Nature Communications

https://doi.org/10.1038/s41467-026-72859-x

Article in Press

A Bayesian modelling framework to improve antibody titer estimation applied to RSV dilution series data

Received: 14 August 2025

Accepted: 24 April 2026

Cite this article as: Wang, Y., Wang, Q., Wymant, C. *et al.* A Bayesian modelling framework to improve antibody titer estimation applied to RSV dilution series data. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-72859-x

Yan Wang, Qianli Wang, Chris Wymant, Junyi Zou, Lan Yi, Meng Xu, James A. Hay & Hongjie Yu

A Bayesian modelling framework to improve antibody titer estimation applied to RSV dilution series data

Yan Wang[1], Qianli Wang[2], Chris Wymant[3], Junyi Zou[1], Lan Yi[2], Meng Xu[1], James A. Hay[3,*], Hongjie Yu[1,2,4,*]

Affiliations:

1. School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai 200032, China
2. Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai 200032, China
3. Pandemic Sciences Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, United Kingdom
4. Department of Infectious Diseases, Huashan Hospital, Fudan University, Shanghai 200040, China.

* Correspondence: james.hay@ndm.ox.ac.uk (J.A.H.); yhj@fudan.edu.cn (H.Y.)

Abstract

Accurately measuring antibody levels is important for assessing population immunity and guiding vaccine development. We identify batch-level biases and experimental noise in neutralizing antibody (nAb) titer estimates from respiratory syncytial virus (RSV) foci reduction neutralization tests (FRNTs) when using off-the-shelf methods such as the Kärber formula and four-parameter logistic (4PL) model. To address this, we develop a Bayesian hierarchical model (BHM) to estimate nAb titers, correcting for batch effects and other sources of experimental variation. We evaluate model performance using both simulated and experimental FRNT data. In simulation, nAb titers are most accurate using the BHM (Spearman ρ=0.96 compared to simulation truth, P<0.001; root mean square error [RMSE]=0.41), outperforming the Kärber formula (ρ=0.63, P<0.001; RMSE=1.64) and 4PL model (ρ=0.87, P<0.001; RMSE=1.09). The Kärber formula produces more false negatives (9.85%) than the 4PL model (2.42%) and BHM (0.93%), and the 4PL model often produces biased titers for weakly positive samples. In experimental data, population-level measures, such as geometric mean titers (GMTs), seroprevalence and seroincidence differ substantially depending on the method. This framework can be adapted to other antibody assays producing dilution series data and improves the accuracy and robustness of titer estimates across a range of experimental settings.

Introduction

Serological data are crucial for quantifying population exposure and immunity against a range of pathogens, particularly vaccine preventable diseases where humoral immunity provides a key measure of protection.[1-4] By measuring pre-existing antibody levels induced by past infection and vaccination, public health interventions can be targeted to susceptible individuals and populations at greatest risk of disease. This approach to characterizing population immunity is particularly timely for respiratory syncytial virus (RSV) following recent advances in RSV prophylactics, including the approval of several vaccines and a half-life-extended monoclonal antibody (mAb).[5-9] The public health burden of RSV is considerable, causing acute respiratory infections in individuals of all ages, contributing to over 100,000 deaths annually among children under five and significant rates of mortality among older adults.[10,11] Thus, reliable serological evidence is needed to guide optimal vaccination strategies to reduce disease burden, particularly in children and infants.

Neutralizing antibodies (nAbs) in particular play a key role in immunity by effectively blocking viral entry into host cells and inhibiting a virus's within-host activity, and are recognized as key serological correlates of protection for many pathogens, including RSV.[12-15] Multiple neutralization tests, such as the plaque reduction neutralization test (PRNT), foci reduction neutralization test (FRNT), and pseudovirus neutralization test (PVNT), have been established to quantify antibody concentrations in serum.[12,16-19] A common design for these assays involves incubating a virus with a series of diluted serum samples and measuring the experimental outcome (e.g., the number of plaques or foci of infected cells) at each dilution. As assay readouts vary across dilutions, a quantitative relationship (i.e., a dilution curve) can be inferred for each sample, serving as the basis for estimating the nAb titer, a summary statistic of the antibody concentration in a sample. Specifically, the titer is often defined as the reciprocal of the dilution that results in 50% neutralization.

Although a neutralizing antibody titer is ideally an accurate summary measurement of antibody levels in a sample from a dilution series, differences in experimental design and the absence of standard controls can limit the comparability of data across studies.[3,20-22] International initiatives to standardize methods and reporting have improved inter-lab comparisons, but other sources of random and systematic error can still persist. For example, intra-laboratory variation arises from differences in reagent lots, cell passages, operators, and environmental conditions across experimental batches even if the protocol is unchanged. Moreover, even within a single batch, technical errors, such as pipetting errors and variation in incubation conditions, can introduce uncertainty in titer estimation. Replicate measurements of the same sample are often used to average over this technical variation, but this does not adjust for systematic biases arising between batches or laboratory testing rounds.

Despite the attention paid to experimental design, less care is given to the statistical method for calculating antibody titers from dilution series data.[20] To generate a single antibody titer estimate, a dose-response analysis is usually performed using the raw dilution series data such that the output is a single parameter or summary from a fitted model.[23] Careful statistical analysis at this stage is arguably as important for calculating estimates such as seroprevalence or geometric mean titers (GMTs) as the experimental design, but modelling specialists are often only involved after the raw data processing stage. A number of off-the-shelf methods exist for estimating antibody titers, including the Kärber formula and the four-parameter logistic (4PL) model, as well as more sophisticated, custom mathematical models.[17,23-27] However, their performance for estimating antibody titers from raw serial dilution data has not been explored, and uncertainty in the model fits are not usually propagated into downstream analyses.

This study aims to provide a practical mathematical framework that can be readily adopted by laboratories and researchers to improve the estimation of antibody titers from dilution series data. We developed a Bayesian hierarchical model (BHM), generating posterior distributions for individual antibody titers while correcting for between-lab or between-batch level effects. We applied the framework to RSV FRNT data generated from a robust assay protocol with positive serum and virus controls, demonstrating how raw data generated from a standardized procedure can be paired with a sophisticated model to improve the estimation of neutralizing antibody titers. To evaluate the performance of our framework, we conducted a simulation study comparing the proposed model with standard estimation methods. We also examined how different titer estimation approaches affect population-level serological outcomes from two seroepidemiological studies in Anhua County, Hunan Province, China, including age-stratified seroprevalence, GMTs, seroconversion rates, and fold-rise rates.

Results

Description of experimental data

This study was based on 1,777 serum samples combined from two independent studies: 485 from a cross-sectional seroepidemiological survey (participants aged 9 months to 95 years)[28] and 1,292 from a mother-neonate cohort (143 mother-neonate pairs followed for 5-8 years post-delivery)[29]. The overall age distribution of the combined samples was 0-6 months: 508 (28.6%), 7 months-5 years: 541 (30.4%), and >5 years: 728 (41.0%). See Methods and Table S1 for details. Samples were tested in 28 separate batches using RSV FRNTs (with workflow illustrated in Figure 1A and detailed in the Supplementary Information). In each batch, multiple 96-well plates were processed in parallel under the same experimental conditions, with an assay-level covariate (i.e., virus working dilution; see Methods for details) varied across batches. Each batch included three types of samples. First, a virus control (VC) prepared at a fixed dilution of working virus stock without serum. Second, serially diluted positive controls to ensure reproducibility, including an internal positive control (PC) and an International Standard control

(IS; see Methods for details). Third, serum samples from study participants with unknown antibody concentration (Figure 1B). The virus working dilution, number of replicates per assay control, as well as the number and source of samples in each batch, are summarized in Table S2. The raw dilution series data (i.e., foci count) from the FRNT assay are illustrated in Figure 1C.

The objective of the FRNT is to estimate the antibody concentration (i.e., nAb titer) in each sample. In this process, the virus control provides a measure of the background signal (i.e., the count of foci observed in the absence of serum) across different experimental conditions and batches. The positive controls, which should have the same concentration in each experimental run, provide information on the effects of between-batch variation. The true antibody concentration in each sample is therefore a latent variable of interest. However, the assay provides only noisy measurements of this latent variable. The purpose of the modelling is to obtain reliable estimates of the true antibody titer after accounting for noise and errors introduced in the measurement process.

RSV FRNT titers show considerable variation between and within batches

Using experimental data from assay controls, we observed considerable variability in RSV FRNT measurements across virus working dilutions, as well as between and within batches. Results are presented for both raw assay readouts and nAb titers (estimated using the Kärber method and the 4PL model; see Methods for details; an illustration of the 4PL model is provided in Figure S1). Specifically, the assay-level covariate, i.e., virus working dilution, directly influenced the observed foci counts, resulting in different levels of background signals, with higher virus working dilutions associated with lower foci counts in the VC. However, considerable variation in raw VC measurements was still observed across batches even when the same virus working dilution was used (Figure S2A). This between-batch variability was also observed for PC and IS. We selected two representative batches as illustrative examples: batch 20 consistently showed lower foci counts across serial dilutions, whereas batch 22 showed higher foci counts, despite using the same virus working dilution (Figure S2B-C). A commonly used method for the control of batch effects is to normalize the data by calculating foci reduction at each serial dilution relative to the mean foci count of VC within the same batch, as defined in Methods (Equation 1). However, we still observed notable differences in both the foci reduction data and the fitted 4PL curves across batches (Figure S2D-E), likely reflecting the fact that the background signal provided by VC does not capture all sources of batch-to-batch variability across series dilutions.

To further assess the impact of these batch effects on antibody titers, nAb titers were estimated for each PC replicate (2-4 replicates per batch; Table S2) using the Kärber formula and the 4PL model. Both methods summarize the relationship between the dilution factor and the reduction in foci counts to generate a single antibody titer estimate per sample. In a perfect experimental and modelling setup, the measured nAb titer should be the same across all batches, as each run uses

the same PC sample. However, the total variance of the 4PL titer estimates (on the $\log_2$ scale) was 0.45, 30.8% of which was between-batch variance and 69.2% within-batch variance. The Kärber formula resulted in even higher variability (a total variance of 0.89 on the $\log_2$ scale), with between- and within-batch variance accounting for 44.1% and 55.9%, respectively (Figure S3). Typically, nAb titers are derived from the averaged responses of multiple replicates to mitigate the impact of within-batch random errors, but the above figure refers to separate PC replicates to illustrate the potential impact of different sources of variation.

Description of the Bayesian hierarchical model

In the above analysis we observed substantial variation in both the experimental data and the titer estimates derived from the Kärber formula and 4PL model. While averaging responses across multiple replicates reduced within-batch variance, significant variability remained due to batch effects, which are often overlooked and cannot be fully addressed with standard methods such as the Kärber formula and 4PL model. To provide a method that could further account for between-batch variation, we developed a Bayesian hierarchical model (BHM) based on the 4PL model (Figure 2). This method is applicable to serological assays in which a serially diluted positive control is tested alongside serum samples across batches and it accommodates assay-level covariates by adjusting for the background signal provided by the virus control without serum. Briefly, for each set of counts, the foci reduction was first calculated relative to the modelled mean foci counts in VC for each virus working dilution. Then, we modelled the relationship between the dilution factor and reduction in foci counts using a hierarchical 4PL model, allowing for random effects both between samples and between batches on each parameter. Each sample has its own set of parameter values drawn from a population-level distribution, whereas the PC and IS serve as references, with batch-specific parameters centered on their true values. Using this hierarchical approach, we fit a 4PL curve to the dilution series data from each sample, adding both sample-level and batch-level effects. Thus, the observed data are described by a 4PL curve accounting for all of these sources of variation, whereas the true, unadjusted 4PL curve for each sample is obtained by subtracting the batch effects. The corrected antibody titer can then be calculated directly as a function of the unadjusted 4PL curve (see Methods for details; Model parameters are listed in Table S3). Notably, this method can also account for random effects arising from differences in assay protocols, equipment, reagents, or personnel, which are common in large-scale or multicenter studies. However, since the experiments in this study were carefully controlled and conducted by the same team in a single laboratory, such lab- or experimental-level corrections were not necessary.

