

# Bayesian Analysis of Serial Dilution Assays

Andrew Gelman,<sup>1,\*</sup> Ginger L. Chew,<sup>2</sup> and Michael Shnaidman<sup>1</sup>

<sup>1</sup>Department of Statistics, Columbia University, New York 10027, U.S.A.

<sup>2</sup>Department of Environmental Health, Columbia University, New York 10032, U.S.A.

\**email*: gelman@stat.columbia.edu

**SUMMARY.** In a serial dilution assay, the concentration of a compound is estimated by combining measurements of several different dilutions of an unknown sample. The relation between concentration and measurement is nonlinear and heteroscedastic, and so it is not appropriate to weight these measurements equally. In the standard existing approach for analysis of these data, a large proportion of the measurements are discarded as being above or below detection limits. We present a Bayesian method for jointly estimating the calibration curve and the unknown concentrations using all the data. Compared to the existing method, our estimates have much lower standard errors and give estimates even when all the measurements are outside the “detection limits.” We evaluate our method empirically using laboratory data on cockroach allergens measured in house dust samples. Our estimates are much more accurate than those obtained using the usual approach. In addition, we develop a method for determining the “effective weight” attached to each measurement, based on a local linearization of the estimated model. The effective weight can give insight into the information conveyed by each data point and suggests potential improvements in design of serial dilution experiments.

**KEY WORDS:** Assay; Bayesian inference; Detection limit; Elisa; Measurement error models; Serial dilution; Weighted average.

## 1. Introduction

### 1.1 *Serial Dilution Assays*

A common design for estimating the concentrations of compounds in biological samples is the serial dilution assay, in which measurements are taken at several different dilutions of a sample, giving several opportunities for an accurate measurement. Currently, serial dilution is a standard tool in the fields of toxicology and immunology. Our experience is in enzyme-linked immunosorbent assays (Elisa) of allergens in house dust samples.

Assays are performed using microtiter plates (for example, see Table 1) that contain two sorts of data: *unknowns*, which are the samples to be measured and their dilutions; and *standards*, which are dilutions of a known compound, used to calibrate the measurements. Figure 1 shows data of measurements versus dilutions from a single plate (assays of the cockroach allergen Bla g1), for the standards and each of 10 unknown samples (which in this case were house dust collected from inner-city apartments). The estimation of the curves relating dilutions to measurements is described in Section 3 of the article. The 10 unknown concentrations are estimated so that the measurements line up with the calibration curve.

Recent formulations of dilution assays appear in Finney (1976), Hamilton and Rinaldi (1988), Racine-Poon, Weihs, and Smith (1991), Higgins et al. (1998), and Lee and Whitmore (1999). Giltinan and Davidian (1994) and Davidian

and Giltinan (1995) present a simulation study suggesting potential improvements using Bayesian methods, and Dellaportas and Stephens (1995) describe Bayesian computations for a model with a single unknown concentration. We continue these ideas here, setting up a hierarchical model including variation among compounds and plates and validating with two sets of experimental data.

This article develops a Bayesian method for estimating concentrations of unknown samples in serial dilution assays. In Section 1.2 we describe a problem with the currently used estimation method, which is used in numerous laboratories across the country and worldwide. Section 2 presents our model, which is based on those of Racine-Poon et al. (1991), Giltinan and Davidian (1994), and Higgins et al. (1998). Section 3 explains how to use Bayesian inference to obtain estimates and uncertainties for the different sources of variation and for the unknown concentrations in the assay, illustrating with a re-analysis of existing data. Having developed the new method, in Section 4 we test it against the existing approach using a laboratory experiment in which different samples are diluted by known amounts, and then we see which method performs better at estimating the true dilutions. Section 5 presents a statistical method, based on linearization of the calibration curve, to estimate the amount of information provided by each measurement in our estimate. We conclude in Section 6 with suggestions about implementation of the new method and the implications for assay designs.

