events characteristic of infections. This cut off is typically a fourfold rise or greater for HAI, IgG, and IgM titers between the paired sera, [29,30] since it is expected that measurement error may lead to observed changes of up to a two fold rise. These cut offs have been widely adopted across serological analyses [14,30,31] even though it is acknowledged that their performance may be affected by factors mentioned previously including the timing of sample collection and previous flavivirus exposure. Alternative statistical approaches have been developed to move away from these cut off methods and have been successfully applied to paired sera from a single serological assay to define seroconversions [21,32]. However, these methods do not allow for the incorporation of multiple serological assays which can, in practice, increase the amount of available information through which estimates can be made. Using results from different serological assays has been shown to lead to tradeoffs in sensitivity and specificity [33–36], a balancing act characteristic of this classification problem.

Here we develop a multivariate gaussian mixture model approach for paired serological data (acute and convalescent samples) to probabilistically classify seroconversions from multiple serological assays. We test our method using simulated data, apply it to data from two longitudinal cohorts in Thailand, and compare it to existing approaches currently implemented in most DENV surveillance studies. These methods are flexible in the number of data streams that can be incorporated as well as allowing for the probability of infection to be assigned to an individual. Accurately diagnosing infections is a vital step in determining risk factors for infection and severe outcomes.

Table 1. Dataset characteristics for analyzed data split by source as well as across the entire dataset.

Covariate	Source		
	KPS1 (n=202)	KFCS (n=475)	Total (n=677)

Age (years)	Mean (SD)	10.1 (1.5)	11.0 (13.7)	10.7 (11.6)
	Median [Min, Max]	10 [7,15]	6 [0,91]	9 [0,91]
	[1,5)	0 (0%)	182 (38.3%)	182 (26.9%)
	[5,18)	202 (100%)	206 (43.4%)	408 (60.3%)
	[18,30)	0 (0%)	47 (9.9%)	47 (7.0%)
	[30,100+)	0 (0%)	36 (7.6%)	36 (5.3%)
Hospitalized Dengue	+	40 (19.8%)	48 (10.1%)	88 (14.9%)
	-	162 (81.2%)	427 (89.9%)	589 (85.1%)
RT-PCR	+	157 (77.7%)	75 (15.8%)	232 (34.3%)
	-	45 (22.3%)	400 (84.2%)	445 (65.7%)
HAI	Primary infection	79 (39.1%)	31 (6.5%)	110 (16.2%)
	Post-primary infection	109 (54.0%)	64 (13.5%)	173 (25.6%)
	No infection	8 (3.9%)	369 (77.7%)	377 (55.7%)
	Recent infection	6 (3.0%)	11 (2.3%)	17 (2.5%)
EIA	Primary infection	8 (3.9%)	16 (3.4%)	24 (3.5%)
	Post-primary infection	170 (84.2%)	71 (14.9%)	241 (35.6%)
	No infection	23 (11.4%)	378 (79.6%)	401 (59.2%)
	JEV infection	1 (0.5%)	10 (2.1%)	11 (1.6%)

129

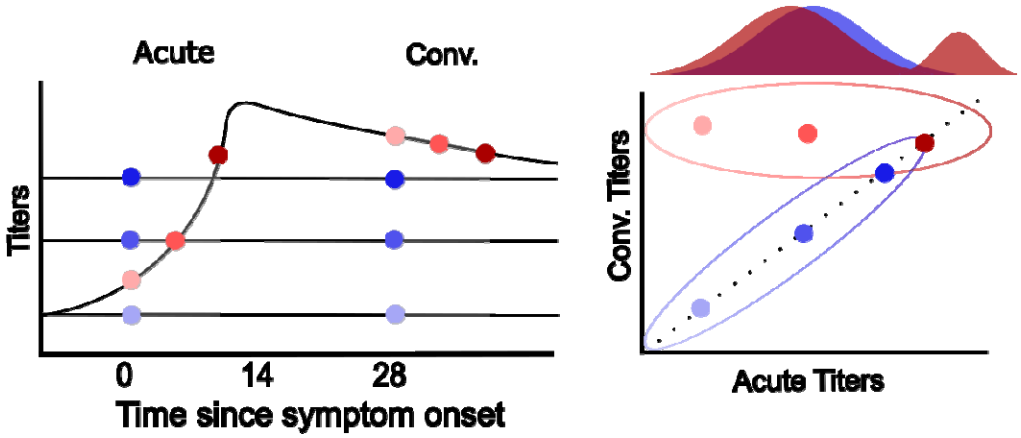

Figure 1. Theoretical relationship between time since symptom onset and titers in infected (red) and uninfected (blue) individuals. These titers are mapped onto a two-dimensional representation of acute and convalescent titers. Ellipsoids represent the expected distribution of acute and convalescent titers for infected and uninfected individuals while the underlying marginal distribution of these acute titers is presented at the top. Opacity of each pair of points represents titers at each acute sample. For the infected individual (red) the lightest point represents how titers compare when sampling close to the date of symptom onset while the darkest red point represents how titers compare when sampling more than a week from symptom onset, often defined as a "recent" infection. The marginal distribution of acute titers colored by infected and uninfected individuals is presented above the acute and convalescent titer plot to demonstrate how these recent infections can present.