Comparison of titer estimation methods applied to experimental data

We applied the BHM to our experimental data and compared the results with those obtained from standard methods, including the Kärber formula and the 4PL model. Visual inspection of trace plots suggested good mixing and convergence for population-, batch-, and sample-level

parameters (Figure S4), with no divergent transitions across all chains and $\hat{R}$ values close to 1 for all fitted parameters (0.999-1.005). The BHM demonstrated strong performance in fitting experimental data, with model-fitted foci reduction (posterior medians) highly correlated with the experimental observations (Spearman ρ = 0.89-0.94, P<0.001 for all), and root mean square errors (RMSEs) between 0.13 and 0.14. (Figure S5). For the VC, the model-based mean foci count closely matched the observed mean for each batch (Figure 3A). For the PC and IS, the BHM captured batch effects and yielded batch-adjusted 4PL curves for the PC and IS (Figure 3B). For serum samples, BHM-adjusted titers were strongly correlated with those estimated using the Kärber formula (ρ = 0.89, P<0.001) and the 4PL model (ρ = 0.97, P<0.001), while producing a well-constrained distribution of nAb titers at the population level (Figure S6). However, titer estimates obtained from the Kärber method showed greater discrepancies with the BHM-adjusted titers (RMSE = 0.99), whereas the 4PL model exhibited smaller discrepancies (RMSE = 0.57) but produced a higher number of negative titers than BHM (N = 503 vs. N = 423 out of 1,777 serum samples; Figure 3C).

To illustrate how the BHM corrects for batch effects, we visualized model fits for three representative batches (Figure S7). In batches 11 and 21, the unadjusted PC and IS curves were higher than the adjusted curves, leading the BHM to apply downward adjustments to samples from these batches. In contrast, for batch 19, the unadjusted curves were lower, resulting in upward adjustments. The adjusted titers for serum samples were subsequently estimated based on these batch-adjusted curves at 50% foci reduction. In most cases, the BHM-unadjusted fits were similar to those from the standard 4PL model; however, the latter sometimes failed to capture the underlying curve patterns, leading to negative titer estimates, as illustrated by samples 826 and 907 in Figure S7B.

Comparison of titer estimation methods applied to simulated data

Despite the BHM performing well in fitting the experimental data, the accuracy of the titer estimates could not be directly assessed because the true titers were unknown. We therefore conducted a simulation study to evaluate model performance where the true parameters for each sample were known. Mirroring the structure of the experimental data, we generated a simulated dataset with four levels of working virus dilution, seven batches per level, and 20 samples per batch, yielding a total of 28 batches and 560 serum samples; each batch included VC, PC, and IS. We simulated moderate batch effects, setting the standard deviations of the batch effects for each parameter as $[\sigma_{\gamma_b}, \sigma_{\gamma_c}, \sigma_{\gamma_d}, \sigma_{\gamma_e}, \sigma_\varsigma]' = [0.028, 0.211, 0.009, 0.250, 3.670]'$ (see Methods and Table S4 for detailed simulation settings). The simulated dataset successfully reproduced the key features of the experimental data, including the heterogeneity across virus working dilutions, batches and serum samples (Figure 4).

Based on the simulated VC data, we obtained reasonable estimates for the parameters of each virus working dilution (Figure 4A and Figure S8). Using these estimates, we transformed the

simulated data into foci reduction to fit the BHM. Trace plots showed good mixing and convergence (Figure S9), with no divergent transitions and all $\hat{R}$ values close to 1 (0.999-1.008). Posterior distributions were centered around the true parameter values (Figures S10-S12). Empirical coverage of the 95% credible intervals (CrIs) ranged from 0.82 to 1 for batch-level parameters and from 0.90 to 1 for sample-level parameters (Figure S13). The strong correlation between true and fitted foci reduction values (Spearman ρ=0.99-1, P<0.001 for all; Figure S14) indicated close recovery of the true curves for both serum samples and assay controls. In Figure 4B-C, we present the true and fitted PC curves from two representative batches, along with the curves from three simulated serum samples drawn from these batches. Batch 1 showed higher levels of foci reduction in the PC, whereas batch 2 showed lower levels. Consequently, the Kärber formula and 4PL model overestimated titers in batch 1 and underestimated titers in batch 2, while the BHM model consistently provided batch-adjusted estimates closest to the true titers.

Next, we compared the estimated nAb titers against the true values across all simulated samples (Figure 5). Among the methods evaluated, the Kärber formula exhibited the greatest uncertainty (ρ = 0.63 compared to simulation truth, P<0.001; RMSE = 1.64) and led to more false negatives (9.85%) than the 4PL model (2.42%) and the BHM (0.93%). Although the 4PL model and BHM without batch adjustment showed improved accuracy (ρ = 0.87 and 0.84 compared to simulation truth, P<0.001 for both; RMSE = 1.09 and 1.12, respectively), both were susceptible to biased titer estimates in the presence of large batch effects (e.g., Sample 1 in Figure 4C). In addition, the 4PL model tended to produce false negatives for samples with low antibody levels (e.g., Samples 2 in Figure 4C). Notably, among the 13 false negatives generated by the 4PL model, 38.5% (Samples 4, 6, 8, 10, and 11 in Figure S15) were likely due to inappropriate curve-fitting patterns. When such weakly positive samples make up a large proportion of the dataset, these errors may bias the overall titer distribution. In contrast, the BHM-adjusted estimates achieved the highest accuracy in recovering true titers (Spearman ρ = 0.96 compared to simulation truth, P<0.001; RMSE = 0.41), with a coverage probability of 95% CrIs to be 0.97.

Impact on seroepidemiology

Finally, we performed seroepidemiological analysis using different titer estimates from the experimental data. We observed notable differences across methods in the estimates of GMTs and seropositivity rates. These discrepancies were particularly evident for the Kärber estimates among children aged 0-5 years, a high-risk group for severe RSV infection. For example, in infants aged 0-6 months, the GMTs estimated using the Kärber formula, 4PL model, and BHM-adjusted method were 6.33 (95% confidence interval [CI]: 6.17-6.49), 5.66 (95% CI: 5.48-5.84), and 5.87 (95% CI: 5.70-6.04), respectively. The corresponding seroprevalence estimates were 66.73% (95% CI: 62.52%-70.69%), 48.23% (95% CI: 43.91%-52.57%), and 54.92% (95% CI: 50.57%-59.20%), respectively (Figure 6A-B).

Seroconversion or fold rises in nAb titers of paired serum samples are commonly used outcomes to identify new serological infections, or to assess the immunogenicity of vaccines. Based on the 715 paired serum samples from the cohort study, BHM-adjusted method detected 114 seroconversions, 71 two-fold titer rises and 39 four-fold titer rises. Assuming the BHM-adjusted titers reflected the true serostatus, we found that both the Kärber formula and the 4PL model showed reduced sensitivity in detecting seroconversion (68.4% and 85.1%), two-fold titer rises (64.8% and 60.6%), and four-fold titer rises (61.5% and 59.0%), despite demonstrating high specificity (>90%) across all outcomes (Figure 6C). Given the community-based study design, true seroconversions and fold rises are likely to be infrequent, and thus the choice of method yielded only small differences in seroresponse rates (e.g., four-fold rise: 3.5% [25/715] by 4PL vs. 5.5% [39/715] by BHM-adjusted method; Figure S16). However, within the same 715 paired samples, BHM-adjusted method identified 39 four-fold rises compared with 25 by 4PL model, representing a 56% relative increase in detected seroresponses.

Discussion

In this study we developed a Bayesian hierarchical modelling framework based on FRNT, one of the most widely used serological assays for quantifying neutralizing antibody levels.[12] Our primary objective was to correct for batch effects, which were found to be an influential source of variability in our experimental data but have often been overlooked in serological studies. However, the model can be adapted with minor modifications for lab- and batch-effect correction in any similar assay that produces serial dilution curves, enhancing the accuracy of serological testing. Overall, our method performed well in capturing the patterns of experimental data. Compared with previous approaches, the proposed model produced unbiased and more accurate titer estimates in the simulation study. It could offer more reliable evidence to inform vaccine development and public health policymaking.

Batch effects are a common source of technical variation in biological experiments, arising from differences in assay protocols, equipment, reagents, operators, or experimental conditions. Previous studies have attempted to correct for these effects using Bayesian or empirical Bayes models, particularly in genomic and epigenomic experiments, where methods such as ComBat and its extensions have been widely applied to adjust for both known and latent batch effects.[30-33] However, to our knowledge, no method has been specifically developed for correcting batch effects in serological dilution series data. Although mathematical models have been developed to fit dilution curves in a hierarchical framework, they do not incorporate batch effects into the hierarchical structure.[26,27]

Both the Kärber formula and 4PL model are simple and widely used methods for estimating nAb titers from raw count data in assays such as PRNT or FRNT.[17,23-25] However, we found that the Kärber formula can lead to high variability in individual titer estimates and result in more false

negatives in the simulated dataset. Although the true titer distribution for samples in our experimental data was unknown, the Kärber formula yielded higher GMTs and seroprevalence (in infants and young children) than other methods, potentially reflecting bias in titer estimates. We also found that the 4PL model, despite providing a flexible, model-based approach for estimating nAb titers, was suboptimal in low-titer samples, where the estimated curve often deviated substantially from the true underlying curve, leading to inaccurate titer estimates. This bias may arise because it fits each curve solely based on the data from an individual serum sample, whereas the hierarchical structure of the proposed Bayesian model provides more reasonable fits by borrowing strength across samples through partial pooling. In addition, our simulation study showed that large batch effects may cause the fitted 4PL curve to fail to reach the 50% neutralization threshold, resulting in biased titer estimates. This, in turn, could lead to biased population-level parameter estimates. However, in our experimental data, GMT and seroprevalence estimates based on the 4PL model were only slightly lower than those obtained using the BHM-adjusted titers. This is likely because all FRNT assays in this study were conducted under standardized procedures, where batch effects, as well as other sources of technical variation, were highly controlled. In more complex study settings, such as multicenter seroepidemiological studies or vaccine trials, systematic errors introduced by the 4PL model may have a greater impact on population-level parameter estimates.

It should be noted that, for "true negative" samples, foci counts do not change across serial dilutions, resulting in a flat response in the experimental data. In contrast, in our simulation, 4PL parameters are drawn from a population distribution to generate 4PL curves for each sample and thus there are no simulated true negatives. The negative samples in the simulation represent only a small subset of samples with very low antibody titers below the positivity threshold (i.e., 1:40; N = 22 out of 560). Among these 22 negative samples, the Kärber formula frequently produced positive titers ($\geq$5.32; i.e., $\log_2(40)$; N = 11) that deviated substantially from the true values (GMT=5.51; RMSE = 2.00). For the 4PL model and BHM, although some samples were classified as "clear negatives" and excluded from curve fitting (see Methods for details), both models still produced positive titers for some samples (N = 9 and 10 out of 22, respectively), with the 4PL model showing greater deviations (GMT = 4.86; RMSE = 1.05) compared with BHM (GMT = 5.07; RMSE = 0.63; Figure 5).