**Table 1**

Typical setup of a plate with 96 wells for a serial dilution assay. The first two columns are dilutions of “standards” with known concentrations, and the other columns are 10 different “unknowns.” The goal of the assay is to estimate the concentrations of the unknowns, using the standards as calibration.

Std	Std	Unk 1	Unk 2	Unk 3	Unk 4	Unk 5	Unk 6	Unk 7	Unk 8	Unk 9	Unk 10
1	1	1	1	1	1	1	1	1	1	1	1
1/2	1/2	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3
1/4	1/4	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9
1/8	1/8	1/27	1/27	1/27	1/27	1/27	1/27	1/27	1/27	1/27	1/27
1/16	1/16	1	1	1	1	1	1	1	1	1	1
1/32	1/32	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3
1/64	1/64	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9
0	0	1/27	1/27	1/27	1/27	1/27	1/27	1/27	1/27	1/27	1/27

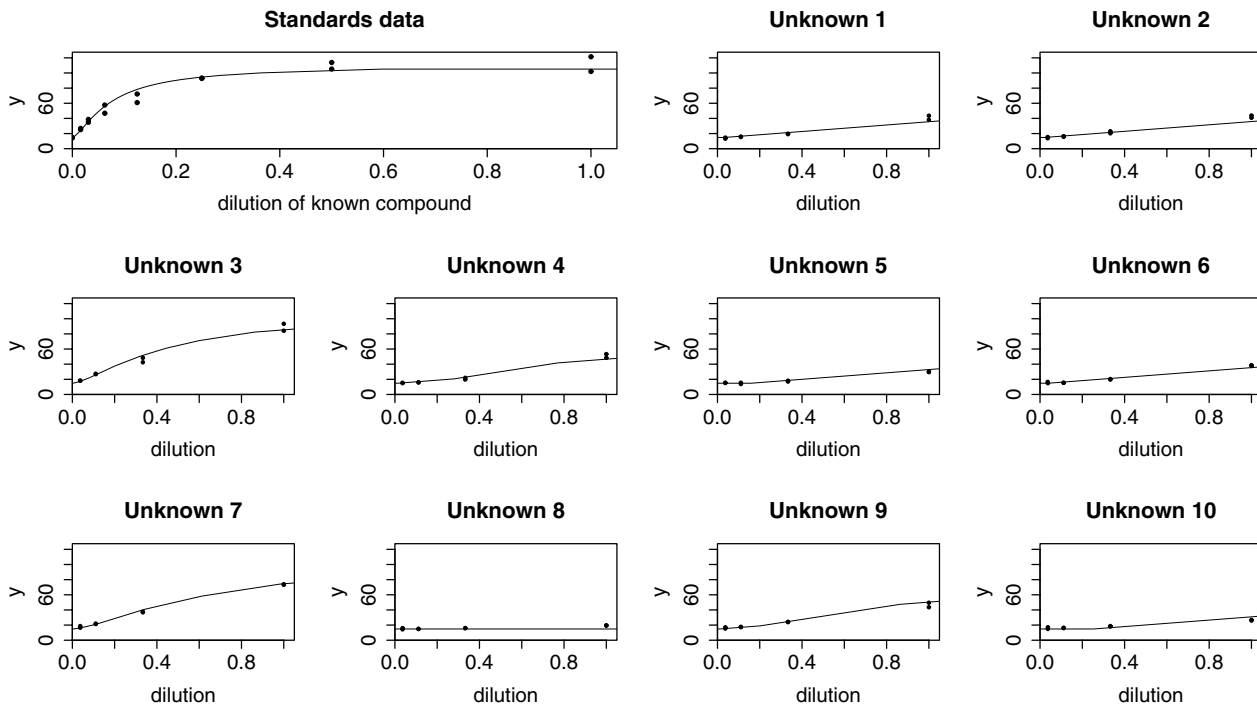
1.2 Difficulties with the Current Method of Estimation

The usual approach to analysis of dilution assays, as implemented in widely used commercial software (Molecular Devices, 2002) follows two steps. First, the standards data are used to estimate the curve relating concentrations to measurements—typically assumed to be a four-parameter logistic function—using least squares. Second, this estimated curve is used to read off the concentration that corresponds to each of the measurements of the unknowns. Estimates of diluted samples are scaled back to the original scale, and these are averaged to obtain an estimated concentration for each unknown sample.