Methods

Mixture model

Gaussian mixture models have successfully been applied to individual antibody titers in multiple pathogens including DENV and can provide an unbiased and unsupervised approach to quantifying seropositivity in a population of interest [19–23,28]. These methods implement a hierarchical probabilistic clustering framework where the data generating process is assumed to be produced in some predefined manner while associated parameters are subsequently estimated. We extend previous methodologies that either investigated multiple serological assays at a single time point, or a single assay at multiple time points to allow for the analysis of multiple serological assays measured at an acute and convalescent sample. The objective being to determine whether the patient is or is not acutely infected with the pathogen of interest, in this case DENV. In Figure 1 a set of theoretical antibody trajectories for a single assay are presented for infected (red) and uninfected (blue) individuals. Given an acute and convalescent pair of samples we developed a method that can disentangle these two groups of infected and uninfected individuals. Here we will outline the methods used to infer infections from individual samples from a single assay, paired samples from a single assay, and then how to incorporate multiple assays taken at paired samples.

Mixture models to serological data from a single acute or convalescent sample

We fit univariate mixture models to acute or convalescent samples independently (acute and convalescent geometric mean HAI, IgG, and IgM). We call these the acute or convalescent (GM HAI or IgG) univariate mixture model hereafter. This is done using a single univariate mixture model with M mixtures each of which has a mean and variance of μ_m^R and σ_m^R respectively. The mixture with the highest average titers contained individuals who had the most recent boost in titers. These models were implemented to demonstrate the performance of each serological assay when only acute or convalescent samples are available. When fitting these models, we choose the number of mixtures to be two, one for the infected and one for the uninfected classes.

Mixture models for serological data from paired acute and convalescent samples

For paired serological samples we will first outline how this model is derived under the assumption that acute and convalescent samples were analyzed with a single assay. For each individual, i, in the sampled population of size n, we have measured acute and convalescent titers (y_i^A and y_i^C respectively). Let us assume that the likelihood function for convalescent titers is dependent on two normal distributions, one representing the distribution of convalescent titers given acute titers if an infection occurs and one when no infection occurs. We write this as follows:

$$f(y_i^C) = \theta_0\ Normal(y_i^C | y_i^A, \sigma_0) + \theta_1\ Normal(y_i^C | \mu + \beta\ y_i^A, \sigma_1\ ),$$

Where $\theta_0 + \theta_1 = 1$, and θ_0 represents the probability of not being infected while θ_1 represents the probability of being infected. Let μ represent the minimum baseline boost of serological titers and β be the relative increase in convalescent titers as a function of acute serological titers. In addition, σ_0 and σ_1 represent the variance of each normal distribution around their respective means..

For the multivariate case, where more than one assay's titers are incorporated, we extend the above to a multivariate normal distribution (MVN). We assume that each individual has acute and convalescent samples for k total assays measured as y_{ij}^A and y_{ij}^C, where j represents the assay number. Let $\hat{y}_i^A = (y_{i1}^A, y_{i2}^A, \ldots, y_{ik}^A)^T$ and $\hat{y}_i^C = (y_{i1}^C, y_{i2}^C, \ldots, y_{ik}^C)^T$ be the vectors containing all the observed acute and convalescent serological responses for an individual across all assays. These will contribute to the likelihood via two mixtures that are assumed to be multivariate normally distributed:

$$f(\hat{y}_i^C) = \theta_0\ MVN(\hat{y}_i^A, \Sigma_0) + \theta_1\ MVN(\hat{\mu} + \hat{\beta} \circ \hat{y}_i^A, \Sigma_1)$$

Again, let θ_0 be the probability of not being infected while θ_1 is the probability of being infected such that $\theta_0 + \theta_1 = 1$. Let $\hat{\mu}$ and $\hat{\beta}$ now be vectors the length of the number of assays sampled (k) that represent the minimum baseline boosted serological titers and the relative increase in convalescent titers as a function of acute serological titers. Finally, let Σ_0 and Σ_1 be the covariance matrices associated with each mixture.

Combining with univariate mixture models to identify recent infections

We fit additional univariate models for the identification of recent infections (as illustrated in dark red in Figure 1). These were added to identify infections that might not be identified from the

paired acute/convalescent data due to the acute sample being obtained after the boosting has occurred. This approach utilized the same mixture model approach outlined above with five mixtures to ensure differentiation of the high acute titers. For the identification of these recent infections we used a stepwise approach to combine the acute univariate model with the paired multivariate model. We first determined if an infection can be considered recent using the univariate mixture model (any individual with an assigned probability of infection greater than 50%), and then for those that were not recent we found the probability that they were classified as an infection using the multivariate mixture model. We attempted to fit alternative approaches that incorporated the time since symptom onset but found these did not perform as well as this stepwise approach as multiple recent infections were missed (Figure S1).

Simulation of serological data

In order to assess the performance of our proposed model we first used simulated data generated using serosim[37], a flexible R package capable of simulating serological data which we ran in R version 4.3.3 [38]. We simulated the serological response to three serological assays for all individuals (Figure 2). The population as a whole underwent an extended period of low exposure to the pathogen during which individuals accumulated varying levels of immunity. This was done to simulate the distribution of immunity expected in a population. Subsequently, a short period of high exposure occurred that led to approximately half of the population being infected with the pathogen. Finally, samples reflective of an acute and convalescent time point were taken. We designated mean and variances around the expected titer boost for all three assays (IgG, IgM, and GM HAI), upper and lower bounds of detection for each assay, and the relationship between an individual's initial titers and the expected boost in titers they will experience. In addition the relationship between titers and protection from infection were defined to better reflect reality. The parameters we chose for all antibody kinetics can be found in Table S1. Note that this approach only simulates an immediate boost post infection followed by a decay in titers meaning the timing of antibody kinetics in response to primary and post-primary infection do not vary for any measure antibody titer. Note that IgG and IgM were simulated on the linear scale while HAI was simulated on the log scale while all were analyzed on the log scale. Three different amounts of observational noise were analyzed to quantify how measurement error impacts model performance. The code we utilized for these simulations can be found in the github repository github.com/marcohamins/linking-multiple-sero-assays.