Moreover, correcting for batch effects is critical for accurately estimating population-level seroincidence and for evaluating vaccine immunogenicity. These outcomes are typically inferred from paired serum samples collected from the same individuals (e.g., acute and convalescent phases, pre- and post-season, or pre- and post-vaccination), based on evidence of seroconversion or fold changes in antibody titers. In our experimental data, which included 715 paired serum samples, we observed substantial discrepancies in individual serostatus depending on the method used. For example, the BHM detected a 56% increase in four-fold titer rises compared with the 4PL model (39 vs. 25). Although these discrepancies yielded only minor differences in seroresponse rates in our community-based study cohort (5.5% [39/715] vs. 3.5% [25/715]), they

could be amplified in settings with higher transmission intensity or vaccination coverage, potentially leading to underestimation of seroincidence and vaccine effectiveness.

Despite the advancements of the proposed model, we recommend minimizing between-batch variation at the experimental design stage. Controlling for the measurement errors through a standard operating procedure (SOP) remains the primary choice, as this reduces the need for complex models. Given that a standardized, universally accepted SOP for RSV FRNT assays is still lacking, we adapted the assay based on an established protocol.[34] Multiple pilot tests were performed to optimize assay procedures and ensure stability and reproducibility. The detailed workflow provided in the Supplementary Information may serve as a reference for future SOP development. Control samples, as emphasized in the World Health Organization (WHO) Guidelines[35], are crucial for standardizing measurements across assay runs and also serve as the basis for correcting batch effects in our Bayesian hierarchical modelling framework. In this study, random batch effects were assumed across all batches, which is reasonable given the implementation of standardized procedures: all assays were performed by the same group of trained personnel in a single laboratory, using reagents from the same manufacturer, thereby minimizing technical variation across batches. As a result, BHM-unadjusted estimates showed only minor differences from the adjusted estimates in the experimental data (Figure 3C), likely reflecting the consistency afforded by standardized assay procedures. However, in a simulation study with amplified batch effects, BHM-unadjusted estimates deviated more pronouncedly from the true titers (Figure 5C), highlighting the importance of batch effect correction under less controlled conditions.

Our experimental setup had one assay-level covariate (i.e., virus working dilution), where concentrations ranging from 1:200 to 1:400 were used across batches. We accounted for this systematic variation in the Bayesian hierarchical model (see Methods for details), though this step would not be necessary if the same working virus stock was used across all FRNT batches. Furthermore, unavoidable within-batch variability, arising from sources such as pipetting errors and random measurement noise, can contribute substantially to uncertainty in the estimated titers. In this study, substantial variability was observed in repeated measurements of assay controls within the same experimental run (Figure S2-S3), emphasizing the importance of incorporating replicates to control for within-batch variability in the modelling stage.

Our study has limitations. First, the readouts for negative samples are theoretically unchanged across serial dilutions. As a result, these samples do not fit the 4PL model and were excluded from the analysis prior to fitting the BHM. Specifically, we defined 'clear negatives' as samples with a foci reduction of less than 30% at the starting dilution, allowing a 20% margin to account for measurement errors (see Methods for details). Although using a mixture model with positives and negatives, and assuming bimodal parameter distributions within the BHM could potentially address this issue[36], we chose not to implement this approach in order to reduce model complexity. Second, the four parameters of the 4PL model may be correlated. In the BHM, we

did not account for potential correlations among these parameters to reduce model complexity. Third, we assumed that batch effects were consistent between serum samples and assay controls. However, minor discrepancies may exist, as assay controls were all placed on the last plates of each batch and may have been exposed to longer fixation and mixing times during the experimental procedure. Furthermore, due to the absence of control sera on each plate, the model cannot adjust for between-plate variability. Lastly, the BHM was applied exclusively to data generated in a single laboratory. Its robustness and generalizability require further validation using data from additional sources.

Accurate antibody titer estimates are essential for understanding the full infection burden of pathogens which are missed through routine syndromic surveillance, and for guiding effective public health interventions. In clinical trials, a significant increase in nAbs is a primary outcome for assessing vaccines or mAbs. For public health policy makers, testing for nAbs is important to determine seroprevalence, evaluate population immunity, assess real-world vaccine effectiveness and design optimal immunization strategies. Our model demonstrates improved nAb titer estimation accuracy over standard methods in both simulation studies and real-world data, and also provides antibody titer estimates as posterior distributions with uncertainty rather than single point estimates. This framework can be readily adapted to other serial dilution assays and has the potential to enhance the accuracy and precision of titer estimation across diverse experimental settings. Future studies should consider the importance of standardizing both experimental design and choice of mathematical model when generating serological datasets from serial dilution assays.

Methods

Ethics statement

This study was based on two community-based seroepidemiological studies independently conducted in Anhua County, Hunan Province, China: a mother-neonate cohort study and a cross-sectional survey, both of which have been published previously.[28,37] The mother-neonate cohort study was approved by the Institutional Review Board (IRB) of the WHO Western Pacific Regional Office (2013.10.CHN.2.ESR), the Chinese Centre for Disease Control and Prevention (201224), the London School of Hygiene & Tropical Medicine (15698) and School of Public Health, Fudan University (IRB#2019-15-0756 and #2022-02-0947). The cross-sectional survey was approved by the IRB of the School of Public Health, Fudan University (IRB#2020-11-0857, #2020-11-0857-S and #2022-02-0948). In both studies, written informed consent was obtained at each time of sample collection. Consent for neonates was provided by their mothers, and consent for participants under 7 years of age was provided by their legal guardians. Participants aged between 7 and 11 years signed simplified forms, with their guardians signing the full versions.

Participants older than 12 years, together with their guardians, signed the full version of the consent form.

Study participants and serum samples

Participants in the mother-neonate cohort were recruited from local hospitals between 2013 and 2015. Peripartum venous blood samples were collected from mothers at delivery, and cord blood samples were obtained from neonates at birth. Mothers were followed up once at 5-8 years postpartum, whereas neonates were followed at 2, 4, 6, 12, 24, and 36 months of age and again at 5-8 years. From the original cohort, 143 mother-neonate pairs who had completed at least six follow-up visits were randomly selected for neutralization tests. The cross-sectional serological survey was conducted between July and November 2021 and enrolled community participants across all age groups. Of these, 485 participants were randomly selected for neutralization tests.

Neutralization assays

RSV neutralizing antibody titers were measured using a foci reduction neutralization test (FRNT), established based on previously published protocols.[34] An overview of the assay workflow is shown in Figure 1A. Briefly, Vero cells (CCL-81, American Type Culture Collection, ATCC) were seeded at a density of 2×10^5 cells per well in 96-well plates 16 hours prior to infection. Heat-inactivated serum samples were serially diluted 3-fold in Minimum Essential Medium (MEM), ranging from 1:20 to 1:43,740. These dilutions were then mixed with an equal volume of the working solution of RSV A2 strain (ATCC), yielding final serum dilutions from 1:40 to 1:87,480 in a total volume of 240 μL per well. After incubation for 2 hours at 4 °C, the mixtures were added to the cell plates in duplicate (100 μL per well). Following a 1.5-hour incubation at 37°C in 5% CO_2, the inoculum was removed, and each well was overlaid with 2% carboxymethylcellulose (low viscosity, Sigma-Aldrich, St. Louis, MO) overlay medium consisting of cell culture medium supplemented with 2% fetal bovine serum (10091148, Thermo Fisher Scientific, USA) at 100 μL per well, then incubated at 37 °C for 48 hours.

The CMC overlay medium was then removed and cell monolayers were fixed with cold methanol at 4 °C for 20 minutes. Endogenous peroxidase activity was blocked for 10 minutes (P0100B; Beyotime Institute of Biotechnology, China), followed by permeabilization and blocking using 0.3% Triton X-100 and 2% goat serum in phosphate-buffered saline (PBS) for 30 minutes at room temperature. Viral foci were detected using a rabbit polyclonal antibody targeting the RSV fusion (F) glycoprotein (11049-T46, Sino Biological, China) at a dilution of 1:500, followed by an HRP-conjugated goat anti-rabbit IgG secondary antibody (A0208, Beyotime, China) at a dilution of 1:800. After fixation and immunostaining, the plates were scanned. The number of foci was counted using the image analyzer (EliSpot Reader, AID Diagnostika GmbH, Germany). The resulting data were recorded in Microsoft Excel (Microsoft

Corporation, USA). All FRNT experiments were performed in an approved Biosafety Level 2 (BSL-2) setting.

The assay was performed in 28 batches, defined as experimental runs comprising multiple 96-well plates processed in parallel under identical assay conditions. The virus working dilution (i.e., the concentration of the working virus stock) directly influences foci counts in the FRNT assay. In the first 15 batches, we tested virus working dilutions ranging from 1:200 to 1:400 (a range based on preliminary experiments) to determine the optimal condition. From these, a dilution of 1:330 was selected and used for the remaining batches, as it consistently produced countable and reproducible foci (approximately 30-55 foci per well), minimized focus overlap, and maintained the assay's dynamic range[38]. Detailed virus working dilutions for each batch are provided in Table S2.

The layout of serum samples and assay controls on each plate within each batch is illustrated in Figure 1B. Specifically, six serum samples were tested per plate, with each sample assayed in duplicate to minimize random variability. All assay controls were placed on the last plate of each batch, including a virus control without serum (VC), an internal positive control (PC) and an International Standard control (IS). The VC was prepared at a fixed dilution for each batch, whereas both the PC and IS were subjected to the same serial dilution scheme as the test samples. The PC consisted of pooled positive sera from hospitalized RSV-infected patients and was used consistently across all batches. The IS used in this study was the first WHO International Standard for Antiserum to RSV (NIBSC code: 16/284)[39], which was tested at concentrations of 500 and 1000 International Units (IU; denoted as IS500 and IS1000, respectively). This experimental design was implemented for between-batch quality control and to validate the assay's ability to detect samples with varying antibody concentrations. The raw dilution series data (i.e., foci count) from the FRNT assay are illustrated in Figure 1C using a representative example.

Standard methods for estimating nAb titers from FRNT

Raw FRNT dilution series data were converted to proportions of foci reduction for analysis, which are generally defined as

$$y_i^{\text{sam.}j} = 1 - \frac{C_{i|k(l)}^{\text{sam.}j}}{\bar{C}_{k(l)}^{\text{VC}}}. \quad (1)$$

Here, i indexes individual observations, j indexes the serum samples, k indexes batches and l indexes the levels of virus working dilutions. $\bar{C}_{k(l)}^{\text{VC}}$ denotes the mean foci count observed in the VC wells from batch k at virus working dilution l. $C_{i|k(l)}^{\text{sam.}j}$ and $y_i^{\text{sam.}j}$ represent the foci count and corresponding proportion of foci reduction for the i^{th} observation of sample j, respectively.

Neutralizing antibody titers were then defined as the reciprocal of the dilution that results in 50% foci reduction. In our FRNT assay design, the lower limit of quantification (LLOQ) was defined as one dilution step below the starting dilution (i.e., 1:40/3), and the upper limit of quantification (ULOQ) was defined as the highest dilution tested (1:87,480). Titers falling below or above these limits were censored at the LLOQ or ULOQ, respectively.

Standard methods commonly used for estimating nAb titers include the Kärber formula and the 4PL model.[17,23-25] These approaches typically estimate titers using a simple formula or by independently fitting a single dose-response model to each sample's individual dilution series. The Kärber formula does not require any data smoothing or curve fitting, making it one of the simplest and most commonly used approaches for calculating FRNT-based antibody titers. The nAb titer for each sample was calculated as:

$$\log_{10}\left(\text{titer}_{j.\text{Kärber}}\right) = \log_{10}(m) - \log_{10}(\Delta)\left(\sum_{x}(1 - \bar{y}_{x}^{\text{sam}.j}) - 0.5\right) \tag{2}$$

where m is the highest dilution, Δ is the constant interval between serial dilutions, $\bar{y}_{x}^{\text{sam}.j}$ represents the mean foci reduction across all replicates of sample j at dilution x.