The first step is not a problem; the four parameters of the curve can generally be estimated accurately using least

squares, given the amount of standards data typically supplied on an assay plate. It is possible to estimate from multiple plates together and pool information, but the usual approach, estimating from one plate at a time, works reasonably well.

Unfortunately, the second step—estimating the unknown concentrations—presents serious difficulties. In reading concentrations directly off a curve, the standard method ignores measurement error, which is particularly serious for very high measurements, where the curve is flat. Furthermore, the equal averaging of estimates is inefficient since measurements of highly diluted samples will have greater variance (e.g., the estimated concentration of a 1/27 dilution is multiplied by 27, which scales up its estimation error accordingly). The usual way these problems are handled is by simply discarding



**Figure 1.** Data from a single plate of a serial dilution assay. The large graph shows the calibration data, and the 10 small graphs show the data for the unknown compounds. The goal of the analysis is to figure out how to scale the *x*-axes of the unknowns so they will line up with the curve estimated from the standards. (The curves shown on these graphs are estimated from the model as described in Section 3.2.)

**Table 2**

Example of some measurements  $y$  from a plate as analyzed by the standard software used for dilution assays. The standards data are used to estimate the calibration curve, which is then used to estimate the unknown concentrations. The measurements indicated by asterisks are labeled as “below detection limit.” However, information is present in these low observations, as can be seen by noting the decreasing pattern of the measurements from dilutions 1 to 1/3 to 1/9.

Standards data			Some of the unknowns data			
Conc.	Dilution	$y$	Sample	Dilution	$y$	Est. conc.
0.64	1	101.8	Unknown 8	1	19.2	*
0.64	1	121.4		1	19.5	*
0.32	1/2	105.2		1/3	16.1	*
0.32	1/2	114.1		1/3	15.8	*
0.16	1/4	92.7		1/9	14.9	*
0.16	1/4	93.3		1/9	14.8	*
0.08	1/8	72.4	Unknown 9	1/27	14.3	*
0.08	1/8	61.1		1/27	16.0	*
0.04	1/16	57.6	1	49.6	0.040	
0.04	1/16	50.0	1	43.8	0.031	
0.02	1/32	38.5	1/3	24.0	0.005	
0.02	1/32	35.1	1/3	24.1	0.005	
0.01	1/64	26.6	1/9	17.3	*	
0.01	1/64	25.0	1/9	17.6	*	
0	0	14.7	1/27	15.6	*	
0	0	14.2	1/27	17.1	*	

measurements that are above or below detection limits, which are defined based on the measurements of the standards.

Table 2 illustrates the difficulties with the current method of estimating unknown concentrations. The left part of the figure shows standards data (corresponding to the first graph in Figure 1): The two initial samples have known concentrations of 0.64, with each followed by several dilutions and a zero measurement. The right part of Table 2 shows, for 2 of the 10 unknowns on the plate, the measurements  $y$ , and corresponding concentration estimates as estimated from the fitted curve.

All the estimates for unknown 8 are shown by asterisks, indicating that they were recorded as “below detection limit,” and the standard computer program for analyzing these data gives no estimate at all. A casual glance at the data (see the plot of unknown 8 in Figure 1) might suggest that these data are indeed all noise, but a careful look at the numbers reveals that the measurements decline consistently from concentrations of 1 to 1/3 to 1/9, with only the final dilutions apparently lost in the noise (in that the measurements at 1/27 are no lower than at 1/9). A clear signal is present for the first six measurements.

