Biometrics, 2026, 82(3), ujag134
https://doi.org/10.1093/biomtc/ujag134
Published: 24 July 2026
Biometric Practice

OXFORD

A Bayesian contamination model for serial dilution assays

Siquan Wang[1], Qixuan Chen[1,*], Matthew S. Perzanowski[2], Adnan Divjan[3], Andrew Gelman[3]

[1]Department of Biostatistics, Columbia University, New York, NY 10032, United States
[2]Department of Environmental Health Sciences, Columbia University, New York, NY 10032, United States
[3]Departments of Statistics and Political Science, Columbia University, New York, NY 10027, United States

[*]Corresponding author: Qixuan Chen, Department of Biostatistics, Columbia University, 722 West 168th Street, New York, NY 10032, United States (qc2138@cumc.columbia.edu)

ABSTRACT

Serial dilution assays are essential tools in biomedical research. Current analytical methods work well when measurements are in the steep part of the calibration curve but are not so efficient when measurements are near the upper and lower bounds. Moreover, sample contamination is common in serial dilution assays and can result in problematic concentration estimation. To address these challenges, we propose a Bayesian contamination model that simultaneously flags contaminated samples and estimates the concentrations of unknown samples, while quantifying estimation uncertainty. In extensive simulation studies under various contamination scenarios, our method consistently outperforms commonly used approaches. We further demonstrate a step-by-step application of a Bayesian workflow for modeling contamination using raw immunoassay data from the New York City Neighborhood Asthma and Allergy Study. This work advances previous research by introducing a practical method for detecting contaminated samples and estimating unknown concentrations in immunoassays and by presenting a general framework for incorporating unknown contamination processes into the Bayesian modeling workflow.

Keywords: Bayesian robust inference, immunoassays, indoor allergens, measurement error models, mixture models, sample contamination

1 Introduction

Epidemiologic investigations of environmental-disease associations rely on accurate measurements of environmental exposures. These exposures are typically assessed by collecting samples and measuring concentrations in the laboratory using immunoassays with serial dilutions. Such assays are widely used to detect and quantify pollutants; measure pesticide concentrations; monitor contaminants; and assess exposure to environmental chemicals.

An immunoassay is a biochemical test used to measure the presence or concentration of an analyte based on its interaction with an antibody or antigen (Vashist and Luong, 2018). Among the various types of immunoassays, the Enzyme-Linked Immunosorbent Assay (ELISA), developed in the 1970s, remains a foundational technique for detecting and quantifying proteins and other antibody-binding substances (Engvall and Perlmann, 1972; Van Weemen and Schuurs, 1971). In recent years, there has been growing interest in multiplex assays, which enable the simultaneous detection of multiple target analytes within a single reaction volume (Carter et al., 2001; Olmedo et al., 2011; Phipatanakul et al., 2020). These assays offer significant advantages, including reduced sample consumption and increased efficiency by eliminating the need for multiple singleplex assays. In 2025 alone, more than 1500 articles can be identified in PUBMED with the search term “ELISA” and about 500 articles with “multiplex assays.”

The immunoassay test is typically performed using serial dilutions of both standard and unknown samples in a microtiter plate. Signal responses from antigen–antibody interactions are measured at each dilution. Serial dilutions of the standard sample are used to construct a calibration curve that relates signal intensity to known concentrations. This curve is then used to estimate concentrations of the unknown samples. Performing multiple dilutions of unknowns allows for improved estimation accuracy and provides a measure of uncertainty.

Received: 10 May 2025. **Revised:** 12 June 2026. **Accepted:** 3 July 2026

The classical calibration method is the industry-standard approach for estimating concentrations of unknown samples and is typically implemented in three steps. First, a calibration response curve is constructed by regressing signal intensity, y, on known concentrations of standard samples x using $y = g(x) + \text{error}$, where $g(x)$ is commonly fit using a four- or five-parameter logistic model (Prentice, 1976). Second, concentrations of the diluted unknown samples are estimated by reading values from the estimated calibration curve using $\widehat{g}^{-1}(y)$. Finally, these estimated concentrations are rescaled by dividing by the corresponding dilution factors and then averaged to obtain a final estimate of the concentration for each unknown sample. Although it is easy to implement, the classical calibration approach has several limitations. First, estimating the concentration by inverting the calibration curve does not account for calibration uncertainty and can lead to significant measurement error (Lee and Whitmore, 1999). Second, variances in measurements at the extreme ends of the calibration curve can place sample measurements below limits of detection (LOD), and these measurements are typically replaced with a constant such as 0, LOD, LOD/2, or LOD/$\sqrt{2}$, but any of these options can lead to serious bias in statistical inference (Guo et al., 2010). Third, averaging the estimates of diluted samples is inefficient since measurements of highly diluted samples can have greater variance after scaling the measurements by dilution levels. Calibration methods that use weighted average across estimates of diluted samples partially address this problem by introducing variance-based weights, so that more informative dilutions contribute more to the final concentration estimate (Cheung et al., 2015; Xu et al., 2016).

Another major challenge in immunoassay analysis is sample contamination, in which samples contain not only the analytes of interest but also interfering substances that affect assay performance. Web Figure 1 shows the results of an experiment evaluating the impact of sample contamination in measuring a protein from dust mites, Der p 1, using immunoassays. In this experiment, a commercially available laundry detergent was added as a contaminant to 16 dust samples at 8 different contamination concentrations, ranging from 1 in 10 to 1 in 100 million. Serial dilution was then applied to each of the 16 contaminated samples. The signal response curves look very different across different contamination concentrations, with flatter response curves associated with more contaminated samples. Thus, when contamination exists in unknown samples, the classical calibration inference method is no longer applicable, because it assumes that unknown samples share the same response curve as the calibration samples, whereas the contaminated samples may exhibit signal response curves that differ from the calibration curve. Since the concentration or even presence of these substances is often unknown, methods that can detect and account for potential contamination in the samples are needed when analyzing immunoassay data.

Previous work on concentration estimation in serial dilution assays has explored a range of statistical approaches. Early efforts include parameter estimation of the dose-count curve under a model assuming variance proportional to the mean count in radioligand assays (Finney, 1976), and the use of summary statistics, including mean, median, and standard deviation, to estimate threshold concentrations while accounting for grouped observations (Hamilton and Rinaldi, 1988). Other methods have modeled concentration-response relationships by assuming additive sequential dilution errors on the log-scale of concentrations between reference and test preparation (Racine-Poon et al., 1991), or have modeled serial dilution error by incorporating propagation of random measurement error in the dilution process (Higgins et al., 1998). Lognormal measurement error models have also been proposed for serial dilution data (Lee and Whitmore, 1999).