The 4PL model is also a common way to fit experimental data generated by FRNT and thereby estimate nAb titers. In this approach, as shown in Figure S1, the proportion of foci reduction ($y_i^{\text{sam}.j}$) at dilution $x_i^{\text{sam}.j}$ is modeled using a 4PL function $f(\cdot)$, specified as:

$$f(x_i^{\text{sam}.j}, \boldsymbol{\theta}_{\mathbf{j}}) = c_j + \frac{d_j - c_j}{1 + e^{b_j\left(\log\left(x_i^{\text{sam}.j}\right) - \hat{e}_j\right)}} \tag{3}$$

where $\boldsymbol{\theta}_{\mathbf{j}} = \left[b_j, c_j, d_j, \hat{e}_j\right]'$ is the vector of sample-specific parameters. Specifically, b_j represents the slope of the sigmoidal curve; c_j and d_j denote the lower and upper asymptotes, respectively; $\hat{e}_j$ denotes the log-transformed midrange dilution. The nAb titer, derived from the 4PL curve at 50% foci reduction, can be expressed as a function of the fitted parameters, such that $\log\left(\text{titer}_{j.4\text{PL}}\right) = g(\boldsymbol{\theta}_{\mathbf{j}})$, where the function $g(\cdot)$ is given by

$$g(\boldsymbol{\theta}_{\mathbf{j}}) = \frac{\log\left(\frac{d_j - c_j}{0.5 - c_j} - 1\right)}{b_j} + \hat{e}_j. \tag{4}$$

It should be noted that "clear negative" samples, defined as those exhibiting less than 30% foci reduction at the starting dilution (allowing a 20% margin to account for potential measurement error due to intra- or inter-batch variation), were excluded from curve fitting and assigned a titer equal to the LLOQ. This is reasonable because foci reduction decreases monotonically with

dilution, indicating that the true response curves for such samples are highly unlikely to reach the 50% foci reduction threshold. Similarly, samples with fitted curves that failed to achieve 50% foci reduction were considered seronegative, and their titers were imputed as the LLOQ.

Bayesian hierarchical modelling framework

In this study, we developed a Bayesian hierarchical modelling framework based on the 4PL model to jointly account for assay-level fixed effects (i.e., virus working dilutions) and batch-level random effects in RSV FRNT assays. The overall model structure is illustrated in Figure 2. Prior distributions of the model parameters are summarized in Table S3. The model is comprised of three levels:

(1) Estimating assay-level fixed effects

First, we model the count of foci observed in each VC well, denoted by $C_{i|k(l)}^{VC}$, as Poisson distributed:

$$C_{i|k(l)}^{\mathrm{VC}} \sim \mathrm{Poisson}\left(\mu_{k(l)}^{\mathrm{VC}}\right) \tag{5}$$

where $\mu_{k(l)}^{\mathrm{VC}}$ indicates the mean count of foci for VC wells from batch k at virus working dilution l. We model heterogeneity in $\mu_{k(l)}^{\mathrm{VC}}$ as arising from two sources: a fixed assay-level effect for the virus working dilution, denoted by δ_l, and a random effect for the batch, denoted by ς_k. We build a Bayesian hierarchical model to provide robust estimates of the mean foci count for each virus working dilution (δ_l). The model is specified as:

$$\mu_{k(l)}^{\mathrm{VC}} = \delta_l + \varsigma_k \tag{6}$$

We calculate the foci reduction as the foci count at each serial dilution relative to the mean foci count of VC for the corresponding virus working dilution (i.e., the estimated parameter δ_l). This definition preserves the same interpretation of foci reduction as in Equation 1, while replacing the batch-specific VC mean with a virus working dilution-specific baseline, given as follows:

$$y_{i|k}^{T} = 1 - \frac{C_{i|k(l)}^{T}}{\delta_l} \tag{7}$$

where T indicates the type of sample, i.e., $T = \mathrm{PC}, \mathrm{IS500}, \mathrm{IS1000}$ or $\mathrm{sam.}\,j$.

(2) Estimating batch-level random effects

Next, the model estimates the batch-specific effects based on the serially diluted data from the PC and IS. The foci reduction defined in Equation 7 is modelled as normally distributed:

$$y_{i|k}^{T} \sim N\left(\mu_{i|k}^{T}, \sigma_y\right) \tag{8}$$

$$\mu_{i|k}^{T} = f(x_{i|k}^{T}, \boldsymbol{\theta}_{\mathbf{k}}^{\mathbf{T}}) \tag{9}$$

where T indicates the type of control samples, i.e., $T =$ PC, IS500 or IS1000. $f(\cdot)$ denotes the 4PL function as specified in Equation 3. $\boldsymbol{\theta}_{\mathbf{k}}^{\mathbf{T}} = [b_k^T, c_k^T, d_k^T, \hat{e}_k^T]'$ represents the set of unadjusted parameters for batch k. We model batch effects as having an additive effect on $\boldsymbol{\theta}_{\mathbf{k}}^{\mathbf{T}}$, such that:

$$\boldsymbol{\theta}_{\mathbf{k}}^{\mathbf{T}} = \boldsymbol{\theta}^{\mathbf{T}} + \boldsymbol{\gamma}_{\mathbf{k}} \tag{10}$$

where $\boldsymbol{\theta}^{\mathbf{T}} = [b^T, c^T, d^T, \hat{e}^T]'$ denotes the mean parameters for the control sera across batches, and $\boldsymbol{\gamma}_{\mathbf{k}} = [\gamma_{b,k}, \gamma_{c,k}, \gamma_{d,k}, \gamma_{\hat{e},k}]'$ denotes the random effects specific to batch k.

Adjusted nAb titers for the PC and IS are then estimated based on the mean parameters $\boldsymbol{\theta}^{\mathbf{T}}$, such that $\log(\text{titer}_T) = g(\boldsymbol{\theta}^{\mathbf{T}})$, as specified in Equation 4. In this study, the nAb titer of the IS1000 was used to define a conversion factor $\varphi = \text{titer}_{\text{IS1000}}/2000$, which can further be applied to convert individual nAb titers obtained from FRNT assays to WHO International Units per milliliter (IU/mL), thereby enabling comparison of our results with those from other studies.[40]

(3) Estimating adjusted nAb titers for serum samples

Finally, we model the observed foci reduction data for each serum sample j ($y_{i|k}^{\text{sam.}j}$) as:

$$y_{i|k}^{\text{sam.}j} \sim N\left(\mu_{i|k}^{\text{sam.}j}, \sigma_y\right) \tag{11}$$

$$\mu_{i|k}^{\text{sam.}j} = f(x_{i|k}^{\text{sam.}j}, \boldsymbol{\theta}_{\mathbf{j|k}}^{*}) \tag{12}$$

where $\boldsymbol{\theta}_{\mathbf{j|k}}^{*} = [b_{j|k}^{*}, c_{j|k}^{*}, d_{j|k}^{*}, \hat{e}_{j|k}^{*}]'$ represents the set of unadjusted parameters for sample j from batch k. We model each parameter as independently normally distributed, centered at the population-level mean parameters, denoted by $\boldsymbol{\theta}^{\mathbf{pop}} = [b^{\text{pop}}, c^{\text{pop}}, d^{\text{pop}}, \hat{e}^{\text{pop}}]'$.

To account for both sample-level and batch-level heterogeneity embedded in $\theta_{j|k}^{*}$, we derive the adjusted sample-specific parameters $\boldsymbol{\theta}_{\mathbf{j}} = [b_j, c_j, d_j, \hat{e}_j]'$ by removing the estimated batch effects $\boldsymbol{\gamma}_{\mathbf{k}}$ (estimated from PC and IS) from $\boldsymbol{\theta}_{\mathbf{j|k}}^{*}$:

$$\boldsymbol{\theta}_{\mathbf{j}} = \boldsymbol{\theta}_{\mathbf{j|k}}^{*} - \boldsymbol{\gamma}_{\mathbf{k}} \tag{13}$$

The adjusted nAb titers are then calculated using the batch-corrected parameters as $\log(\text{titer}_{j.\text{BHM.adj}}) = g(\boldsymbol{\theta}_{\mathbf{j}})$. For comparison, we also calculate the unadjusted titers based on the unadjusted parameters, $\log(\text{titer}_{j.\text{BHM.unadj}}) = g(\boldsymbol{\theta}_{\mathbf{j|k}}^{*})$, where $g(\cdot)$ denotes the function used to

calculate nAb titers, as specified in Equation 4. Notably, although BHM generates a posterior distribution of the nAb titer for each sample, we used the posterior median as a point titer estimate.

Bayesian inference was performed using four parallel Markov chains across four computational cores. Each chain was run for 6,000 iterations, including 3,000 warm-up iterations for algorithm adaptation and 3,000 post-warm-up iterations retained for posterior sampling. We set the target acceptance probability to 0.95 to improve the stability of the Hamiltonian Monte Carlo (HMC) algorithm and reduce the likelihood of divergent transitions. We employed non-centered parameterizations for all population-level, batch-level, and sample-level parameters to improve sampling efficiency.

Statistical simulation

We conducted a simulation study with a total of 560 simulated samples tested in 28 batches to recover the structure of RSV FRNT experimental data and assess the model performance. Specifically, to match the real data, we simulated four levels of virus working dilution, with the mean foci count in the virus control to be $\boldsymbol{\delta} = [\delta_1, \delta_2, \delta_3, \delta_4]'$. For each level of virus working dilution, seven experimental batches were simulated. In each batch, 20 serum samples, an internal positive control, an International Standard control (at different concentrations, i.e., IS500 and IS1000), and a virus control were included. Each serum sample, along with the PC and IS, was tested in duplicate across 3-fold serial dilutions ranging from 1:40 to 1: 87,480. For the VC, 22 replicates were simulated per batch.

The 4PL parameters for PC, IS500 and IS1000 (denoted by $\boldsymbol{\theta}^{\mathbf{PC}}$, $\boldsymbol{\theta}^{\mathbf{IS500}}$and $\boldsymbol{\theta}^{\mathbf{IS1000}}$, respectively), as well as the population-level mean parameters for serum samples (denoted by $\boldsymbol{\theta}^{\mathbf{pop}}$) were set in the simulation to match the posterior median estimates from the real data. For each batch k, the batch-level random effects on the 4PL parameters ($\gamma_{b,k}, \gamma_{c,k}, \gamma_{d,k}$and $\gamma_{\hat{e},k}$) and the effect on VC foci counts (ς_k) were combined into a vector $\boldsymbol{\Gamma}_{\mathbf{k}} = [\gamma_{b,k}, \gamma_{c,k}, \gamma_{d,k}, \gamma_{\hat{e},k}, \varsigma_k]'$. Each $\boldsymbol{\Gamma}_{\mathbf{k}}$ was independently drawn from a multivariate normal distribution with correlation matrix $\mathbf{R}_{\boldsymbol{\Gamma}}$. Similarly, for each sample j, sample-level random effects ($\lambda_{b,j}, \lambda_{c,j}, \lambda_{d,j}\ and\ \lambda_{\hat{e},j}$), collected into a vector $\boldsymbol{\lambda}_{\mathbf{j}} = [\lambda_{b,j}, \lambda_{c,j}, \lambda_{d,j}, \lambda_{\hat{e},j}]'$ were also drawn from a multivariate normal distribution with correlation matrix $\mathbf{R}_{\boldsymbol{\lambda}}$. All parameter settings for simulation are listed in Table S4.

The observed foci count for each well was drawn from a Poisson distribution:

$$C_{i|k(l)}^{T} \sim \text{Poisson}\left(\mu_{i|k(l)}^{T}\right) \tag{14}$$

where T indicates the type of sample.