Cohort data description

Data for this analysis was collected from two cohort studies. Both studies were conducted in Kamphaeng Phet province in Northern Thailand and their characteristics are presented in Table 1. The first cohort study, the Kamphaeng Phet Study (KPS-1), recruited children from 12 different primary schools and ran from 1998 to 2002 [39,40]. The study involved blood draws four times a year (January, June, August, and November) as well as active surveillance performed during the rainy season (June - November), where any children who experienced a febrile illness were evaluated and had acute and convalescent blood samples taken, with an interval of 14 days between collections. The acute blood draws from active surveillance underwent dengue RT-PCR while both acute and convalescent samples underwent serological testing (HAI, IgM, and IgG) for DENV and Japanese Encephalitis virus (JEV).

The second cohort study, the Kamphaeng Phet Family Cohort Study (KFCS) began in 2015 and remains active. For this analysis we included all data collected between 2015 and 2022. During this period, 494 multigenerational households were enrolled consisting of 3220 individuals who provided yearly blood samples through routine visits. In addition, if an individual in the cohort reported a febrile event at any point during the year, an acute and two convalescent samples were obtained, with intervals of approximately 14 and 28 days. Only episodes where both the acute and first convalescent sample were collected were utilized in this analysis to ensure consistent sample timing across datasets. Similar to the previous cohort, acute samples underwent RT-PCR while acute and convalescent samples underwent HAI, IgM, and IgG testing for DENV and JEV. If this individual's acute sample was a confirmed DENV infection via RT-PCR, acute and convalescent specimens were collected from all other members of the household and underwent identical serological and molecular testing.

It is important to highlight that the data from these two cohorts were taken from two different subpopulations in Kamphaeng Phet and had different inclusion protocols. Individuals in KPS1 are those who had a confirmed DENV infection via RT-PCR, HAI, or EIA. However, in KFCS if someone in a household had a RT-PCR confirmed infection the rest of the household were also sampled in a cluster investigation. This can be found in the higher percentage of cases that were RT-PCR confirmed in KPS1 compared to KFCS.

The protocol for KPS1 was approved by the Human Use Review and Regulatory Agency of the Office of the Army Surgeon General, the Institutional Review Board of the University of Massachusetts School of Medicine, and the Thai Ethical Review Board of the Ministry of Public Health, Thailand (protocol #654). The study protocol for KFCS was approved by Thailand Ministry of Public Health Ethical Research Committee, Siriraj Ethics Committee on Research Involving Human Subjects, Institutional Review Board for the Protection of Human Subjects State University of New York Upstate Medical University, and Walter Reed Army Institute of Research Institutional Review Board (protocol #2119).

Laboratory methods and serological interpretations

A complete outline of the laboratory methods used for each serological assay are outlined in the supplemental information and can be found in Hamins-Puertolas et al. (2023)[8]. When using HAI, infection is usually defined by a 4-fold or greater rise in HAI titers for any of the four DENV serotypes and JEV when comparing acute and convalescent sera. A higher seroconversion titer for JEV than for any DENV serotype is interpreted as a JEV infection event. All HAI titers we report are geometric mean HAI titers (GM HAI). For EIA units, an IgM ≥ 40 is usually used as a positive cut-off value. Evidence of DENV infection was classified by a ratio of DENV IgM/JEV IgM ≥ 1.0, and JEV infection when the ratio was < 1.0. Further classification into primary and post-primary for each assay is outlined in the supplemental information.

Model fitting

All models were estimated using Rstan v2.26.23 and Cmdstanr v0.5.3 with four independent chains, each of length 1000 after 2000 discarded for warm-up. Convergence was defined to be

when R-hat<1.1 and the effective sample size >300 for model parameters, after which posterior draws were pooled for reported parameter estimates. To assess accuracy, we performed four-fold cross-validation. Testing set predictions were used to assess accuracy.

We removed any individuals whose GM HAI titers had reduced between acute and convalescent samples by more than 25%. These waning titers may reflect a DENV infection in the previous few months, either missed by the active surveillance program or subclinical. However, since no determination can be made, we removed these individuals from classification (n=41) leading to a total of 677 total samples for which accuracy is assessed.

Assessing model performance

To directly compare the results from each of the models against both the simulated and the cohort data, we utilized sensitivity, specificity, area under the reporter operator curves (ROC AUC), and F1-score. The F1-score is a measure of predictive performance that weights the number of correctly identified infections (TP) by the number of missed infections (FN) and incorrectly identified non-infection events (FP). The precise formula is TP/(TP+0.5*(FP+FN)). For the Thai data we presented these evaluation metrics for two definitions of what a true positive infection was for classification purposes. The first only considered RT-PCR positive infections while the second definition considered an individual to be infected if either RT-PCR, HAI, or EIA were positive.

Results

Assessment of model performance to identify infections using simulated data

We first evaluated the performance of models on multiple sets of simulated data. We simulated paired (acute/convalescent) serological data for three independent biomarkers (GM HAIs, IgM and IgG) for 500 individuals of which approximately 50% had been recently infected.