Bayesian methods have been shown to yield improved estimates of unknown concentrations in serial dilution assays. Gelman et al. (2004) proposed a Bayesian calibration framework using hierarchical models that account for multiple sources of variation. Klauenberg et al. (2015) extended that work by adding an informative prior to the calibration model. A Bayesian lognormal measurement error model with constant variances has been shown to outperform the classical calibration curve inversion method (Feng et al., 2011). Additional robust Bayesian approaches include model averaging weighted by posterior probabilities (Morales et al., 2006) and a multilevel model that separates experimental noise into detection sensor noise and stochastic natural noise (Fong et al., 2012). Despite these advances, none of the existing methods explicitly address contamination in serial dilution assay data.

The Bayesian approach to contaminated data is to use a mixture model so that each observation has a large probability of coming from the uncontaminated model and a small probability of coming from a contaminated distribution (Box and Tiao, 1968). The probability of contamination can be estimated from the data. In this paper, we develop a novel Bayesian contamination model for concentration estimation in serial dilution assays that accounts for potential contamination in the unknown samples. The goal is to enable users to estimate underlying concentrations while simultaneously identifying samples likely affected by contamination. We propose to address the challenges of Bayesian contamination models using several steps. First, we fit a mixture model that allows the modeling of sample contamination. Second, we use estimated mixture probabilities to flag potentially contaminated samples while also aiming for the contamination model to provide rough inference for underlying parameters of interest. Third, we apply posterior predictive checks to assess model fit; any deviations identified through these checks can inform refinements to the contamination model.

In practice, both the extent and nature of contamination are unknown, which introduces additional complexity. As a result, we anticipate some systematic error in inference under contamination. However, identifying likely contamination through our model can serve as a valuable diagnostic tool and motivate further investigation or experimental validation.

2 Motivating application

Our study focuses on immunoassay data measuring indoor allergen levels from the New York City Neighborhood Asthma

Table 1 Standards and selected unknown samples from a microtiter plate with the concentrations of the new samples estimated using the classical calibration method (OOR<: out of range; FI: fluorescence intensity units)

Sample ID	Dilution	Signal Resp (FI)	Conc. (ng/ml)	Sample ID	Dilution	Signal Resp (FI)	Conc. (ng/ml)
Standard	1	16 418	125	Standard	1/128	906	0.98
	1	18 977	125		1/128	1141	0.98
	1/2	16 350	62.5		1/256	397	0.49
	1/2	17 960	62.5		1/256	450	0.49
	1/4	14 573	31.25		1/512	164	0.24
	1/4	14 625	31.25		1/512	166	0.24
	1/8	12 380	15.63		1/1024	72.17	0.12
	1/8	13 310	15.63		1/1024	77.83	0.12
	1/16	8728	7.81		1/2048	34.33	0.06
	1/16	9175	7.81		1/2048	39.27	0.06
	1/32	5152	3.91		0	4	0
	1/32	5313	3.91		0	4	0
	1/64	2353	1.95				
	1/64	2424	1.95				
Unk 10	1/10	4259.5	32.47	Unk 23	1/10	1904	15.87
	1/100	241	32.21		1/100	783	78.78
	1/10 000	6	OOR<		1/10 000	39	647.51
Unk 3	1/10	660	6.92	Unk 22	1/10	51	0.87
	1/100	646	68.04		1/100	7	OOR<
	1/10 000	74	1233.87		1/10 000	3	OOR<
Unk 9	1/10	475	5.39	Unk 16	1/10	9	OOR<
	1/100	251	33.23		1/100	5	OOR<
	1/10 000	6	OOR<		1/10 000	8	OOR<

and Allergy Study (NYC NAAS), which includes 7- and 8-year-old children with and without asthma (Perzanowski et al., 2008). Homes of participating families were visited, and dust samples were collected from each child's bed by vacuuming the fitted sheet on the upper half of the bed and both sides of the pillows. Samples were extracted with PBS 0.05% Tween, pH 7.4 at a concentration of 50 mg/ml and stored at −20°C until analysis. Allergens from cockroach, cat, dog, rat, mouse, and dust mites were measured by multiplex assays using microtiter plates.

Each microtiter plate contains 96 wells (12 columns and 8 rows). Web Table 1 shows the layout of standards and unknown samples of a plate, which includes 2 replicates of standards (columns 1–3) and 23 unknown samples (columns 4–12). Each standard sample is prepared at a fixed known concentration and diluted using 12 two-fold dilutions ranging from 1 to 1/2048. Each unknown sample is measured at three dilution levels (1/10, 1/100, and 1/10 000), chosen to enable estimation across a wide range of concentrations.

Table 1 shows standard data for the dust mite allergen, Der f 1, and concentration estimates for 6 selected unknown samples obtained using the classical calibration method on a microtiter plate. The top portion shows signal responses and true concentrations for the standards, with an initial concentration of 125 ng/ml. The bottom portion shows signal responses and estimated concentrations for 6 unknown samples on the same plate. The table highlights several limitations of the classical calibration. First, concentration estimates for different dilutions of the same sample can vary substantially. For example, for unknown sample 3, estimated concentrations range from 6.92 ng/ml at a 1/10 dilution to 1233.87 ng/ml at a 1/10 000 dilution. The latter arises from dividing a calibration-based estimate of 0.123 (corresponding to a signal response of 74) by the dilution factor. Averaging such estimates across dilutions can therefore be biased and inefficient. Second, concentrations are not estimable when signal responses fall at the ends of the calibration curve. For example, for unknown sample 16, all three dilutions lead to responses below the lowest concentration of the standards (0.06 ng/ml), resulting in out of range (OOR<) estimates.

Contamination is a concern because allergen samples were collected from settled bed and floor dust, where unknown contaminants may affect the assay. Figure 1 plots log-transformed signal response versus true or estimated Der f 1 concentrations for the standards and three unknown samples from Table 1. The signal response curves for unknown samples 3 and 23 differ from the calibration curve estimated from the standards, with line crossings indicating that no estimated concentration can align the curves. In contrast, unknown sample 10 exhibits a shape similar to the calibration curve. This suggests potential contamination in samples 3 and 23, but not in sample 10.