For each VC well (i.e., $T = \text{VC}$),

$$\mu_{i|k(l)}^{T} = \delta_l + \varsigma_k \tag{15}$$

For each PC or IS well (i.e., $T = \text{PC, IS500 or IS1000}$) at a given dilution x,

$$\mu_{i|k(l)}^{T} = \delta_l\left(1 - f\left(x_{i|k(l)}^{T}, (\boldsymbol{\theta}^{\mathbf{T}} + \boldsymbol{\gamma}_{\mathbf{k}})\right)\right) \tag{16}$$

where $f(\cdot)$ denotes the 4PL function as specified in Equation 3, and $\boldsymbol{\gamma}_{\mathbf{k}} = [\gamma_{b,k}, \gamma_{c,k}, \gamma_{d,k}, \gamma_{\hat{e},k}]'$ represents the batch effect on the 4PL parameters.

For each well of serum sample j (i.e., $T = \text{sam.}\, j$) at a given dilution x:

$$\mu_{i|k(l)}^{T} = \delta_l\left(1 - f\left(x_{i|k(l)}^{T}, (\boldsymbol{\theta}^{\mathbf{pop}} + \boldsymbol{\lambda}_{\mathbf{j}} + \boldsymbol{\gamma}_{\mathbf{k}})\right)\right) \tag{17}$$

The true 4PL parameters for sample j were defined as $\boldsymbol{\theta}_{\mathbf{j}} = \boldsymbol{\theta}^{\mathbf{pop}} + \boldsymbol{\lambda}_{\mathbf{j}}$, which leads to the true nAb titer $\log\left(\text{titer}_{j.\text{true}}\right) = g(\boldsymbol{\theta}_{\mathbf{j}})$.

Evaluating model performance

Model performance was evaluated using both real experimental data and simulated data. For the real data, where the true parameter values are unknown, model performance was assessed by evaluating goodness of fit, quantified using the Spearman correlation coefficient (ρ) and the root mean squared error (RMSE) between the observed foci reduction data and the model-fitted values (posterior medians). Model convergence was also evaluated using trace plots and checking that the Potential Scale Reduction Factor ($\hat{R}$) for each parameter was less than 1.01.

The simulation study enabled evaluation of model performance under controlled conditions with known true parameter values. We assessed the consistency between the true and model-fitted foci reduction values at each serial dilution to evaluate whether the model could accurately capture the true curves. We examined the posterior distributions of the model parameters and assessed the model's ability to recover the true model parameters using bias, RMSE, Spearman ρ, and the coverage probability (the proportion of simulations in which the true value falls within the corresponding 95% posterior credible interval). Model accuracy in reproducing the true antibody titers was further evaluated by comparing the estimated nAb titers (posterior medians) with the true simulated titers using Spearman ρ, RMSE (on the $\log_2$scale), and coverage probability.

Comparing different methods

We used a linear model to characterize variation in RSV FRNT nAb titers estimated using the Kärber formula and the standard 4PL model. Spearman correlation coefficients were used to assess the agreement between titer estimates from different methods. We calculated age-specific

geometric mean titers (GMTs) and seroprevalence, along with their 95% confidence intervals (CIs), to assess the impact of different titer estimation methods on population-level serological outcomes. Seropositivity was defined as a titer ≥40, corresponding to the starting dilution. Using paired serum samples from two consecutive follow-up visits, we assessed agreement among titer estimation methods in detecting seroconversion and two-fold and four-fold titer rises. Sensitivity, specificity, and overall agreement were calculated using BHM-adjusted estimates as the reference standard.

Statistics and Reproducibility

Technical replicates were used to verify the reproducibility of the experiments. Assay controls (including VC, PC, and IS) were tested across all batches, with ≥2 replicates per batch. All samples were assayed in duplicate to control for random error. No statistical method was used to predetermine sample size. The sample size for the experimental data (1,777 samples from 28 batches) was sufficient to detect batch-level and sample-level effects. A simulation study assuming 560 samples tested in 28 batches was conducted to assess model performance. No data were excluded from the analyses. All analyses were performed in R version 4.4.0 (R Foundation for Statistical Computing, Vienna, Austria, https://www.r-project.org/). The Bayesian hierarchical model was implemented using the rstan package (version 2.32.6).[41] The 4PL model was performed using the drc package (version 3.0-1).[23,42]

Data availability

Due to privacy and ethical reasons, raw experimental data from study participants cannot be made public, but they are available from the corresponding author (H.Y.) upon reasonable request. We have provided raw experimental data for the assay controls, aggregated de-identified data for study participants, as well as simulated datasets that replicate the structure of the experimental data in this study. These datasets have been deposited in Zenodo at https://zenodo.org/records/19445779[43]. Source data are provided with this paper.

Code availability

The code used for modelling, simulation, and reproducing the main results in this study have been deposited in Zenodo at https://zenodo.org/records/19445779[43]. A detailed user guide to facilitate implementation of the proposed model is provided at: https://kristywang.github.io/BHM_Titer_Correction/.

References

1 Cutts, F. T. & Hanson, M. Seroepidemiology: an underused tool for designing and monitoring vaccination programmes in low- and middle-income countries. *Trop. Med. Int. Health*. 21, 1086-1098 (2016).

2 Wilson, S. E., Deeks, S. L., Hatchette, T. F. & Crowcroft, N. S. The role of seroepidemiology in the comprehensive surveillance of vaccine-preventable diseases. *Cmaj*. 184, E70-76 (2012).

3 Horby, P. W. *et al.* CONSISE statement on the reporting of Seroepidemiologic Studies for influenza (ROSES-I statement): an extension of the STROBE statement. *Influenza Other Respir. Viruses*. 11, 2-14 (2017).

4 Mina, M. J. *et al.* A Global Immunological Observatory to meet a time of pandemics. *Elife*. 9, e58989 (2020).

5 U.S. Food and Drug Administration. FDA Approves New Drug to Prevent RSV in Babies and Toddlers. https://www.fda.gov/news-events/press-announcements/fda-approves-new-drug-prevent-rsv-babies-and-toddlers (2023).

6 U.S. Food and Drug Administration. FDA Approves First Respiratory Syncytial Virus (RSV) Vaccine. https://www.fda.gov/news-events/press-announcements/fda-approves-first-respiratory-syncytial-virus-rsv-vaccine (2023).

7 U.S. Food and Drug Administration. FDA Approves First Vaccine for Pregnant Individuals to Prevent RSV in Infants. https://www.fda.gov/news-events/press-announcements/fda-approves-first-vaccine-pregnant-individuals-prevent-rsv-infants (2023).

8 U.S. Food and Drug Administration. MRESVIA. https://www.fda.gov/vaccines-blood-biologics/vaccines/mresvia (2024).

9 Fong, Y. *et al.* Antibody Correlates of Protection From Severe Respiratory Syncytial Virus Disease in a Vaccine Efficacy Trial. *Open Forum Infect. Dis.* 10, ofac693 (2023).

10 Li, Y. *et al.* Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis. *Lancet*. 399, 2047-2064 (2022).

11 Shi, T. *et al.* Global Disease Burden Estimates of Respiratory Syncytial Virus-Associated Acute Respiratory Infection in Older Adults in 2015: A Systematic Review and Meta-Analysis. *J. Infect. Dis.* 222, S577-s583 (2020).

12 Piliper, E. A., Reed, J. C. & Greninger, A. L. Clinical validation of an RSV neutralization assay and analysis of cross-sectional sera associated with 2021-2023 RSV outbreaks to investigate the immunity debt hypothesis. *Microbiol. Spectr.* 12, e0211524 (2024).

13 Ma, C. *et al.* Immune correlates analysis of mRNA-1345 RSV vaccine efficacy clinical trial. *Nat. Commun.* 16, 6118 (2025).

14 Sun, K. *et al.* SARS-CoV-2 correlates of protection from infection against variants of concern. *Nat. Med.* 30, 2805-2812 (2024).

15 Regev-Yochay, G. *et al.* Correlates of protection against COVID-19 infection and intensity of symptomatic disease in vaccinated individuals exposed to SARS-CoV-2 in households in Israel (ICoFS): a prospective cohort study. *Lancet Microbe*. 4, e309-e318 (2023).

16 Ai, J. *et al.* Antibody evasion of SARS-CoV-2 Omicron BA.1, BA.1.1, BA.2, and BA.3 sub-lineages. *Cell Host Microbe*. 30, 1077-1083.e1074 (2022).

17 Cohen, B. J., Audet, S., Andrews, N. & Beeler, J. Plaque reduction neutralization test for measles antibodies: Description of a standardised laboratory method for use in immunogenicity studies of aerosol vaccination. *Vaccine*. 26, 59-66 (2007).

18 Vaidya, S. R. *et al.* Development of a focus reduction neutralization test (FRNT) for detection of mumps virus neutralizing antibodies. *J. Virol. Methods*. 163, 153-156 (2010).

19 Vaidya, S. R., Kumbhar, N. S. & Bhide, V. S. Detection of measles, mumps and rubella viruses by immuno-colorimetric assay and its application in focus reduction neutralization tests. *Microbiol. Immunol.* 58, 666-674 (2014).

20 Salje, H. *et al.* Variability in dengue titer estimates from plaque reduction neutralization tests poses a challenge to epidemiological studies and vaccine development. *PLoS Negl. Trop. Dis.* 8, e2952 (2014).

21 ROSES-S: Statement from the World Health Organization on the reporting of seroepidemiologic studies for SARS-CoV-2. *Influenza Other Respir. Viruses*. 15, 561-568 (2021).

22 Bergeri, I. *et al.* WHO's Investigations and Studies, Unity Studies: A global initiative creating equitable opportunities for enhanced surveillance, operational research, capacity building, and global knowledge sharing. *Influenza Other Respir. Viruses*. 18, e13256 (2024).

23 Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-Response Analysis Using R. *PLoS One*. 10, e0146021 (2015).

24 Malyutina, A., Tang, J. & Pessia, A. drda: An R Package for Dose-Response Data Analysis Using Logistic Functions. *J. Stat. Softw.* 106, 1-26 (2023).

25 Palmer, P. *et al.* AutoPlate: Rapid Dose-Response Curve Analysis for Biological Assays. *Front. Immunol.* 12, 681636 (2021).

26 Feng, F., Sales, A. P. & Kepler, T. B. A Bayesian approach for estimating calibration curves and unknown concentrations in immunoassays. *Bioinformatics*. 27, 707-712 (2011).

27 Gelman, A., Chew, G. L. & Shnaidman, M. Bayesian analysis of serial dilution assays. *Biometrics*. 60, 407-417 (2004).

28 Wang, Q. *et al.* Antibody responses to respiratory syncytial virus: A population-based cross-sectional serological study in southern China, 2021. *Clin. Microbiol. Infect.* 30, 1183-1189 (2024).

29 Wei, X. *et al.* The transfer and decay of maternal antibodies against enterovirus A71, and dynamics of antibodies due to later natural infections in Chinese infants: a longitudinal, paired mother-neonate cohort study. *Lancet Infect. Dis.* 21, 418-426 (2021).

30 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 8, 118-127 (2007).

31 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 28, 882-883 (2012).

32 Zhang, Y., Jenkins, D. F., Manimaran, S. & Johnson, W. E. Alternative empirical Bayes models for adjusting for batch effects in genomic studies. *BMC Bioinformatics*. 19, 262 (2018).

33 Tomo, Y. & Nakaki, R. iComBat: An incremental framework for batch effect correction in DNA methylation array data. *Comput Struct Biotechnol J*. 27, 4121-4131 (2025).