We first fit models incorporating data from the three simulated assays (GM HAIs, IgM and IgG) to try and infer the infection status of each individual (Figure 2). The model incorporating all three assays had an F1 score ranging from 98.0-93.9% as observational noise was increased. This model consistently outperformed the models that incorporated any one or two of the assays when comparing sensitivity, F1, and AUC ROCs (Table S2). Specificity was also highest in this model in all but one observational noise simulation. We then fit models including only combinations of the assays or single assays. Models including any two combined assays outperformed any singular assay, with increases in F1 scores and AUC ROC ranging from 0.1 to 27.3 and 0.1 to 17.9 percentage points respectively across the various observational noise simulations. Two models incorporating just IgM were unable to be fit at higher levels of observational noise, likely due to a loss of signal from titer boosts post infection. All models performed well in scenarios ranging from small to large amounts of observational noise, with decreases in model performance as a function of increasing observational noise.

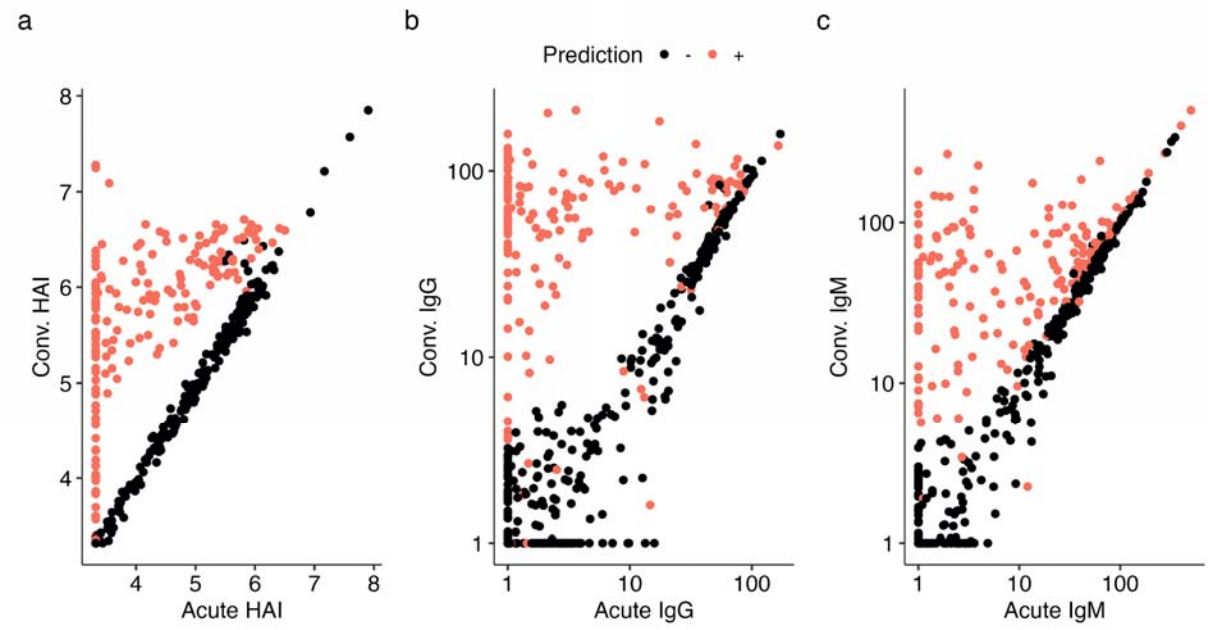

Figure 2. Simulated serological data using `serosim` and subsequent results using the full mixture model on (a) HAI, (b) IgG and (c) IgM data with additional observational noise. For each we present the predicted infection status using the mixture model approach. Parameters chosen for forward simulation are presented in Table S1 and resulting metrics of accuracy are presented in Table S2.

Application of the model to paired serological data from Thailand

We then assessed the performance of our model on data from the two longitudinal cohorts based in Kamphaeng Phet, Thailand, where 718 paired sera from 585 individuals were taken that had 232 RT-PCR confirmed infections. We assessed the performance under two different "gold-standard" definitions of infection, described in the Methods section. This allowed us to compare a more specific gold standard (infections confirmed by RT-PCR) where few false positives and more false negatives are expected, to one that is more sensitive (infections defined by RT-PCR or serological results).

For the first definition, using RT-PCR results, we found that all models except one (Model 2 with just IgM data) performed as well if not better than the previously implemented classifications based solely on HAI (HAI interpretation), IgM and IgG values (EIA interpretation), or combinations of these interpretations (Table 2). For this definition we found that the best performing approach according to the F1 metric and AUC ROC was the model that incorporated information from just IgM and IgG data. However, the mixture that combined all data sources had better sensitivity (96.1% in Model 9 vs 91.8% in Model 6), which may be a more relevant metric of performance for this highly specific definition. With the more sensitive definition of infection that incorporates RT-PCR, HAI, and EIA information we found that using just GM HAI data (Model 3) resulted in the highest F1 and AUC ROC scores. All models with two or more data streams except Model 6 (IgG and IgM) had high F1 and AUC ROC scores (>90%). Note that the comparisons to the HAI- and ELISA-based interpretations for this gold-standard

definition of infection are not meaningful since the interpretation of infection is part of the gold-standard definition, however we reported these values for full transparency (i.e. specificity compared to this definition was 100% in all of these clinical interpretations by construction).

Overall, we found that combining multiple data streams using this mixture model approach consistently led to higher performance when compared to currently implemented interpretations. Model 9 (GM HAI, IgG, IgM, and an acute GM HAI univariate mixture model, Figure 3) consistently performed well across both gold-standard definitions of infection (Table 2), where it had high sensitivity (96.1% and 85.2%) and specificity (86.2% and 97.7%). We also found that when a DENV infection was classified by one, two or all three assays (RT-PCR, HAI, and EIA illustrated in Figure S2), this model identified an infection in 48.7% (19/39), 97.0% (65/67), and 100% (208/208) of individuals respectively (Figure S3, Table S3). This model also identified three additional infections that were missed by all three interpretations, two of which corresponded to high acute and convalescent GM HAI titers while one was near the decision boundary for the IgM paired samples and had an inferred probability of infection of 57% (Figure S4). In addition, Model 7 (IgM, IgG, and a separate acute IgG classifier), consistently performed as well or better than the currently implemented algorithms. In individuals where one, two, or all three serological assays (RT-PCR, HAI, and EIA) classify there to be an infection, this model correctly identified an infection in 30.8% (12/39), 85.1% (57/67), and 97.1% (202/208) of individuals respectively with no additional infections inferred when all three serological assays did not classify there to be an infection.