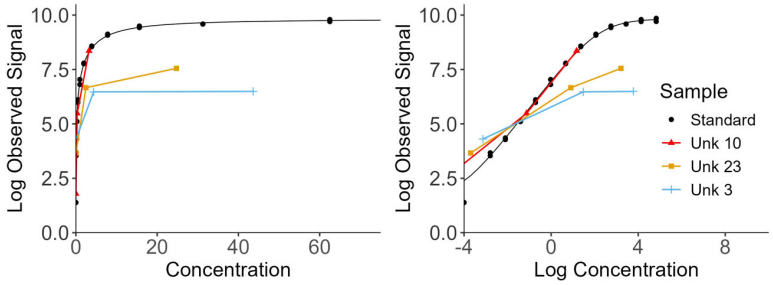

Figure 1 Examples of sample contamination from the New York City Neighborhood Asthma and Allergy Study. The left plot shows log-transformed signal responses read from the machine versus true or estimated concentrations of Der f 1 for the standards and 3 unknown samples from Table 1. To better visualize the lower concentrations, log transformation is applied to the *x*-axis in the plot on the right. The dots represent the data from the standards, overlaid with the calibration curve. The values on the *x*-axis for the three unknown samples were derived from their estimated concentrations using the classical calibration method times the corresponding dilution levels.

3 Bayesian contamination model for serial dilution assays

3.1 Notation and setup

We consider data of a single allergen in a single microtiter plate. Let x denote allergen concentration and y denote signal response to the antibody-antigen interaction, measured as color intensities read from the machine. Define θ_0 to be the known concentration of the undiluted standard sample, θ_j to be the unknown true concentration for new sample j on the same plate, and d_i to be the dilution level of observation i.

Studies have shown that log transformation of the measurements has the effect of regularizing the variance in the measurement error model (Feng et al., 2011). Therefore, we model the log-transformed y instead of y as a function of x in the serial dilution assays. To allow a smooth and increasing dose-response relationship between $\log(y)$ and x, we consider a four-parameter logistic curve model, which is also commonly used in the classical calibration approach:

$$E(\log(y)|x,\boldsymbol{\beta}) = \log(g(x,\boldsymbol{\beta})) = \log\left(\beta_1 + \frac{\beta_2}{1+(x/\beta_3)^{-\beta_4}}\right), \quad 1$$

where β_1 represents the color intensity at zero concentration, β_2 represents the increase to saturation, β_3 represents the concentration where the curve turns, and β_4 represents the rate at which saturation occurs (Higgins et al., 1998).

For observations y_{i0} from the standard sample with known initial concentration θ_0 and known dilution level d_i, we assume a lognormal measurement error model:

$$\log(y_{i0}) \overset{ind}{\sim} \text{normal}(\log(g(\theta_0 d_i, \boldsymbol{\beta})), \sigma_y). \quad 2$$

3.2 Bayesian contamination model

We define contamination in serial dilution assays as a situation in which dose-response curves of some unknown samples differ from that of the standard calibration sample. Our primary inference goal is twofold: to ensure that concentration estimates for uncontaminated samples remain unaffected by contaminated samples, and to flag contaminated samples. Achieving this requires a robust and flexible model capable of accounting for potential contamination. With each unknown sample, we model it using a mixture of two normal measurement error models in the log scale. For observations y_{ij} from sample j with known dilution level d_i but unknown concentration θ_j, the model is

$$\begin{aligned}\log(y_{ij}) \overset{ind}{\sim} &(1-\lambda)\cdot \text{normal}(\log(g(\theta_j d_i, \boldsymbol{\beta})), \sigma_y) \\ &+\lambda \cdot \text{normal}(\log(g(\theta_j d_i, \boldsymbol{\beta}_j)), \sigma_{yj}), \end{aligned} \quad 3$$

where

$$\boldsymbol{\beta_j} = (\beta_1, \beta_{2j}, \beta_3, \beta_4), \quad \beta_{2j} = \beta_2 e^{\delta_{\beta_{2j}}}, \quad \sigma_{yj} = \sigma_y e^{\delta_{\sigma_j}}. \quad 4$$

The first mixture component models the unknown sample using the same measurement error model as the standard data. The second mixture component allows the unknown sample to have different β_2 and σ_y in the measurement error model than that for the standard sample. The intuition for only allowing β_2 to be different from the standard sample is that β_2 controls the increase to saturation, which in the motivating data explains the most difference between the standard calibration curve and the signal response curve of contaminated samples. Also, y_{ij} tends to show more variation among contaminated samples than the standard and uncontaminated samples. We let the variance of the second mixture component, σ_{yj}, to be larger than the variance of the standard sample, σ_y. If we believe that most of the unknown samples are uncontaminated, we can assign an informative prior to λ with a mean close to 0, for example $\lambda \sim \text{beta}(1, 10)$.

3.3 Prior specification

In our setting, all four components of $\boldsymbol{\beta}$ must be positive, so we consider weakly informative priors: $\log(\boldsymbol{\beta}) \sim \text{normal}(0, 10)$. For σ_y, we use a positively constrained Gaussian prior to ensure support on the positive real line while maintaining a proper, weak informative, and interpretable specification with computational convenience: $\sigma_y \sim \text{normal}^+(0, 10)$. To model contamination, we assign a weakly informative normal prior for $\delta_{\beta_{2j}} \sim \text{normal}(0, 1)$ and a weakly informative exponential prior for $\delta_{\sigma_j} \sim \text{exponential}(1)$, reflecting the assumption that contaminated samples have larger variance than uncontaminated ones. If contaminated samples may instead have larger or smaller variance, a weakly informative normal prior can be used for δ_{σ_j}.

We assign hierarchical exponential priors for the concentrations θ across different unknown samples j and let the model estimate the hyperparameter from the data using a weakly informative hyperprior. The exponential prior is consistent with the possibility of zero or effectively zero concentrations for some samples while allowing high concentrations for others:

$$\theta_j \sim \text{exponential}(\mu), \quad \mu \sim \text{normal}^{+}(0, 0.1). \tag{5}$$

3.4 Alternative model specification

We assume weakly informative hierarchical exponential priors for θ_j so that we can model both large concentrations and the lower positive end well. Alternative prior distributions for θ_j include less diffuse exponential distributions, t distributions on the log scale, horseshoe distribution, and skewed-normal distributions (Azzalini, 1985; Carvalho et al., 2010). For the error distribution, we assume a normal error model, with a t distribution as an alternative to accommodate heavier tails. We further discuss the performance of these alternative priors in the real data application and simulations. For the mixture proportion parameter λ, we assign a beta prior with a mean close to zero to generally represent the situation where most of the unknown samples are uncontaminated. Alternatively, we can modify such prior by changing the values of the shape parameters in the beta distribution to accommodate situations where the contamination proportions are high. Also, remember that the main goal of our contamination model is not to accurately model the underlying unknown contamination mechanism, but to generate potential signals for lab technicians to consider the causes behind the existence of contamination. So it is natural to adapt our contamination model based on specific lab experimental settings and needs. For example, the model could be extended by allowing all the β parameters to be different between the standard sample and unknown samples.