34 Beran, J. *et al.* Safety and Immunogenicity of 3 Formulations of an Investigational Respiratory Syncytial Virus Vaccine in Nonpregnant Women: Results From 2 Phase 2 Trials. *J. Infect. Dis.* 217, 1616-1625 (2018).

35 World Health Organization. Guidelines on the quality, safety and efficacy of respiratory syncytial virus vaccines, Annex 2, TRS No 1024. https://www.who.int/publications/m/item/respiratory-syncytial-virus-vaccines-annex-2-trs-no-1024 (2020).

36 Swart, A., Maas, M., de Vries, A., Cuperus, T. & Opsteegh, M. Bayesian Binary Mixture Models as a Flexible Alternative to Cut-Off Analysis of ELISA Results, a Case Study of Seoul Orthohantavirus. *Viruses*. 13, 1155 (2021).

37 Xu, M. *et al.* Transplacental transfer efficiency and longitudinal dynamics of antibodies against RSV in Chinese children from birth to 8 years: a paired mother-neonate cohort study. Preprint at http://medrxiv.org/content/early/2025/10/17/2025.10.15.25338064.abstract (2025).

38 Park, Y. *et al.* Comparison of plaque reduction and focus reduction neutralization tests for the measurement of neutralizing antibody titers against japanese encephalitis virus. *J. Virol. Methods*. 306, 114540 (2022).

39 McDonald, J. U., Rigsby, P., Dougall, T. & Engelhardt, O. G. Establishment of the first WHO International Standard for antiserum to Respiratory Syncytial Virus: Report of an international collaborative study. *Vaccine*. 36, 7641-7649 (2018).

40 Simões, E. A. F. *et al.* Prefusion F Protein-Based Respiratory Syncytial Virus Immunization in Pregnancy. *New. Engl. J. Med.* 386, 1615-1626 (2022).

41 Stan Development Team. The R interface to Stan. https://cloud.r-project.org/web/packages/rstan/index.html (2025).

42 Ritz, C. Analysis of Dose-Response Curves. https://cran.r-project.org/web/packages/drc/refman/drc.html (2016).

43 Wang, Y. BHM_Titer_Correction. Zenodo https://zenodo.org/records/19445779 (2026).

Acknowledgements

We thank all participants for providing their information and serum samples for this study. We acknowledge the staff of the Anhua County Center for Disease Control and Prevention (CDC) and the township health centers at the study sites for their support with field investigations and sample collection. The computations in this research were performed using the CFFF platform of Fudan University. Drawings in Figure 1A was created using BioRender.com.

Funding

This study was supported by the Prevention and Control of Emerging and Major Infectious Diseases-National Science and Technology Major Project (2025ZD01901300 to H.Y.), a Wellcome Trust Early Career Award (225001/Z/22/Z to J.A.H) and a fellowship from the China Postdoctoral Science Foundation (2024M750556 to Y.W).

Author Contributions Statement

H.Y. conceived and designed the study. H.Y. and J.A.H. co-supervised the project. Q.W. and M.X. established and optimized the laboratory methods. Q.W., M.X., L.Y. and J.Z. performed the laboratory experiments. L.Y. and J.Z. prepared the experimental data. Y.W., J.A.H., and C.W. developed the model. Y.W. analyzed the experimental data, conducted the simulations, and prepared the first draft of the manuscript. H.Y., J.A.H., and C.W. critically revised the manuscript for important intellectual content. All authors contributed to the review and revision of the manuscript, approved the final version as submitted, and agreed to be accountable for all aspects of the work.

Competing Interests Statement

H.Y. has received research funding from Sanofi Pasteur, Shenzhen Sanofi Pasteur Biological Products Co., Ltd, Shanghai Roche Pharmaceutical Company, and SINOVAC Biotech Ltd. None of the research funding is related to this work. J.A.H. has received consulting fees from GLG (Gerson Lehrman Group, Inc). All other authors report no competing interests.

Figure Legends

Fig. 1 | Workflow and data interpretation of the RSV FRNT assay. (A) Overview of the RSV FRNT experimental procedure. (B) Layout of serum samples and assay controls on 96-well plates in each experimental batch. Shading from dark to light indicates serial dilutions from high to low concentration. The VC was prepared at a fixed dilution. **(C) Example readouts showing count of foci from serially diluted serum samples.** Orange points and dashed lines represent results from two replicates; red points and solid line indicate replicate means. **Abbreviations:** RSV, respiratory syncytial virus; FRNT, foci reduction neutralization test; VC, virus control; PC, internal positive control; IS, International Standard.

Fig. 2 | Bayesian hierarchical modelling framework for correcting batch effects in RSV FRNT measurements. (A) A simplified experimental design for RSV FRNT. Serum samples are tested across batches under different groups of assay-level covariates (e.g., virus working dilution). A virus control (VC) and a positive control (PC) were included consistently in all batches. **(B) Schematic model diagrams along with the corresponding model fits applied to VC, PC, and serum sample data.** Error bars represent the 95% confidence interval (CI) of foci counts from replicated VC wells in each batch, with the central point indicating the mean. The model diagram used for fitting International Standards (IS) data is the same as that used for the PC and is therefore not shown. Model notation is defined in Methods and Table S3.

Fig. 3 | Posterior estimates from the Bayesian hierarchical modelling framework based on experimental data. (A) Foci count in VC across batches using different working virus stocks. Open circles represent observed counts of foci from 603 replicated VC wells across 28 batches. Colored points with error bars represent the means and 95% CIs of observed data. Red points with error bars represent posterior medians and 95% credible intervals (CrIs). **(B) Comparison of BHM-adjusted and unadjusted curves for PC and IS across batches.** Red curves with error bands represent the posterior medians and 95% CrIs of BHM-adjusted curves. Orange curves represent the posterior medians of BHM-unadjusted curves. Black points at 50% foci reduction (i.e., the grey dashed line) indicate BHM-adjusted titers, while error bars represent the range of BHM-unadjusted titers based on replicates of PC (N = 67) and IS (N = 24 for IS500; N = 65 for IS1000) across batches. Notably, the BHM-adjusted nAb titer for the IS1000 is 842.85, suggesting that the nAb titers in this study can be converted to International Units per milliliter (IU/mL) by multiplying a conversion factor of 0.421 (i.e., 842.85/2000)[40]. **(C) Pairwise agreement between BHM-adjusted titer estimates and those derived from Kärber formula, 4PL model, and BHM-unadjusted model based on 1,777 serum samples.** The vertical and horizontal dashed lines indicate the seropositivity threshold. Data were analyzed using a two-sided Spearman rank correlation test. Spearman's correlation coefficient (ρ) and P values are reported. RMSE was calculated from the titer differences between the two methods on the $\log_2$ scale. **Abbreviations:** 4PL, four-parameter logistic model; BHM, Bayesian hierarchical

model; RMSE, root mean square error; IS500 and IS1000, International Standard tested at 500 and 1000 International Unit (IU), respectively.

Fig. 4 | Posterior estimates from the Bayesian hierarchical modelling framework based on simulated data. (A) Comparison of true and estimated mean foci counts in VC across 28 simulated batches (22 replicates per batch; N = 616). Colored points with error bars indicate posterior medians and 95% CrIs. **(B) Comparison of true and fitted curves for PC in representative batches.** Solid curves show the simulated true curves; dashed curves with error bands show the posterior medians and 95% CrIs. **(C) Fitted curves and nAb titer estimates for three serum samples from the same batches in (B).** Black lines indicate the simulated true curves. Green circles represent foci reduction data calculated relative to the statistical mean foci count of VC in each batch, which were used for the Kärber formula and 4PL model, as defined in Methods (Equation 1); Orange circles represent the same data but calculated relative to the posterior mean foci count of VC in each working virus stock, as defined in Methods (Equation 7). Intersections of the curves with the horizontal dashed line at 50% foci reduction mark the corresponding nAb titer estimates for each method; estimates from the Kärber formula are shown as inverted triangles. Samples with fitted curves that did not achieve a 50% foci reduction were considered seronegative and had their titers imputed as the lower limit of quantification (LLOQ). Titer estimates exceeding the final dilution point (87,480) were censored at 87,480 (i.e., upper limit of quantification [ULOQ]).

Fig. 5 | Agreement between estimated and true nAb titers across methods based on simulated data (N=560). (A) Kärber formula. (B) 4PL model. (C) BHM-unadjusted model. (D) BHM-adjusted model. (E) Differences between estimated and true nAb titers. The vertical and horizontal dashed lines in (A)-(D) indicate the seropositivity threshold. Points in the upper-left quadrant relative to the intersection represent false positives, whereas points in the lower-right quadrant represent false negatives. Data were analyzed using a two-sided Spearman rank correlation test. Spearman's correlation coefficient (ρ) and *P* values are reported. RMSE: root mean square error between the true and estimated nAb titers on the $\log_2$ scale. Coverage: the proportion of simulations in which the true titer falls within the corresponding 95% CrI. Violin and boxplots in (E) showing the distribution of residuals between estimated and true titers across different methods. Boxes represent the interquartile range (IQR) with the central line indicating the median. Whiskers extend to the most extreme values within 1.5 × IQR.

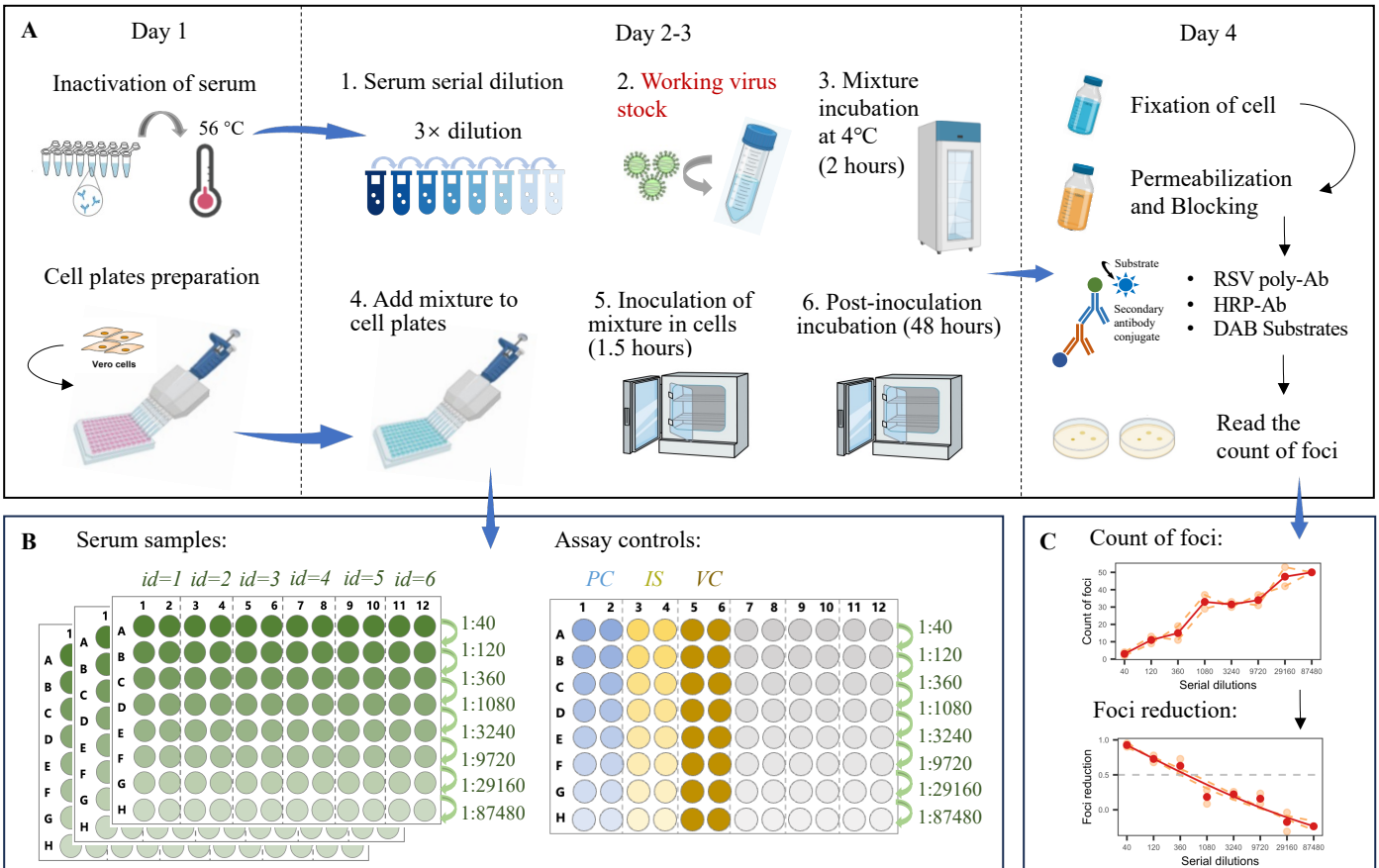
Fig. 6 | Comparison of seroepidemiological indicators based on RSV nAb titer estimates from each method. (A) Age-specific geometric mean titers (GMTs). The horizontal dashed line indicates the seropositivity threshold. Colored points represent the GMTs, with error bars showing the 95% CIs. **(B) Age-specific seroprevalence.** Colored points represent the seroprevalence, with error bars showing the 95% CIs. **(C) Agreement of the Kärber formula, 4PL model, and BHM-unadjusted model with the BHM-adjusted model in identifying**