Unknown 9 shows a better outcome, in which four of the eight measurements are within detection limits. Once again, however, information seems to be present in the lower measurements, which decline consistently with dilution. As can be seen in Figure 1, unknowns 8 and 9 are not extreme cases but rather are somewhat typical of the data from this plate. In measurements of allergens, even low concentrations can be important (e.g., for asthma sufferers) and we need to be able to distinguish between zero concentrations and values that are merely low.

Bayesian inference has the potential to make better use of this information, for two reasons. First, the likelihood function (and thus the posterior distribution) automatically accounts for the greater uncertainty at very low and very high concentrations, without requiring that the extreme data be completely discarded as below or above detection limits. We explore this issue further in Section 5. The second advantage of the Bayesian approach is that it can incorporate several sources of variation without requiring point estimation or linearization, either of which can cause uncertainties to be underestimated in this nonlinear errors-in-variables model (Davidian and Giltinan, 1995; Dellaportas and Stephens, 1995).

## 2. The Model

### 2.1 Curve of Expected Measurements Given Concentration

We use the notation  $x_i$  for concentrations and  $y_i$  for observed color intensities. The expected value of the measurement  $y$  is an increasing function of the concentration  $x$ , and a four-parameter model is typically used (see, e.g., Higgins et al., 1988):

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

where  $\beta_1$  is the color intensity at zero concentration,  $\beta_2$  is the increase to saturation,  $\beta_3$  is the concentration at which the curve turns, and  $\beta_4$  is the rate at which saturation occurs. This model is equivalent to a logistic function of  $\log(x)$ . All four of the  $\beta$  parameters must be positive, and so we model them on the logarithmic scale.

### 2.2 Measurement Error

We follow Higgins et al. (1998) and model the measurement errors as normally distributed with unequal variances:

$$y_i \sim N \left[ g(x_i, \beta), \left( \frac{g(x_i, \beta)}{A} \right)^{2\alpha} \sigma_y^2 \right], \quad (2)$$

where the parameter  $\alpha > 0$  models the pattern that variances are higher for larger measurements (e.g., see Figure 1). The constant  $A$  in (2) is arbitrary and is set to some value in the middle of the range of the data. It is included in the model so that the parameter  $\sigma_y$  has a more direct interpretation as the error standard deviation for a “typical” measurement. We assign a uniform prior distribution to  $\sigma_y$  in the Bayesian analysis.

The model (2) reduces to an equal-variance normal model if  $\alpha = 0$  and approximately corresponds to the equal-variance model on the log scale if  $\alpha = 1$ . In fitting the model, we assign  $\alpha$  a uniform prior distribution in the range  $[0, 2]$ , thus allowing variance relations in a fairly wide range centered at proportionality. Getting the variance relation correct is important here because many of our data are at very low concentrations, and we do not want our model to overstate the precision of these measurements.

### 2.3 Dilution Errors

The dilution process introduces errors in two places: the *initial dilution*, in which a measured amount of a sample is mixed with a measured amount of an inert liquid; and *serial dilutions*, in which a sample is diluted by a fixed factor such as 2 or 3. Higgins et al. (1998) found the initial dilution error to be

detectable but the serial dilution error to be essentially zero (i.e., there was no noticeable autocorrelation in the errors). We found the same, and so we include initial dilution error but not serial dilution error in our model.

We use a normal model on the log scale for the initial dilution error. (Initial dilutions are an issue only for the standards, not for the unknowns, which are typically measured at full strength with no initial dilution.) For any sample that is subject to an initial dilution, we label  $\theta$  as the true concentration of the undiluted sample,  $d^{\text{init}}$  as the (known) initial dilution, and  $x^{\text{init}}$  as the (unknown) concentration of the initial dilution, with

$$\log(x^{\text{init}}) \sim N[\log(d^{\text{init}} \cdot \theta), (\sigma^{\text{init}})^2]. \quad (3)$$

For the 12 samples on a plate such as shown in Table 1, we define  $\theta_0$  as the (known) concentration of the standards (corresponding to the first two columns of the plate) and  $\theta_1, \dots, \theta_{10}$  to be the unknown concentrations that are the estimands of interest.