3.5 Computation

We fit the model using the NUTS algorithm in Stan as called from cmdstanr (Carpenter et al., 2017; Hoffman and Gelman, 2014). NUTS is a variation of Hamiltonian Monte Carlo that combines the ideas of Markov chain Monte Carlo and deterministic simulation methods by using the derivatives of the density function being sampled, thus allowing the random walk behavior to move more efficiently to the target distribution. We track mixing of the chains using effective sample size and the $\widehat{R}$ diagnostic (Vehtari et al., 2021).

4 Workflow for analyzing a dust mite allergen immunoassay plate

In this section, we present a workflow for modeling potential contamination to one of the microtiter plates measured in the NYC NAAS for Der f 1. The workflow includes initial model construction, model fitting and evaluation, and concludes with sensitivity analysis and model refinement.

4.1 Initial model construction

At this step, we could treat the components of the Bayesian model using the idea of modular construction, for example, starting from something simple and then generalizing and expanding the components when needed. In practice, we need to choose our initial model based on our understanding of the applied problem and the data. We start with a Bayesian model for serial dilution assays with y as the raw observation from the machine, x as the concentration, and $\boldsymbol{\beta}$ as the dose-response curve parameters (Gelman et al., 2004):

$$\begin{aligned} E(y|x, \boldsymbol{\beta}) = g(x, \boldsymbol{\beta}) &= \beta_1 + \frac{\beta_2}{1 + (x/\beta_3)^{-\beta_4}}, \\ y &\overset{ind}{\sim} \text{normal}\left(g(x, \boldsymbol{\beta}), \left(\frac{g(x, \boldsymbol{\beta})}{A}\right)^{\alpha} \sigma_y\right), \end{aligned} \tag{6}$$

where A is a regularizing constant set based on the scale of the data.

We sequentially adapt and improve the base model (6). First, we have observed data skewness, a log transformation to the data would make the inference more interpretable and have the effect of regularizing the variance in the measurement error model (Feng et al., 2011). In order to make our parameters interpretable in the sense of shape and rate for the dose-response curve, we perform the corresponding log transformations for the curve parameters. Also, since we switch to the log scale, the original multiplicative variance component now becomes linear and simpler, and we name it the intermediate model:

$$\begin{aligned} E(\log(y)|x, \boldsymbol{\beta}) = \log(g(x, \boldsymbol{\beta})) &= \log\left(\beta_1 + \frac{\beta_2}{1 + (x/\beta_3)^{-\beta_4}}\right) \\ \log(y) &\overset{ind}{\sim} \text{normal}(\log(g(x, \boldsymbol{\beta})), \sigma_y). \end{aligned} \tag{7}$$

We then extend the intermediate model by adding an additional mixture component to account for potential sample contamination, yielding the final model (3) in Section 3.2.

To follow the idea of modular construction, we put flexible models and weakly informative priors as placeholders for future modification if we gain some insights later in the workflow. In Figure 2, we can see that the model fitting results are getting much better as we go from the initial model (6) to the intermediate model (7) and to the final model (3). For the standard calibration sample, our initial model underestimates the variance component, whereas the intermediate model overestimates it. Sample 23 is highly suspected to be contaminated based on our discussion in Figure 1 and its fit improves noticeably as the model improves. The initial model fits the data poorly, particularly for sample 23. Applying a log transformation to y dramatically improves the model fit. Further allowing potentially contaminated samples to have distinct response curves leads to estimated curves for the standard samples and Sample 23 that better match the observed data and produce smaller, more appropriate variance estimates. The classical calibration approach would produce a poor estimate for Sample 23 because it estimates concentrations by inverting the standard curve.

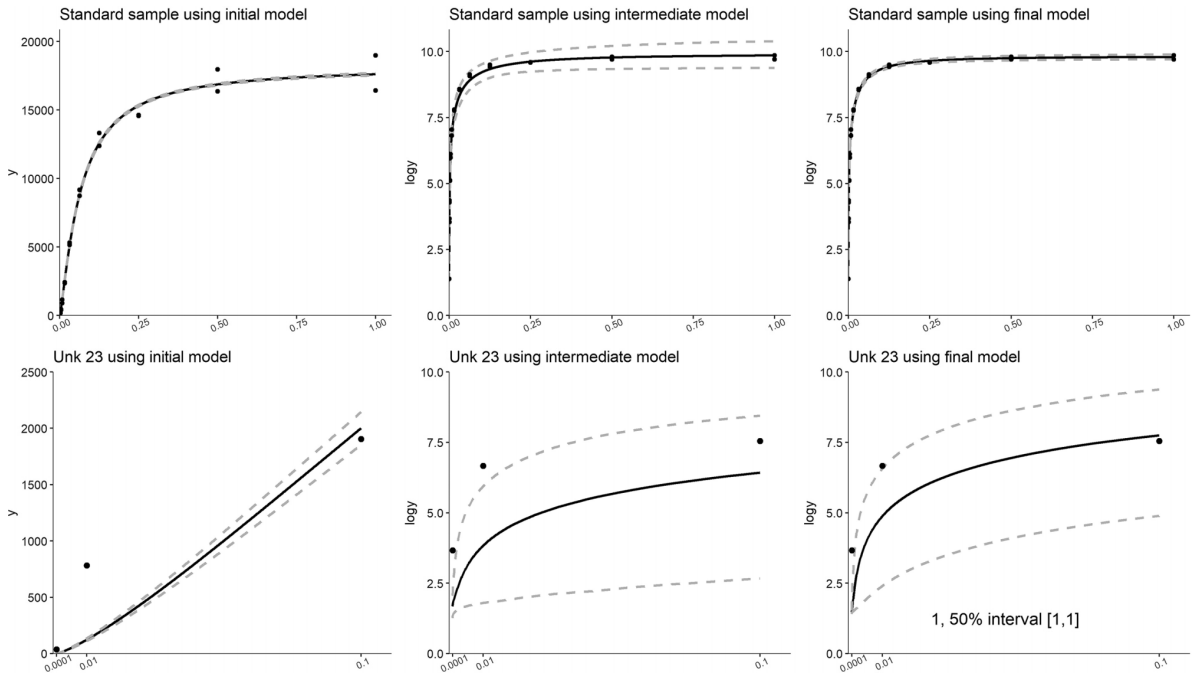

Figure 2 Demonstration of sequential model improvement under Bayesian workflow. From left to right, the initial model uses the basic four-parameter logistic curve; the intermediate model applies a log transformation to stabilize variance; and the final model adds a mixture component to allow for potential model contamination.