In addition to performing equally or better than a standard algorithm, a useful feature of the proposed mixture model methodology is that beyond classifying individuals, it provides an estimate for the probability that an individual belongs to the infected class, and thus the "certainty" of classification (Figure S5). In the aforementioned Models 7 and 9 we find that 85.5% (579/677) and 96.9% (656/677) respectively of paired sera have an inferred probability of infection below 5% or above 95%, indicating high certainty in a sizeable majority of cases.

Finally, we compared the performance of our paired acute/convalescent models to univariate mixture models independently fit on acute or convalescent data for each serological assay (Table S4). We found that each acute data stream performed poorly against both definitions of infection. The univariate convalescent mixture models performed better than their acute counterparts with F1 and AUC ROC scores ranging from 19.8 to 56.7 and 39.1 to 48.1 percentage points higher respectively. Convalescent IgG and IgM both performed at similarly high levels within these univariate models. For example, when using the RT-PCR definition of infection with F1 scores of 66.4% and 67.4% and AUC ROC scores of 92.6% and 91.3% while the convalescent GM HAI had scores of 56.8% and 90.9%. None of these models performed better than their equivalent models that incorporated both acute and convalescent titers demonstrating the importance of acute samples in providing background immunological context for individuals, but that reliance on acute samples alone is not sufficient for accurate diagnosis.

384 **Table 2**. Comparison of each accuracy metric (sensitivity, specificity, F1, and area under the
385 curve for reporter operator curve (AUC ROC)) on each model. Models are defined by which
386 assays were incorporated (i.e. HAI, IgG, and/or IgM).

Model		Infection definition							
		RT-PCR+ vs. RT-PCR -				RT-PCR, HAI, or EIA + vs. RT-PCR, HAI, and EIA -			
#	Data	Sens.	Spec.	F1 score	AUC ROC	Sens.	Spec	F1 score	AUC ROC
–	HAI clinical interp.	98.3%	82.5%	83.7%	–	*94.8%*	*100%*	*97.4%*	–
–	EIA clinical interp.	90.1%	85.6%	81.8%	–	*84.5%*	*100%*	*91.6%*	–
–	HAI AND EIA clinical interp.	89.7%	88.3%	83.7%	–	*80.3%*	*100%*	*89.1%*	–
–	HAI OR EIA clinical interp.	98.7%	79.8%	81.9%	–	*99.1%*	*100%*	*99.5%*	–
1	IgG	90.5%	93.5%	89.2%	95.9%	72.4%	**100%**	84.0%	96.8%
2	IgM	70.0%	94.8%	77.7%	93.2%	55.8%	99.7%	71.5%	93.7%
3	GM HAI	**97.0%**	80.8%	83.0%	92.9%	**90.3%**	96.5%	**93.1%**	**98.4%**
4	GM HAI + IgM	95.7%	86.5%	86.6%	94.4%	85.5%	99.7%	92.0%	98.2%
5	GM HAI + IgG	96.1%	**97.1%**	86.9%	94.4%	85.5%	**100%**	91.6%	97.9%
6	IgM + IgG	91.8%	93.0%	**89.5%**	**96.1%**	73.9%	**100%**	85.0%	96.4%
7	IgM + IgG + acute IgG mixture	94.4%	85.3%	84.9%	95.0%	86.1%	**100%**	92.5%	96.6%
8	All	96.1%	86.7%	86.8%	94.4%	85.2%	99.7%	91.8%	97.6%
9	All + acute GM HAI mixture	96.1%	86.2%	85.4%	94.5%	85.2%	97.7%	90.6%	97.8%