For the further dilutions, we simply set

$$x_i = d_i \cdot x^{\text{init}}, \quad (4)$$

where  $d_i$  is the dilution of observation  $i$  relative to the initial dilution. (The  $d_i$ 's are the numbers displayed in Table 1.)

#### 2.4 Hierarchical Model for Unknown Concentrations

When using the model to estimate unknown concentrations  $\theta_1, \dots, \theta_J$ , we fit a hierarchical model of the form,

$$\log \theta_j \sim N(\mu_\theta, \sigma_\theta^2), \quad \text{for } j = 1, \dots, J.$$

The purpose of this distribution is to bound the estimates of  $\log \theta_j$  using a proper prior distribution. This is particularly useful when estimating extremely low or high concentrations, for which there would otherwise be no way of bounding  $\theta_j$  away from 0 or  $\infty$ . We assign a diffuse hyperprior density (i.e.,  $\mu \sim N(0, 100^2)$ ,  $\sigma \sim U(0, 100)$ ), so that the hyperparameters are estimated from the data. In cases of extremely poor data (i.e., if all the dilutions for all the unknowns are far above or far below detection limit), the posterior distribution would remain diffuse (with substantial posterior probability associated with extremely high or extremely low concentrations), appropriately indicating the lack of information in the data.

#### 2.5 Diffuse Prior Distribution for the $\beta$ 's for Analyzing Data from One Plate

Serial dilution data are most commonly analyzed a single plate at a time. The parameters  $\beta$  of the calibration curve can be estimated using maximum likelihood (weighted least squares); however, for our goal of estimating concentrations beyond the usual detection limits, we use Bayesian inference (as described in Section 3) to more precisely capture the uncertainties in estimation. We assign diffuse prior distributions for the  $\beta$ 's; for example,  $\beta_k \sim N(0, 100^2)$ ,  $k = 1, 2, 3, 4$ . Given the amount of unknowns data on a typical plate (see Table 1), the posterior distribution of the  $\beta$ 's is well identified by the data. If standards data were sparser on each plate, it would be necessary to include substantive prior information

on the ranges and correlations of the  $\beta_k$ 's in the population, or to analyze data from several plates together, as we discuss next.

#### 2.6 Multilevel Model for Variation of the $\beta$ 's among Plates

The parameters  $\beta_1, \beta_2, \beta_3, \beta_4$  themselves vary, which is why there are standards data on each plate—the unknowns must be calibrated with respect to the plate on which they are measured. In addition, experimental conditions vary from day to day (from factors including lot-to-lot variation among reagents, differences in room temperature, and differences in pipetting techniques among technicians), and these are reflected in changes in the calibration curve. For these reasons, we set up a model allowing the parameters of the calibration curve to vary.

All four parameters  $\beta_k$ ,  $k = 1, 2, 3, 4$ , in model (1) must be positive, and we model them with normal distributions on the logarithmic scale. For each plate  $p$ , processed at day  $t(p)$ , we model  $\log \beta_p = (\log \beta_{p1}, \dots, \log \beta_{p4})$  as,

$$\log \beta_p \sim N[\log b_{t(p)}, \Sigma_\beta^{\text{plate}}] \quad \text{for each plate } p.$$

We similarly apply a multivariate normal model to the variation across days  $t$ :

$$\log b_t \sim N[\log b_0, \Sigma_\beta^{\text{day}}] \quad \text{for each day } t.$$

We assign diffuse hyperprior distributions; e.g.,  $\log b_{0,k} \sim N(0, 100^2)$  for  $k = 1, 2, 3, 4$ , and  $\Sigma_\beta^{\text{plate}} \sim \text{Inv-Wishart}_4(I)$ ,  $\Sigma_\beta^{\text{day}} \sim \text{Inv-Wishart}_4(I)$ . The degrees of freedom for the inverse-Wishart are set as low as possible to maintain a proper distribution (see, e.g., Johnson and Kotz, 1972).