4.2 Model fitting while addressing computation issues

We fit our Bayesian models using HMC in Stan and validate the computation by examining convergence diagnostics and computation time. Default Stan settings are used for adaptation and warm-up. The average value of $\widehat{R}$ diagnostic across all parameters is close to 1.001, indicating good chain mixing, and the average effective sample size exceeds 2,000 with no divergent transitions observed, suggesting efficient sampling. Model fitting is computationally efficient, with the final model completing in only a few seconds, enabling scalability to larger datasets. While computational issues can pose challenges in practice, Gelman et al. (2020) present a comprehensive overview of methods for addressing such problems.

4.3 Model evaluation and specific aims for immunoassays data

For model evaluation, we first examine graphical summaries of model fit in Figure 3. The top three plots display zoomed-in estimated dose-response curves for the standards, with log-transformed y on the y-axis and dilution levels on the x-axis. Unknown samples are ordered by the upper bounds of the 95% probability intervals of their response curves. For each unknown sample, we plot the three observed log-scale measurements, the fitted four-parameter logistic curve based on the posterior median of the β parameters, and the corresponding Bayesian 95% probability interval of the fitted curves. We also report the estimated probability of contamination for each sample, shown as the posterior median with 50% probability intervals. Compared to the classical calibration method, which reads off the estimation using the same standard data for all new samples, our Bayesian model allows each new sample to have its own dose-response curve if it is identified by the model to be a contaminated sample. Rather than yielding unstable concentration estimation by averaging highly variable dilution-based estimates, our method fits the observed data well with enough uncertainty to cover the potential variation and contamination. For this plate, the model estimates the probability of contamination to be 1, with the upper and lower bounds of the 50% probability interval both being 1 in the unknown samples 17, 23, 9, 3, and 16. This may suggest contamination in these new samples. Figure 3 also shows that some of the dilution measurements do not align well with the corresponding estimated response curve. Our model estimates a wider 95% probability interval to allow all the observed measurements to fall inside the estimated lower and upper bounds of the curve.

Figure 4 shows the estimated concentrations of the 23 unknown samples, with the posterior median (denoted using dots) and 95% probability interval using the new Bayesian contamination model as well as the point estimate using the weighted classical calibration method (Cheung et al., 2015) (denoted using triangle). The comparison to the conventional classical calibration method is shown in Web Figure 2. For unknown samples 17, 23, and 3, which have high probabilities of contamination, the Bayesian model estimates a different con-

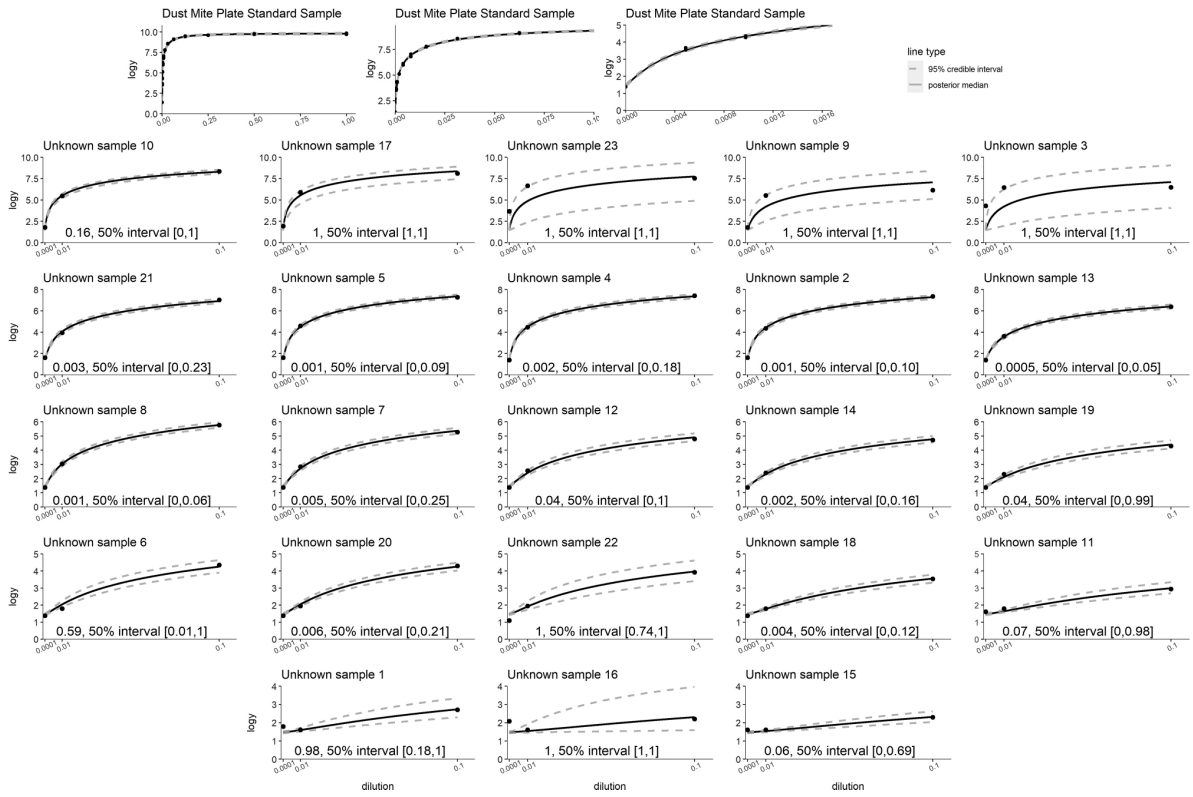

Figure 3 Posterior median and 95% probability interval of the mean dose-response curves for the standards and each new sample estimated using our proposed model. The posterior median and corresponding 50% probability interval for the probability of contamination are shown at the bottom of each plot.