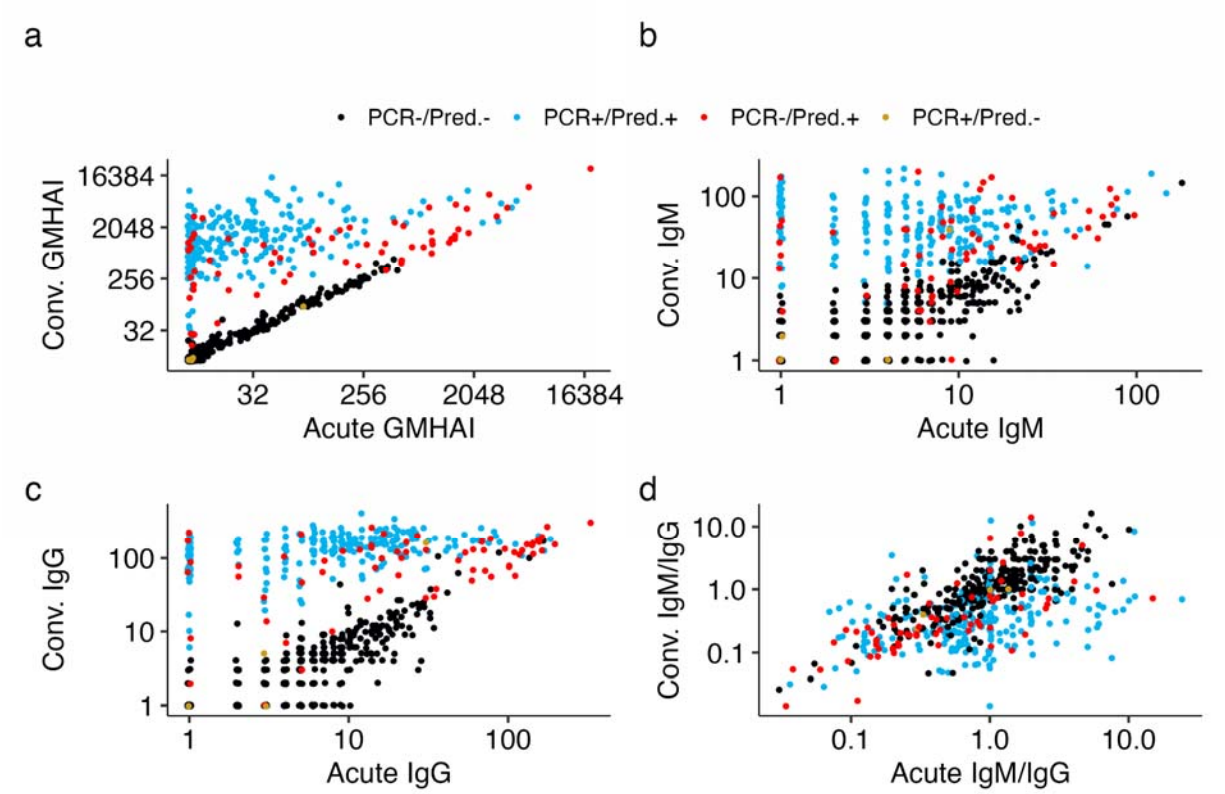

Figure 3. Acute and convalescent serological data from Thai cohorts used in this analyses, colored by true infections status and prediction of infection during model testing. a) Geometric mean titers of a haemagglutination inhibition assay (GM HAI) for all four serotypes of dengue virus. b) Immunoglobulin M (IgM) c) Immunoglobulin G (IgG) d) Ratio of IgM to IgG at both acute and convalescent sera samples.

Estimation of parameters related to antibody kinetics

We also investigated whether these methods can infer parameters associated with antibody kinetics. For the simulated data, we find that we are able to reconstruct the simulated relationship between acute titers and expected boosts across all three assays (Figure S6). The reconstructed relationship remains consistent across varying levels of observational noise. Although estimates do diverge from the ground truth, this is due to the functional form constraint defined by the mixture models. We also estimate this same value for the Thai dataset and find that the expected GM HAI, IgG, and IgM additive titer boost in a fully susceptible individual are 160, 36, and 17 respectively (Figure S7A). The relationship between this expected boost and an individual's acute titers is also quantified (Figure S7B).

Discussion

We have developed a probabilistic framework through which paired serological measurements from multiple assays can be jointly leveraged to predict acute and recent infections. We first used simulated data to demonstrate the performance of these methods under varying levels of

observational noise. We then applied these methods to real world serological data from Kamphaeng Phet, Thailand, and showed that we were able to accurately classify infections as well as quantify boosting patterns as a function of acute titers. In general the incorporation of additional data streams led to higher performance in both simulations and real world applications.

We validated the methodology using simulated paired serological data. We found that observational noise reduces performance, but that these reductions in performance can be minimized when additional data streams are incorporated. In these simulations the observational noise across assays was assumed to be statistically independent, meaning each additional assay provided a unique draw centered around the true serological response. In turn, sharing information across these assays should simultaneously minimize false positives and false negatives. These results are by no means novel as previous work has demonstrated that a simple combination of classification results can lead to increased sensitivity, but at times can at the cost of reduced specificity [33–36]. We show that when applied to simulated data these methods are capable of increasing both sensitivity and specificity as additional immunological data is leveraged under increasing levels of observational noise.

When applied to data from Kamphaeng Phet, Thailand, mixture models are capable of probabilistically assigning whether paired serological results from multiple assays are consistent with an infection. These methods result in high performance when compared to gold standard RT-PCR as well as combined serological interpretations, and can resolve discrepancies in conclusions drawn from RT-PCR, HAI, and EIA methods across a variety of assay combinations. Although these results are context dependent, we find that in Kamphaeng Phet mixture models based solely on the acute and convalescent GM HAI titers perform well in identifying infections and could be used alone in a resource limited setting for this purpose. However, in contexts where HAIs are not available we find that using just paired samples tested with EIA (IgG and IgM) provided some of the best available information from which interpretations could be made. Using acute samples alone leads to drops in classification accuracy while convalescent samples alone perform relatively well in this task even without the additional context that acute samples provide. This itself is not surprising as the majority of rises in titers will be found during the convalescent sample while acute data provides a baseline against which subsequent rises can be compared.

This study's primary strength is the two cohorts of data that provide an ideal dataset on which to assess the performance of this methodology. By combining a cohort focused on individuals around the average of infection (school aged children) and one that sampled individuals across a wide age range and immune profiles using multiple serological assays allowed for this flexible framework to be tested. Due to these study characteristics we were able to show how methods like these can be implemented no matter the number of collected assays, allowing for investigators to infer infections using available acute and convalescent data in a semi-supervised manner. In addition, this provides a systematic way through which the predictive power of each assay can be compared and in turn be used to inform future study design. However, this work has some limitations. One major limitation is that the paired model is at

times not sufficient and two independent models (recent infection and paired model) had to be combined to explain some paired samples with high acute titers. We worked to include additional covariates into the models like baseline titers prior to the illness investigation and time since symptom onset to explain these initially high titers but found these models were still unable to fully capture recent infections. Future work should strive to leverage time since symptom onset as well as available immunological data that provide a baseline measure of immunity. In addition, although this approach was run using a training and testing approach to quantify accuracy in out-of-training samples and we used data from two independent studies, it would be useful to assess the performance of these methods in other populations with different underlying infection histories and immune profiles (i.e. how transferable is classification). Note that these methods have been applied to in house HAI and EIA data that are likely qualitatively consistent with other studies, but means these results cannot simply be exported and applied to other datasets. However, the applicability of this work to other studies both within the DENV literature as well as in different pathogen systems remains.