### 3. Estimation of the Model

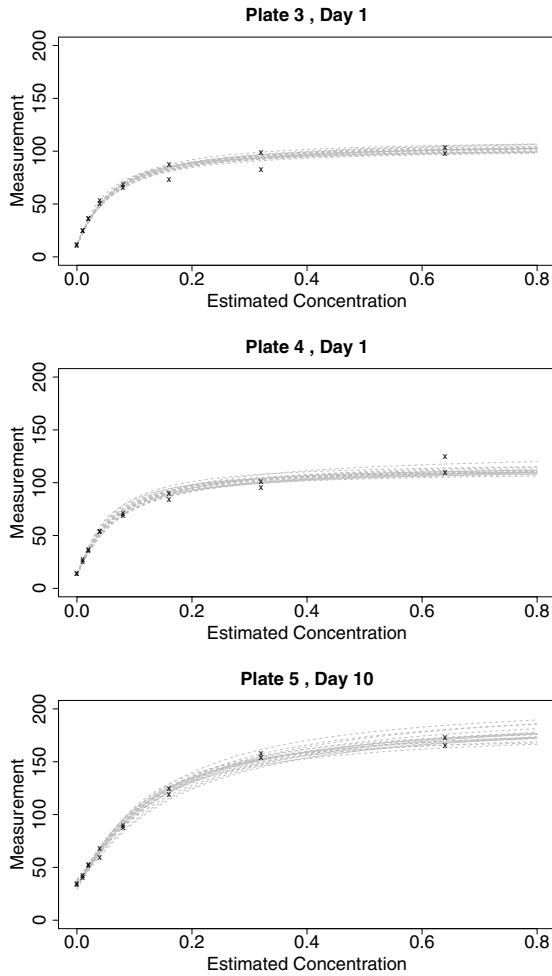
#### 3.1 Estimating the Model Using Standards Data from Several Plates

Before using our model to estimate unknown concentrations, we fit it to standards data in order to check its fit and estimate its hyperparameters with precision. In order to simultaneously identify initial dilution errors in  $x$  and variation of parameters  $\beta$  across plates and days, it is necessary to fit the model to data from several plates, measured at several days with more than one plate per day, and with different initial dilutions.

We estimated the model using standards data from 24 plates, 23 of which were from existing experiments with two columns of unknowns each (as in Table 1) and one of which was a special plate prepared with 10 different initial dilutions of the standard (to allow accurate estimation of the scale  $\sigma^{\text{init}}$  in the model [3] of initial dilution error).

We fit the model (and also the model for a single plate, described in Section 3.2) using the Bugs software for Bayesian inference (Spiegelhalter et al., 1994, 2003), as linked from R (R Project, 2000; Gelman, 2002). We obtained approximate convergence (the potential scale reduction factors of Gelman and Rubin, 1992, were below 1.1 for all parameters) after 150,000 iterations of four parallel chains of the Gibbs sampler. To save memory and computation time, we save every 40th iteration of each chain.

Figure 2 displays a subset of the data used to fit the model, along with the estimated curves and their posterior



**Figure 2.** Standards data  $y_i$  and fitted curves  $E(y_i | x_i)$  versus dilutions for each of three plates selected from the 24 used in the estimation of the hierarchical model. Variation between days is much larger than between plates within a day. (The several curves on each graph represent different random simulation draws of the parameters from the estimated posterior distribution.)

uncertainty. The data, and the curves, vary slightly between plates but much more dramatically between days.