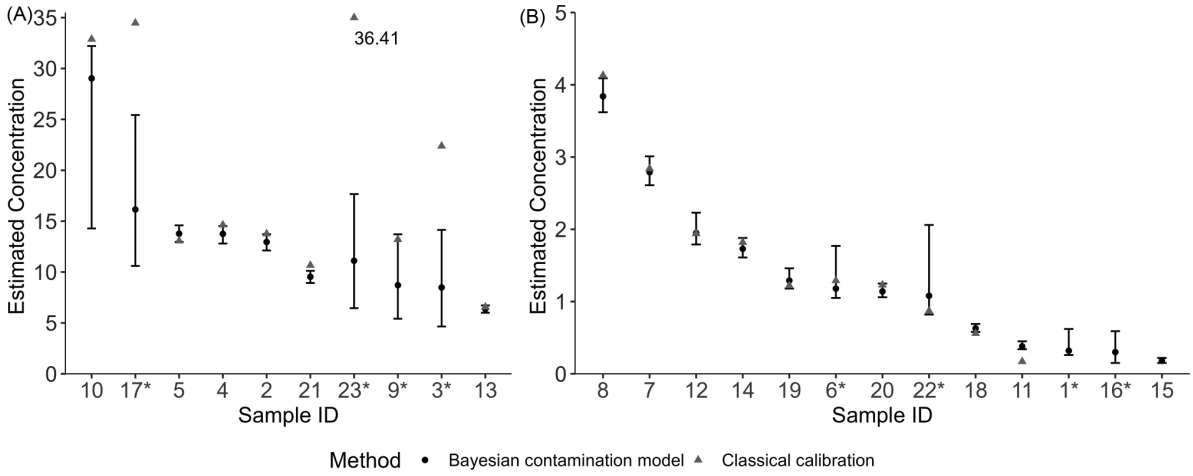

Figure 4 Concentration estimation of Der f 1 for all new samples in a single microtiter plate using the weighted classical calibration and our proposed method with 95% probability interval. Samples with ∗ are estimated to have moderate to high posterior median probabilities of contamination.

centration than using the weighted or conventional classical calibration approaches and a wide 95% probability interval associated with the Bayesian estimate. Unknown samples 1, 16, and 15 are classified as below the LOD using classical calibration, but our Bayesian model provides estimates with very low concentrations.

In real-life lab measurements, experienced technicians may use some unwritten decision rules to check the sample plate

quality. For example, we have multiple measurements for the standard calibration data with known initial concentration and known dilution factors on each plate. When lab technicians apply the conventional classical calibration method to such plates, they usually generate the estimated concentration for each dilution level of the standard calibration sample and compare it to the theoretically expected concentration value at that dilution level. If the estimated concentration is far from the expected concentration, that observation of the standard calibration sample may be discarded. Following this practice, we compare the ratio of estimated and expected concentrations between our proposed method and the classical calibration method (see Web Figure 3). We find that the classical calibration method could not estimate the first two dilutions accurately, but our proposed method performs well for all the dilutions of the standard sample. This further demonstrates the advantage of our method compared to the classical calibration estimation method commonly used in serial dilution lab experiments.

4.4 Sensitivity analysis

To assess the influence of prior information, we have run a sensitivity analysis by refitting the model using various priors mentioned in Section 3.4. In the Web Figures 4 and 5, we show model fitting results analogous to those in Figure 3 but use a stronger exponential prior on the underlying concentration θ_j and a t_4 error model, respectively. The resulting model fits were largely unchanged, except that the estimated probabilities of contamination increase slightly for a few samples under the stronger exponential prior and decrease when the t_4 error model is used. Overall, these results indicate that our model is robust to prior specifications and that even weakly informative priors provide reliable model inference.

Figure 3 shows that, for unknown samples 1 and 16, the signal response at the 1/10 000 dilution is higher than that at the 1/100 dilution. At very low concentration levels, the response curve is essentially flat, and small differences are overwhelmed by measurement noise. While most samples in the last three rows of Figure 3 have response values around 1.5 at the 1/10 000 dilution, unknown sample 22 has a response close to 1, and unknown samples 1 and 16 have responses close to 2. These three samples also have posterior median probabilities of contamination close to 1. To further investigate this behavior, we conduct a sensitivity analysis by manually adjusting these three response values to 1.5, aligning them with the other samples. The results are shown in Web Figure 6. After this adjustment, the posterior median probabilities of contamination decrease to values near 0. This suggests that the high probabilities of contamination on these three samples are driven by measurement error in the responses at the 1/10 000 dilution and that the sensitivity analysis provides insight into why the estimated response curves deviate from the calibration curve.

4.5 Model modification and extension

For model modification, we could easily modify our model with more informative priors justified by experienced lab technicians and also have the ability to incorporate more lab data as well as additional information associated with each observation. For example, if we have another layer of information regarding the sample collection process, such as the relative geographical information of each sample, we could extend our model by modeling the concentration of each sample given that additional information. Besides, our contamination model specification is also pretty flexible, and it could be generalized to allow more complicated contamination models if necessary. If in some other studies there is clear evidence showing that the remaining parameters in the dose-response curve should also vary for the contaminated samples or a different parametric form of contamination exists, we could easily adapt and extend our proposed contamination model to reflect such new information.

5 Simulation studies

We conduct extensive simulation studies to evaluate the performance of the proposed Bayesian contamination model relative to existing approaches. The simulations are designed to closely mimic the NYC NAAS plate analyzed in Section 4, using the same standard concentrations, dilution schemes, and four-parameter logistic response curve calibrated to the real data. Across all scenarios, we compare the proposed model with the base Bayesian model in (6) and the weighted and conventional classical calibration methods.

We consider a range of settings, including varying degrees of contamination among unknown samples and extreme cases with very small or zero concentrations. Performance is assessed using root mean squared error (RMSE) and coverage probability of 95% probability intervals; for the classical calibration methods, which provide only point estimates, evaluation is based on RMSE alone. We additionally conduct prior predictive checks to assess the adequacy of the data-generating model and the robustness of the proposed approach to model misspecification.

Overall, the proposed model consistently achieves lower RMSE and substantially improves coverage compared to competing methods, particularly in the presence of moderate to severe contamination. The model effectively identifies contaminated samples and yields more stable and accurate concentration estimates. In extreme scenarios, the method continues to perform favorably except for samples with very low concentrations, where all methods exhibit increased variability.

Full details of the simulation design, implementation, and results are provided in Web Appendix B and Web Tables 2–5.