In summary, our results suggest that mixture models of paired serological data can be a useful tool to probabilistically infer infection status. Optimizing diagnostic tools is vital in the management of dengue at both the individual and population scales. Beyond dengue and other arboviral pathogens, these methods can be applied to any pathogens where paired data on multiple assays are available. As multiplex assays become more common in the field, methods that leverage all available data simultaneously will become increasingly necessary. We have provided code for both the simulation approach as well as the full mixture model methods, both implemented in R, to stimulate further exploration of these methods when applied to serological data.

Acknowledgments

We are thankful for all efforts from the data collection team as well as the children and adults involved in the study. The authors were supported in this work by the following: NIH Grant P01 AI034533: entire team; Military Infectious Disease Research Program (MIDRP): DB, SF, AF, KBA; 1R01AI175941-01: entire team; and R35 GM138361: MHP and IRB; European Research Council 804744: HS. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author Contributions

The study was conceived and designed by DB, HS, DATC, SF, AF, ALR, TPE, SJT, IRB, and KBA. The data were collected by DB, SF, SK, DK, SI, and AS. The analysis and interpretation of results were performed by MHP, HS, DATC, AF, SJT, AW, ALR, TPE, IRB, and KBA. The draft manuscript was prepared by MHP, IRB, and KBA. All authors reviewed the results and approved the final version of the manuscript.

References

1. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496**, 504–507 (2013).

2. Cattarino, L., Rodriguez-Barraquer, I., Imai, N., Cummings, D. A. T. & Ferguson, N. M.

Mapping global variation in dengue transmission intensity. *Sci. Transl. Med.* **12**, eaax4144 (2020).

3. Colón-González, F. J., Fezzi, C., Lake, I. R. & Hunter, P. R. The Effects of Weather and Climate Change on Dengue. *PLoS Negl. Trop. Dis.* **7**, e2503 (2013).
4. Naish, S. *et al.* Climate change and dengue: a critical and systematic review of quantitative modelling approaches. *BMC Infect. Dis.* **14**, 167 (2014).
5. Ebi, K. L. & Nealon, J. Dengue in a changing climate. *Environ. Res.* **151**, 115–123 (2016).
6. Asish, P. R., Dasgupta, S., Rachel, G., Bagepally, B. S. & Kumar, C. P. G. Global prevalence of asymptomatic dengue infections - a systematic review and meta-analysis. *Int. J. Infect. Dis.* **134**, 292–298 (2023).
7. De Santis, O., Bouscaren, N. & Flahault, A. Asymptomatic dengue infection rate: A systematic literature review. *Heliyon* **9**, e20069 (2023).
8. Hamins-Puértolas, M. *et al.* Household immunity and individual risk of infection with dengue virus in a prospective, longitudinal cohort study. *Nat. Microbiol.* **9**, 274–283 (2024).
9. Raafat, N., Blacksell, S. D. & Maude, R. J. A review of dengue diagnostics and implications for surveillance and control. *Trans. R. Soc. Trop. Med. Hyg.* **113**, 653–660 (2019).
10. Pang, J., Hildon, Z. J.-L., Thein, T. L., Jin, J. & Leo, Y. S. Assessing changes in knowledge, attitude and practices on dengue diagnosis and management among primary care physicians after the largest dengue epidemic in Singapore. *BMC Infect. Dis.* **17**, 428 (2017).
11. Kabir, M. A., Zilouchian, H., Younas, M. A. & Asghar, W. Dengue Detection: Advances in Diagnostic Tools from Conventional Technology to Point of Care. *Biosensors* **11**, 206 (2021).
12. Pang, J., Chia, P. Y., Lye, D. C. & Leo, Y. S. Progress and Challenges towards Point-of-Care Diagnostic Development for Dengue. *J. Clin. Microbiol.* **55**, 3339–3349 (2017).
13. Laboratory Diagnosis and Diagnostic Tests. in *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition* (World Health Organization, 2009).

14. World Health Organization. Manual for the laboratory diagnosis and virological surveillance of influenza. *WHO Glob. Influenza Surveill. Netw. Man. Lab. Diagn. Virol. Surveill. Influenza* (2011).

15. Kafatos, G. *et al.* Is it appropriate to use fixed assay cut-offs for estimating seroprevalence? *Epidemiol. Infect.* **144**, 887–895 (2016).

16. Najioullah, F., Viron, F. & Césaire, R. Evaluation of four commercial real-time RT-PCR kits for the detection of dengue viruses in clinical samples. *Virol. J.* **11**, 164 (2014).

17. Klungthong, C. *et al.* Monitoring and improving the sensitivity of dengue nested RT-PCR used in longitudinal surveillance in Thailand. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* **63**, 25–31 (2015).

18. Teoh, B.-T. *et al.* The Use of NS1 Rapid Diagnostic Test and qRT-PCR to Complement IgM ELISA for Improved Dengue Diagnosis from Single Specimen. *Sci. Rep.* **6**, 27663 (2016).