### 3.2 Estimating Unknowns Using Data from a Single Plate

The hierarchical model in Section 3.1 makes sense, but it is usual and convenient to fit the serial dilution model to data from a single plate at a time using a single initial dilution of the standard compound. With only one plate, we simply estimate the parameters  $\beta_1, \beta_2, \beta_3, \beta_4$  without a hierarchical structure, simultaneously with the set of the unknown concentrations.

The model for a single plate can be constructed from (2)–(4). For data points  $i$  from samples  $j(i)$ ,

$$y_i \sim N \left[ g(x_i, \beta), \left( \frac{g(x_i, \beta)}{A} \right)^{2\alpha} \sigma_y^2 \right]$$

$$x_i = d_i \cdot x_{j(i)}^{\text{init}},$$

and for each sample  $j$ ,

$$\log x_j^{\text{init}} \sim N \left[ \log (d_j^{\text{init}} \cdot \theta_j), (\sigma^{\text{init}})^2 \right]$$

for the standard sample,  $j = 0$

$$x_j^{\text{init}} = \theta_j \quad \text{for the unknown samples, } j = 1, \dots, 10.$$

There is initial dilution of the standards, but the unknowns are started at full strength. The concentration  $\theta_0$  and initial dilution  $d_0^{\text{init}}$  for the standard are known, and the 10 unknown concentrations  $\theta_j$  must be estimated.

A design such as displayed in Table 1 with replications of the standards data at a wide range of dilutions allows us to estimate all the model parameters fairly accurately. When fitting the model in Bugs, it is helpful to use reasonable starting points (which can be obtained using crude estimates from the data) and to parameterize in terms of the logarithms of the parameters  $\beta_j$  and the unknown concentrations  $\theta_j$ .

We illustrate with inference for the data displayed in Figure 1. The posterior median estimates of the parameters of the calibration curve are  $\hat{\beta}_1 = 14.8$  (with a posterior 50% interval of [14.7, 15.0]),  $\hat{\beta}_2 = 94.3$ [89.8, 99.0],  $\hat{\beta}_3 = 0.048$ [0.044, 0.052], and  $\hat{\beta}_4 = 1.41$ [1.37, 1.46]. The median estimates define a curve  $g(x, \hat{\beta})$ , which is displayed in the upper-left plot of Figure 1. As expected, the curve goes through the data used to estimate it.

The variance parameters  $\sigma_y$  and  $\alpha$  are estimated at 2.3 and 1.4 (with 50% intervals of [2.2, 2.4] and [1.3, 1.5], respectively). The parameter  $\sigma^{\text{init}}$  was fixed at 0.02, and the scaling factor  $A$  was set to the geometric mean of the standards data from the plate.

The inferences for the unknown concentrations  $\theta_j$ , along with the estimated calibration curve, were used to draw scaled curves for each of the 10 unknowns displayed in Figure 1.