6 Wet-lab validation study

To validate our simulation results, we conduct a wet-lab experiment using an assay plate for the Der f 1 allergen. The plate design mirrors our motivating NAAS dataset, but uses an ELISA assay rather than NAAS's multiplex assay, shifting the signal response from fluorescence intensity to optical density and reducing costs. The plate includes two replicates of standards with an initial concentration of 125 ng/mL and 12

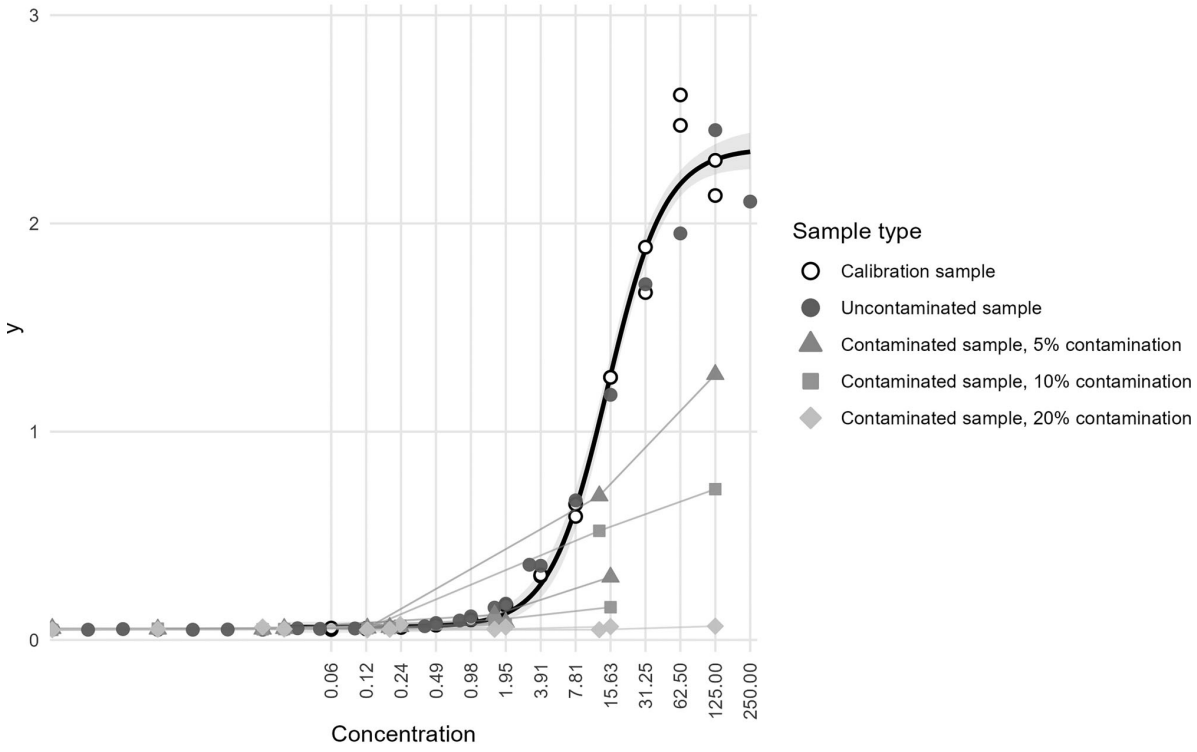

Figure 5 Results of the wet-lab validation study. Calibration samples are shown as open circles, with the fitted calibration curve overlaid. Measurements from diluted samples originating from the same sample are connected by gray lines. Contaminated samples exhibit flatter response curves compared to the calibration curve.

two-fold serial dilutions, as well as 11 uncontaminated and 12 contaminated unknown samples, each measured at three dilution levels (1/10, 1/100, and 1/10 000).

The uncontaminated samples have concentrations of 250, 125, 62.5, 31.25, 15.62, 7.81, 3.91, 1.95, 0.98, 0.49, and 0.24 ng/mL at the first 1/10 dilution level. The contaminated samples have concentrations of 125, 15.62, 1.95, and 0.24 ng/mL at the first dilution level (corresponding to the undiluted concentrations of 1250 ng/mL, etc.) and are generated under three contamination levels (5%, 10%, and 20%) using a household upholstery and fabric cleaner. Although the true concentrations of these samples are known, we treat them as unknown samples for validation purposes.

We apply the Bayesian contamination model to data from this experimental plate. Because signal responses at the upper end of the concentration range are highly variable, we include three uncontaminated unknown samples with concentrations of 250, 125, and 62.5 ng/mL as additional standards to improve calibration curve fitting. Results are shown in Figure 5. Uncontaminated samples lie close to the fitted calibration curve, whereas contaminated samples exhibit responses below the curve. Consistent with Web Figure 1, samples with higher contamination levels display flatter response curves.

Web Table 6 shows that our model correctly classifies all uncontaminated samples as uncontaminated, with posterior probabilities of contamination near zero, and that the 95% posterior intervals for concentrations cover the true values. It also correctly identifies contaminated samples with an initial concentration of 1250 ng/mL (125 ng/mL at the first dilution) under 5% and 10% contamination, assigning posterior median contamination probabilities of one with both the lower and upper bounds of the 50% probability intervals equal to 1.

However, the remaining contaminated samples are classified as uncontaminated, with posterior medians near zero. For the sample with 20% contamination, an extreme and unlikely scenario, the response curve is nearly flat, producing very low signal intensities even at 125 ng/mL; when shifted horizontally, these measurements closely resemble those of an uncontaminated sample with very low concentration. A similar pattern is observed for contaminated samples with first dilution concentrations of 15.62, 1.95, and 0.24 ng/mL under 5% and 10% contamination. In these cases, the observed signal responses are also low, and horizontally shifting the measurements to the left closely aligns them with the fitted calibration curve. Under our definition of contamination, such samples are therefore classified as uncontaminated by the model.

Overall, this validation study confirms that our model effectively detects samples whose response curves deviate substantially from the calibration curve. However, it may fail to identify contaminated samples that exhibit extremely high contamination, leading to very flat response curves, or

contaminated samples with low concentrations at the first dilution.

7 Discussion

Statistical inference under potential data contamination is an essential topic in both the studies of immunoassays and other applied statistics problems. In this paper, we propose a Bayesian contamination model for serial dilution assay data. The proposed model flags potentially contaminated samples while providing reliable estimates for uncontaminated ones, with both empirical analyses and simulation studies demonstrating clear advantages over competing methods. We further develop a general Bayesian workflow for inference under contamination, outlining its key steps and tailoring them to the specific objectives of immunoassay data. This framework highlights essential steps of Bayesian inference in broader applied settings and facilitates users with different backgrounds to quickly adapt our work to their studies.

We define contamination as response curves that deviate from the calibration curve, and under this definition our model can effectively identify contaminated samples. However, there are two scenarios in which the model may misclassify an unknown sample. First, when responses at all three dilution levels are low, the model may classify the sample as uncontaminated with a low concentration because the fitted response curve aligns with the lower end of the calibration curve. This can occur either due to extremely high contamination levels that produce a nearly flat response curve or because the sample has a genuinely low concentration at the first dilution. Addressing this issue would require additional data, such as calibration data for contaminated samples across varying contamination levels, obtained either from the same plate or external plates. Second, when the initial concentration is low, the 1/10 000 dilution may yield near-zero concentrations, where measurement noise can distort the fitted response curve and lead to misclassification as contaminated. In such cases, using a smaller dilution factor would produce more reliable measurements.