19. Hardelid, P. *et al.* Analysis of rubella antibody distribution from newborn dried blood spots using finite mixture models. *Epidemiol. Infect.* **136**, 1698–1706 (2008).

20. Gay, N. J., Vyse, A. J., Enquselassie, F., Nigatu, W. & Nokes, D. J. Improving sensitivity of oral fluid testing in IgG prevalence studies: application of mixture models to a rubella antibody survey. *Epidemiol. Infect.* **130**, 285–291 (2003).

21. Vink, M. A., van de Kassteele, J., Wallinga, J., Teunis, P. F. M. & Bogaards, J. A. Estimating seroprevalence of human papillomavirus type 16 using a mixture model with smoothed age-dependent mixing proportions. *Epidemiol. Camb. Mass* **26**, 8–16 (2015).

22. Hitchings, M. D. T. *et al.* A Mixture Model for Estimating SARS-CoV-2 Seroprevalence in Chennai, India. *Am. J. Epidemiol.* **192**, 1552–1561 (2023).

23. Nielsen, S. S., Toft, N., Jørgensen, E. & Bibby, B. M. Bayesian mixture models for within-herd prevalence estimates of bovine paratuberculosis based on a continuous ELISA response. *Prev. Vet. Med.* **81**, 290–305 (2007).

543 24. Vyse, A. J. *et al.* Interpreting serological surveys using mixture models: the
544 seroepidemiology of measles, mumps and rubella in England and Wales at the beginning of
545 the 21st century. *Epidemiol. Infect.* **134**, 1303–1312 (2006).
546 25. Vyse, A. J., Gay, N. J., Hesketh, L. M., Morgan-Capner, P. & Miller, E. Seroprevalence
547 of antibody to varicella zoster virus in England and Wales in children and young adults.
548 *Epidemiol. Infect.* **132**, 1129–1134 (2004).
549 26. Lam, H. M. *et al.* Serological inference of past primary and secondary dengue infection:
550 implications for vaccination. *J. R. Soc. Interface* **16**, 20190207 (2019).
551 27. Biggs, J. R. *et al.* A serological framework to investigate acute primary and post-primary
552 dengue cases reporting across the Philippines. *BMC Med.* **18**, 364 (2020).
553 28. Cox, V. *et al.* Estimating dengue transmission intensity from serological data: A
554 comparative analysis using mixture and catalytic models. *PLoS Negl. Trop. Dis.* **16**,
555 e0010592 (2022).
556 29. Thongcharoen, P. *Monograph on Dengue/Dengue Haemorrhagic Fever*. (World Health
557 Organization, Regional Office for South-East Asia, New Delhi, 1993).
558 30. Lima, M. R. Q., Nunes, P. C. G. & dos Santos, F. B. Serological DiagnosisSerological
559 diagnosis of DengueDengue. in *Dengue Virus: Methods and Protocols* (ed. Mohana-Borges,
560 R.) 173–196 (Springer US, New York, NY, 2022). doi:10.1007/978-1-0716-1879-0_12.
561 31. Katz, J. M., Hancock, K. & Xu, X. Serologic assays for influenza surveillance, diagnosis
562 and vaccine evaluation. *Expert Rev. Anti Infect. Ther.* **9**, 669–683 (2011).
563 32. Cauchemez, S. *et al.* Influenza Infection Rates, Measurement Errors and the
564 Interpretation of Paired Serology. *PLoS Pathog.* **8**, e1003061 (2012).
565 33. Guzman, M. G. *et al.* Multi-Country Evaluation of the Sensitivity and Specificity of Two
566 Commercially-Available NS1 ELISA Assays for Dengue Diagnosis. *PLoS Negl. Trop. Dis.* **4**,
567 e811 (2010).
568 34. Fry, S. R. *et al.* The Diagnostic Sensitivity of Dengue Rapid Test Assays Is Significantly

569 Enhanced by Using a Combined Antigen and Antibody Testing Approach. *PLoS Negl. Trop.*
570 *Dis.* **5**, e1199 (2011).
571 35. Muller, D. A., Depelsenaire, A. C. I. & Young, P. R. Clinical and Laboratory Diagnosis of
572 Dengue Virus Infection. *J. Infect. Dis.* **215**, S89–S95 (2017).
573 36. Watthanaworawit, W. *et al.* A prospective evaluation of diagnostic methodologies for the
574 acute diagnosis of dengue virus infection on the Thailand-Myanmar border. *Trans. R. Soc.*
575 *Trop. Med. Hyg.* **105**, 32–37 (2011).
576 37. Menezes, A. *et al.* serosim: An R package for simulating serological data arising from
577 vaccination, epidemiological and antibody kinetics processes. *PLOS Comput. Biol.* **19**,
578 e1011384 (2023).
579 38. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation
580 for Statistical Computing (2023).
581 39. Endy, T. P. *et al.* Spatial and Temporal Circulation of Dengue Virus Serotypes: A
582 Prospective Study of Primary School Children in Kamphaeng Phet, Thailand. *Am. J.*
583 *Epidemiol.* **156**, 52–59 (2002).
584 40. Endy, T. P. *et al.* Epidemiology of Inapparent and Symptomatic Acute Dengue Virus
585 Infection: A Prospective Study of Primary School Children in Kamphaeng Phet, Thailand. *Am.*
586 *J. Epidemiol.* **156**, 40–51 (2002).
587 41. Clarke, D. H. & Casals, J. Techniques for hemagglutination and hemagglutination-
588 inhibition with arthropod-borne viruses. *Am. J. Trop. Med. Hyg.* **7**, 561–573 (1958).