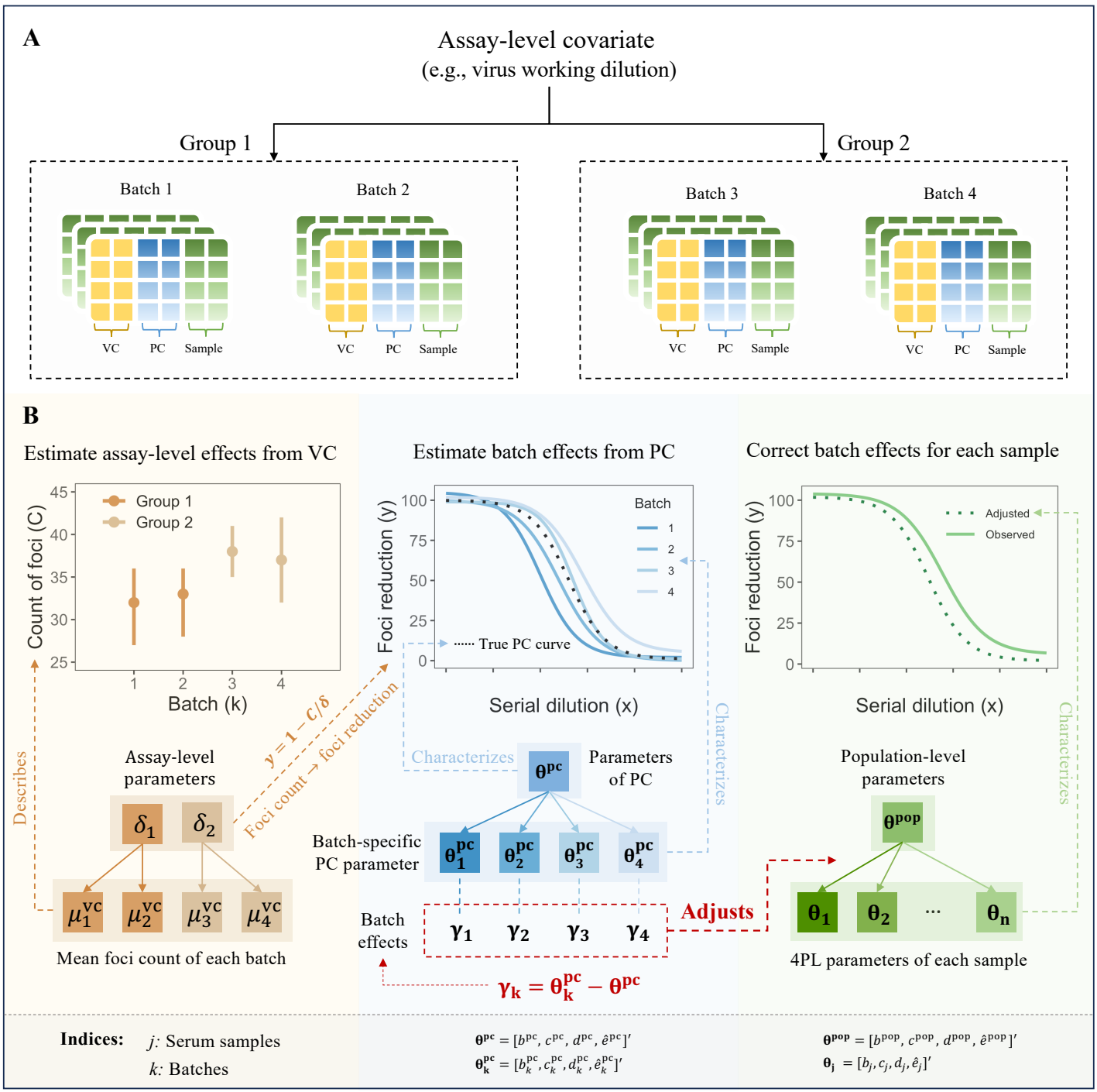
seroconversion and antibody fold rises based on 715 paired serum samples. Colors from dark to light indicate increasing agreement. Source data are provided as a Source Data file.

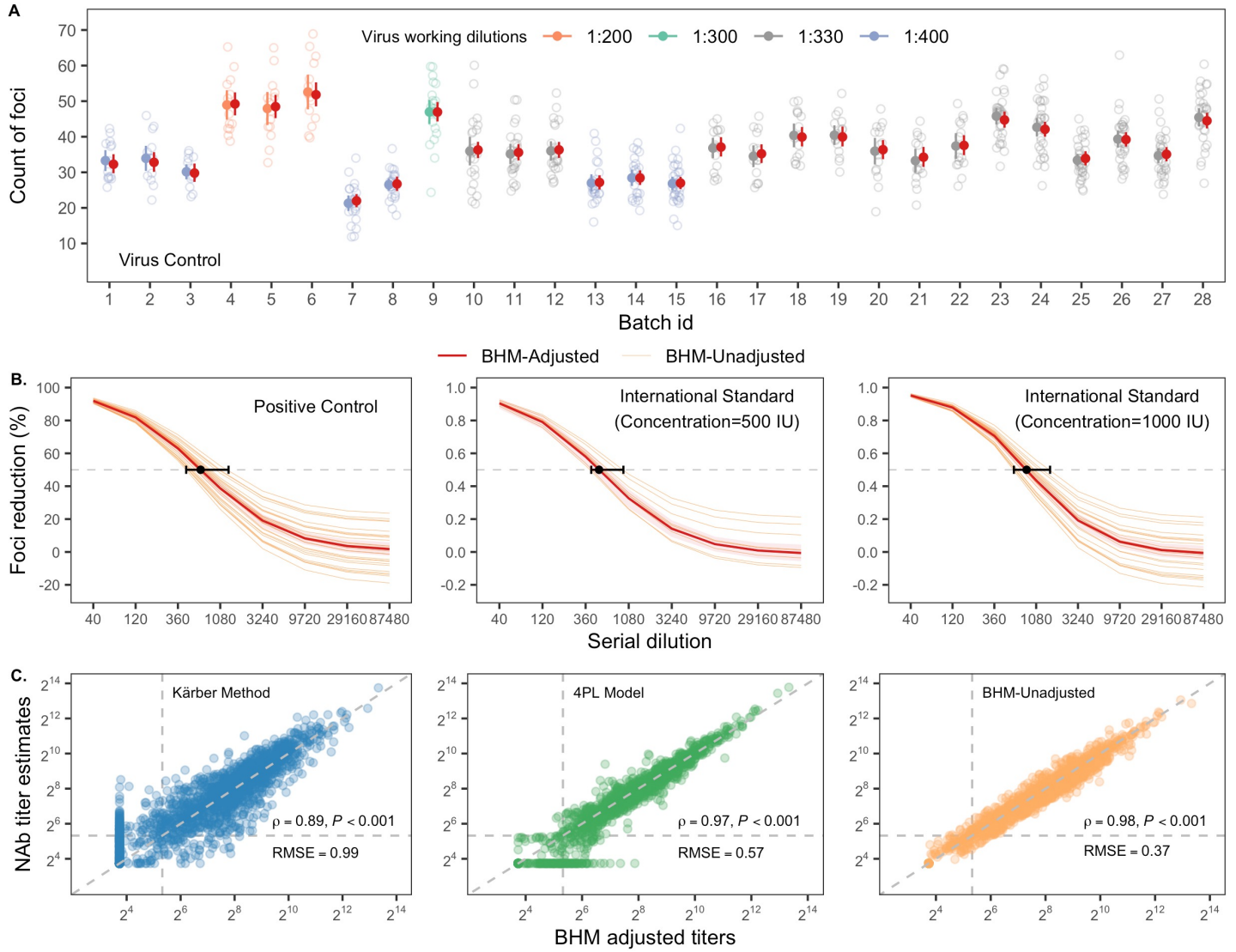
Editorial Summary –

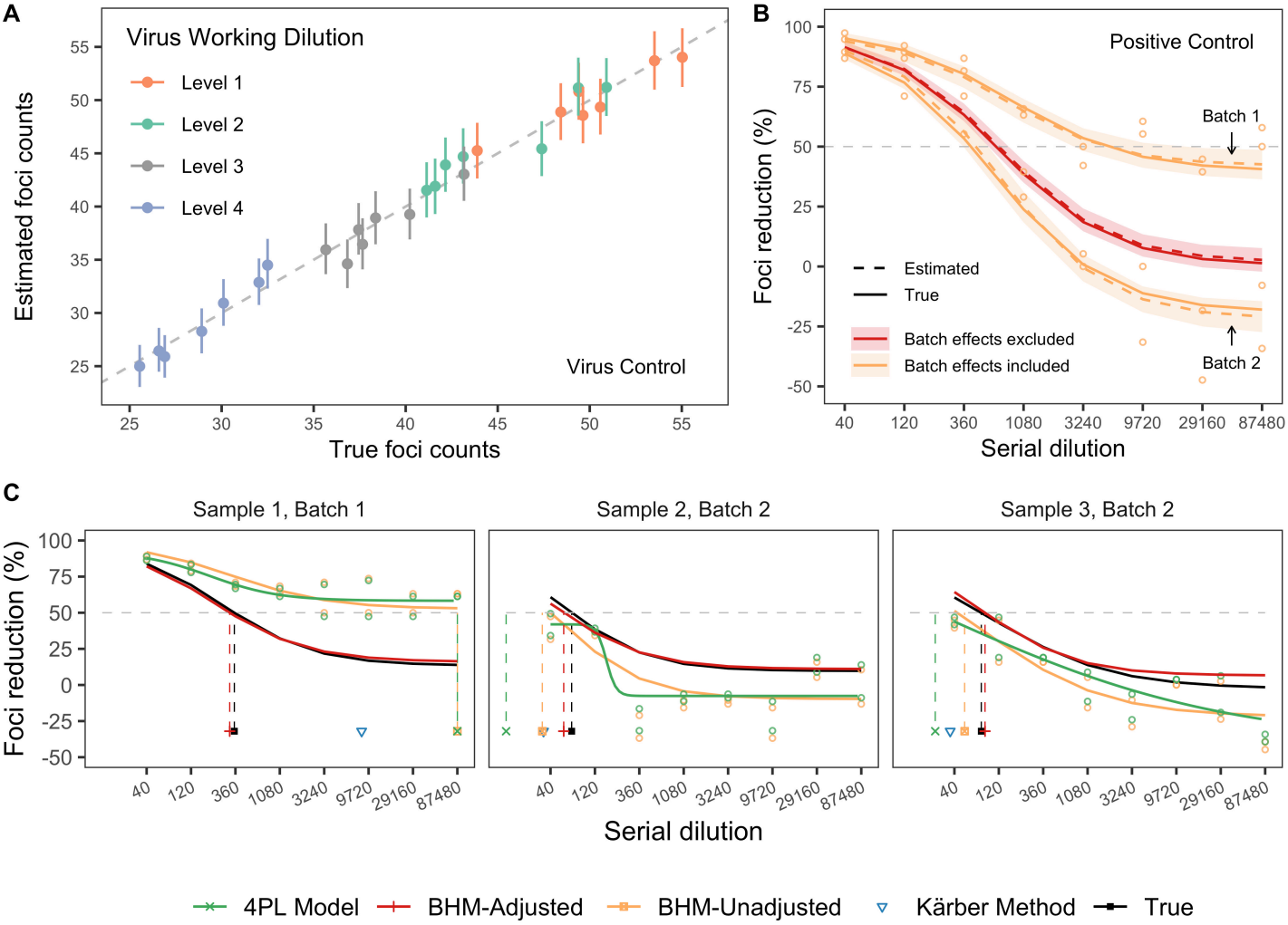
Measuring antibody levels is crucial to assess population immunity and guide vaccine development. Here, the authors develop a Bayesian hierarchical model (BHM) that provides accurate and robust estimates of neutralizing antibody titers.