Our study has a few limitations. First, the contamination model assumes only a subset of mean response curve parameters to differ between calibration and potentially contaminated samples, motivated by observed laboratory data in which contamination primarily affects the saturation level. This might not be comprehensive enough for all types of contamination happening in the sample collection and measurement process. Second, while the proposed Bayesian workflow effectively identifies contaminated samples in both real and simulated data, it does not estimate the contamination level or correctly estimate the concentrations of contaminated samples. Incorporating additional calibration data on contaminated samples could provide insights into contamination severity. Third, each unknown sample in the motivating dataset is measured at only three dilution levels. Using more dilutions could improve response curve estimation and contamination detection, but would increase experimental cost and may produce measurements below detection limits at extreme dilutions. Quantifying the relative information contributed by each dilution could improve experimental efficiency, for example by enabling sample-specific dilution designs rather than fixed dilution schemes such as 1/10, 1/100, and 1/10 000. This represents an important direction for future research.

Our method currently only considers modeling one serial dilution assay plate and one type of allergen at a time. Multiplex assays could simultaneously measure multiple allergens in a sample using a single plate. Because measurement errors of multiple allergens from the same plate can be correlated, all the allergens should be modeled jointly using Bayesian hierarchical models. In future work, we are going to extend our proposed Bayesian contamination model for a single allergen to allow joint modeling of multiple allergens in a single plate.

Acknowledgments

We thank the Editor, Associate Editor, and the two referees for their thoughtful and constructive comments.

Supplementary Materials

Supplementary material is available at *Biometrics* online.

Web Appendices, Tables, and Figures referenced in Sections 1– 2 and 4–6 and code are available with this paper at the Biometrics website on Oxford Academic.

Funding

This project was supported by NIH/NIEHS awards R01ES035784 and P30ES009089.

Conflict of Interest

None declared.

Data Availability

The wet-lab validation data that support the findings in this paper are available from the openICPSR repository at: https://www.openicpsr.org/openicpsr/project/249765/ (Chen, 2026).

REFERENCES

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171–178.

Box, G. E. P. and Tiao, G. (1968). A Bayesian approach to some outlier problems. *Biometrika*, 55, 119–129.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M. and (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76, 1.

Carter, M., Perzanowski, M., Raymond, A. and Platts-Mills, T. (2001). Home intervention in the treatment of asthma among inner-city children. *Journal of Allergy and Clinical Immunology*, 108, 732–737.

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97, 465–480.

Chen, Q. (2026). Benchmark ELISA serial dilution dataset with controlled contamination. Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

Cheung, Y. B., Xu, Y., Remarque, E. J. and Milligan, P. (2015). Statistical estimation of antibody concentration using multiple dilutions. *Journal of Immunological Methods*, 417, 115–123.

Engvall, E. and Perlmann, P. (1972). Enzyme-linked immunosorbent assay, Elisa. 3. Quantitation of specific antibodies by enzyme-labeled anti-immunoglobulin in antigen-coated tubes. *Journal of Immunology*, 109, 129–135.

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

Finney, D. J. (1976). Radioligand assay. *Biometrics.*, 32, 721–740.

Fong, Y., Wakefield, J., De Rosa, S. and Frahm, N. (2012). A robust Bayesian random effects model for nonlinear calibration problems. *Biometrics*, 68, 1103–1112.

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

Gelman, A., Vehtari, A., McElreath, R., Simpson, D., Margossian C., C., Yao, Y. et al. (2026). Bayesian workflow. Chapman & Hall.

Guo, Y., Harel, O. and Little, R. J. (2010). How well quantified is the limit of quantification? *Epidemiology (Cambridge, Mass.)*, 21, S10–S16.

Hamilton, M. A. and Rinaldi, M. G. (1988). Descriptive statistical analyses of serial dilution data. *Statistics in Medicine*, 7, 535–544.

Higgins, K. M., Davidian, M., Chew, G. and Burge, H. (1998). The effect of serial dilution error on calibration inference in immunoassay. *Biometrics*, 54, 19.

Hoffman, M. D. and Gelman, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15, 1593–1623.

Klauenberg, K., Walzel, M., Ebert, B. and Elster, C. (2015). Informative prior distributions for ELISA analyses. *Biostatistics (Oxford, England)*, 16, 454–464.

Lee, M.-L. T. and Whitmore, G. A. (1999). Statistical inference for serial dilution assay data. *Biometrics*, 55, 1215–1220.

Morales, K. H., Ibrahim, J. G., Chen, C.-J. and Ryan, L. M. (2006). Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. *Journal of the American Statistical Association*, 101, 9–17.

Olmedo, O., Goldstein, I. F., Acosta, L., Divjan, A., Rundle, A., Chew, G. and (2011). Home intervention in the treatment of asthma among inner-city children. *Journal of Allergy and Clinical Immunology*, 128, 284–292 e7.

Perzanowski, M. S., Chew, G. L., Divjan, A., Johnson, A., Goldstein, I. F., Garfinkel, R. S. and (2008). Cat ownership is a risk factor for the development of anti-cat IgE but not current wheeze at age 5 years in an inner-city cohort. *Journal of Allergy and Clinical Immunology*, 121, 1047–1052.

Phipatanakul, W., Eggleston, P., Wright, E. and Wood, R. (2020). Mouse allergen. II. The relationship of mouse allergen exposure to mouse sensitization and asthma morbidity in inner-city children with asthma. *Journal of Allergy and Clinical Immunology*, 106, 1075–1080.

Prentice, R. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics*, 32, 761–768.

Racine-Poon, A., Weihs, C. and Smith, A. F. M. (1991). Estimation of relative potency with sequential dilution errors in radioimmunoassay. *Biometrics*, 47, 1235.

Van Weemen, B. and Schuurs, A. (1971). Immunoassay using antigen-enzyme conjugates. *FEBS Letters*, 15, 232–236.

Vashist, S. K. and Luong, J. H.T. (2018). *Handbook of Immunoassay Technologies: Approaches, Performances, and Applications*. 1–18, London: Academic Press.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16, 667–718.

Xu, Y., Milligan, P., Remarque, E. J. and Cheung, Y. B. (2016). Implementing weighted-average estimation of substance concentration using multiple dilutions. *Stata Journal*, 16, 316–330.

Received: 10 May 2025. **Revised:** 12 June 2026. **Accepted:** 3 July 2026