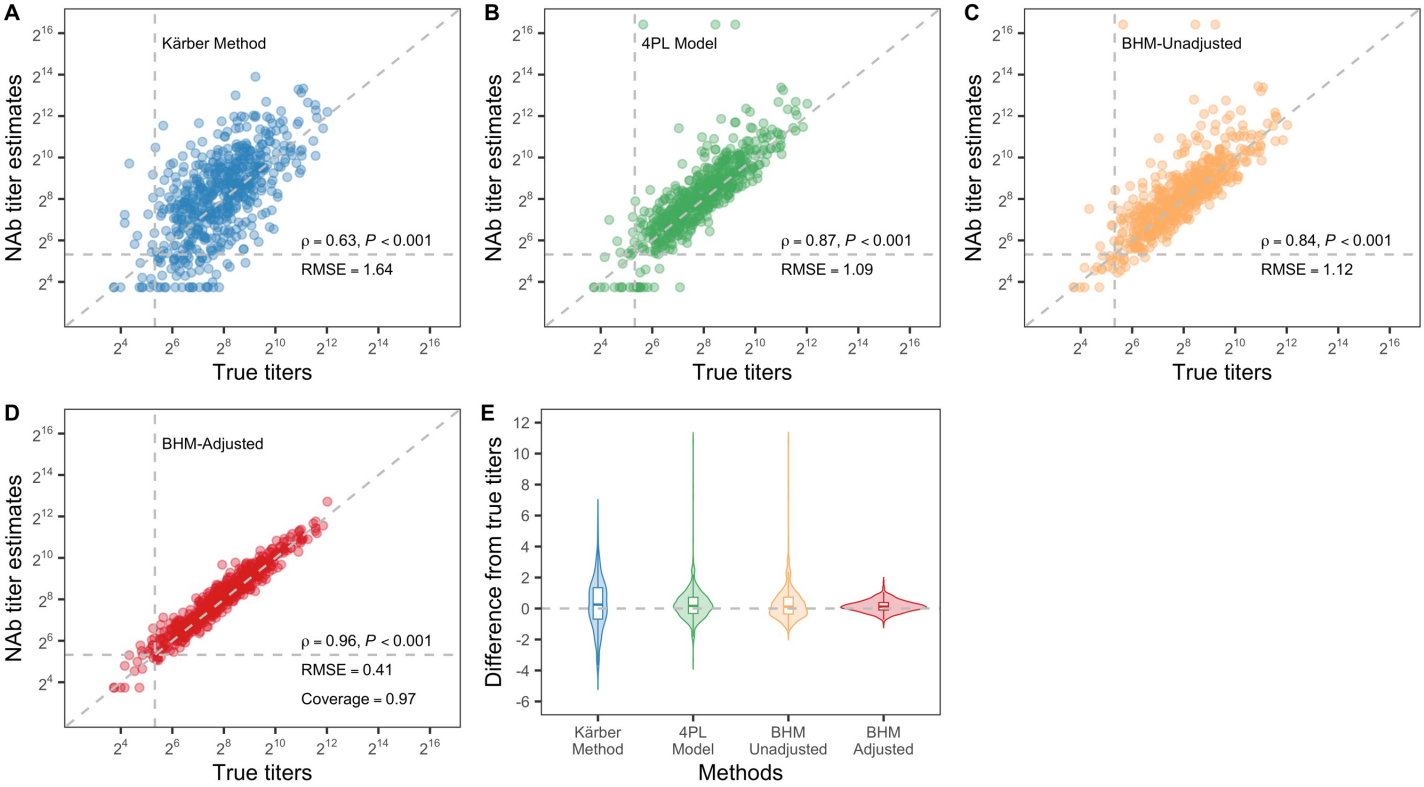
Peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

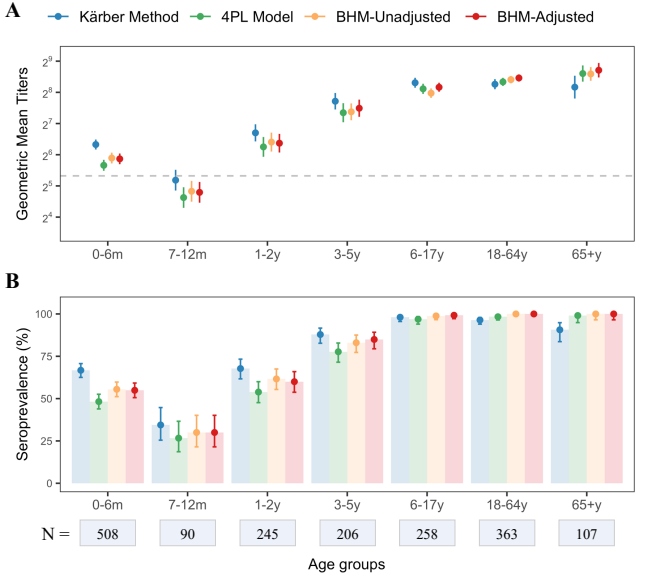

A
Day 1
Inactivation of serum
56 °C
Cell plates preparation
Vero cells
Day 2-3
1. Serum serial dilution
3× dilution
2. Working virus stock
3. Mixture incubation at 4°C (2 hours)
4. Add mixture to cell plates
5. Inoculation of mixture in cells (1.5 hours)
6. Post-inoculation incubation (48 hours)
Day 4
Fixation of cell
Permeabilization and Blocking
Substrate
Secondary antibody conjugate
• RSV poly-Ab
• HRP-Ab
• DAB Substrates
Read the count of foci
B
Serum samples:
id=1 id=2 id=3 id=4 id=5 id=6
1:40
1:120
1:360
1:1080
1:3240
1:9720
1:29160
1:87480
Assay controls:
PC IS VC
1:40
1:120
1:360
1:1080
1:3240
1:9720
1:29160
1:87480
C
Count of foci:
Count of foci
Serial dilutions
Foci reduction:
Foci reduction
Serial dilutions


A
Assay-level covariate
(e.g., virus working dilution)
Group 1
Group 2
Batch 1
Batch 2
Batch 3
Batch 4
VC
PC
Sample
B
Estimate assay-level effects from VC
Count of foci (C)
Group 1
Group 2
Batch (k)
Estimate batch effects from PC
Foci reduction (y)
Batch
1
2
3
4
True PC curve
Serial dilution (x)
Correct batch effects for each sample
Adjusted
Observed
Describes
Characterizes
Assay-level parameters
$y = 1 - C/\delta$
Foci count → foci reduction
δ_1 δ_2
μ_1^{vc} μ_2^{vc} μ_3^{vc} μ_4^{vc}
Mean foci count of each batch
θ^{pc}
Parameters of PC
Batch-specific PC parameter
θ_1^{pc} θ_2^{pc} θ_3^{pc} θ_4^{pc}
Batch effects
γ_1 γ_2 γ_3 γ_4
Adjusts
$\gamma_k = \theta_k^{pc} - \theta^{pc}$
Population-level parameters
θ^{pop}
θ_1 θ_2 ... θ_n
4PL parameters of each sample
Indices:
j: Serum samples
k: Batches
$\theta^{pc} = [b^{pc}, c^{pc}, d^{pc}, \hat{e}^{pc}]'$
$\theta_k^{pc} = [b_k^{pc}, c_k^{pc}, d_k^{pc}, \hat{e}_k^{pc}]'$
$\theta^{pop} = [b^{pop}, c^{pop}, d^{pop}, \hat{e}^{pop}]'$
$\theta_j = [b_j, c_j, d_j, \hat{e}_j]'$


A
Virus working dilutions
1:200
1:300
1:330
1:400
Count of foci
Virus Control
Batch id
BHM-Adjusted
BHM-Unadjusted
B.
Positive Control
International Standard
(Concentration=500 IU)
International Standard
(Concentration=1000 IU)
Foci reduction (%)
Serial dilution
C.
Kärber Method
4PL Model
BHM-Unadjusted
NAb titer estimates
ρ = 0.89, P < 0.001
RMSE = 0.99
ρ = 0.97, P < 0.001
RMSE = 0.57
ρ = 0.98, P < 0.001
RMSE = 0.37
BHM adjusted titers


A
Virus Working Dilution
Level 1
Level 2
Level 3
Level 4
Virus Control
Estimated foci counts
True foci counts
B
Positive Control
Batch 1
Batch 2
Estimated
True
Batch effects excluded
Batch effects included
Foci reduction (%)
Serial dilution
C
Sample 1, Batch 1
Sample 2, Batch 2
Sample 3, Batch 2
Foci reduction (%)
Serial dilution
4PL Model
BHM-Adjusted
BHM-Unadjusted
Kärber Method
True


A
Kärber Method
ρ = 0.63, P < 0.001
RMSE = 1.64
NAb titer estimates
True titers
B
4PL Model
ρ = 0.87, P < 0.001
RMSE = 1.09
NAb titer estimates
True titers
C
BHM-Unadjusted
ρ = 0.84, P < 0.001
RMSE = 1.12
NAb titer estimates
True titers
D
BHM-Adjusted
ρ = 0.96, P < 0.001
RMSE = 0.41
Coverage = 0.97
NAb titer estimates
True titers
E
Difference from true titers
Kärber Method
4PL Model
BHM Unadjusted
BHM Adjusted
Methods



C

BHM-Adjusted Results	N	Agreement with BHM-adjusted results (%)		
		Kärber Method	4PL Model	BHM-Unadjusted
Seroconversion				
Yes	114	68.4%	85.1%	96.5%
No	601	95.7%	97.3%	98.7%
Overall	715	91.3%	95.4%	98.3%
Two-fold titer rise				
Yes	71	64.8%	60.6%	87.3%
No	644	93.8%	98.3%	99.2%
Overall	715	90.9%	94.5%	98.0%
Four-fold titer rise				
Yes	39	61.5%	59.0%	100.0%
No	676	97.8%	99.7%	99.7%
Overall	715	95.8%	97.5%	99.7%