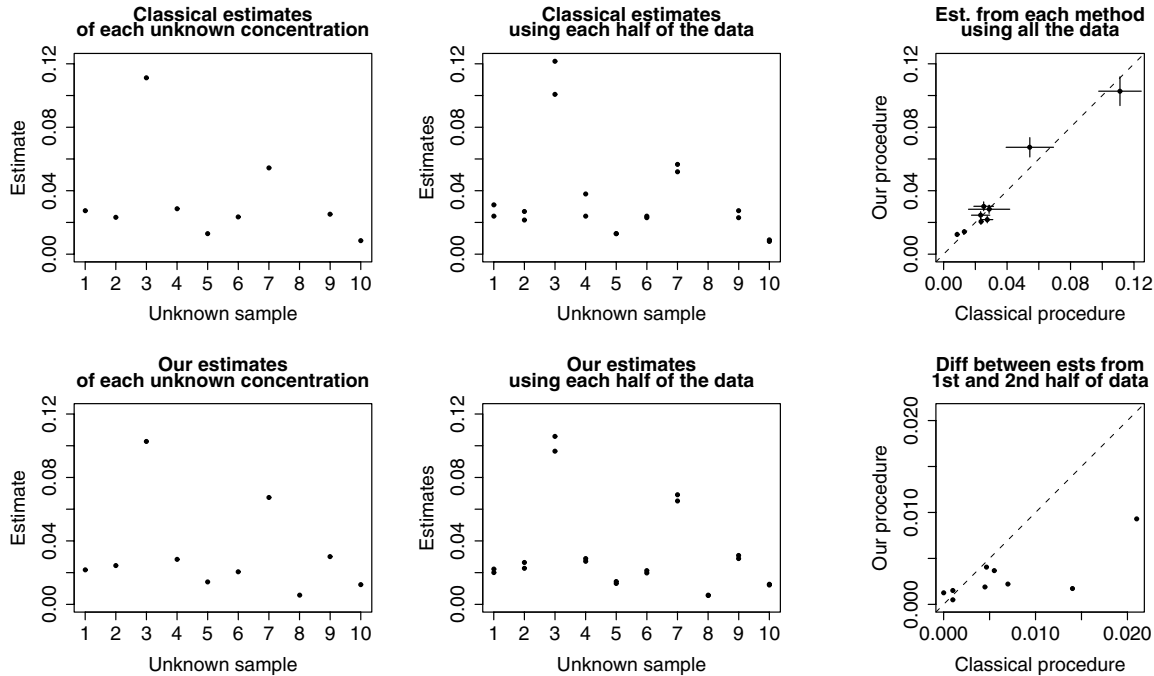
### 3.3 Comparison to the Existing Approach

We can compare our inferences to those obtained from the standard approach of estimating the calibration curve and then transforming each measurement directly to an estimated concentration. For each unknown sample, each estimated concentration is divided by its dilutions, and then the estimates are averaged to obtain a single estimate. For example, for the data displayed in Table 2, the estimated concentration for unknown 9 is  $\frac{1}{4}(0.040 + 0.031 + 3 \times 0.005 + 3 \times 0.005) = 0.025$ . The data from dilutions 1/9 and 1/27 are not used in the estimate since those measurements are below detection limit for this sample.

Figure 3 compares the classical and Bayesian estimates in several ways for the data shown in Figure 1. The leftmost plots in Figure 3 display the estimated concentrations for each of the 10 unknowns. The estimates are generally similar (except that there is no classical estimate for unknown 8 since all its measurements are “below detection limit”).

The middle plots show estimates from each of the two halves of the data (in the setup of Table 1, using only the top four or the bottom four wells for each unknown). For each method, the two estimates are similar, but the consistency is much stronger for our estimate.

Finally, the plots on the right side of Figure 3 compare the two estimates directly. The top plot shows that the two approaches give similar estimates, but ours has lower standard



**Figure 3.** Comparison of classical estimates with our Bayesian procedure for the model fit to data from a single plate. The estimated concentrations are similar under the two methods, but our approach gives smaller standard errors. Our estimates also perform much better in cross-validation: The two estimates using just the first or second set of dilutions for each unknown are much closer under our method than with the classical approach.

errors. The bottom plot displays the absolute difference between the first-half and second-half estimates under each method, and shows that our procedure is consistently more reliable than the classical approach.

This analysis illustrates the potential effectiveness of our approach, but we do not want to make too strong a claim based on data from a single plate. Section 4 describes an experiment specially designed to compare the old and new methods.

**4. A Laboratory Experiment Validating the Method**

*4.1 Design of the Study*

Having developed the inferential method on existing data, we perform a new experiment to check the validity of the Bayesian inferences and compare them to the classical approach. Our validation study involves 10 unknown samples, starting with a sample of very high concentration (extracted from cockroach feces) and then successively diluting it by factors of 4. The unknowns  $\theta_1, \dots, \theta_{10}$  are thus constrained by design to be in the ratio  $1 : \frac{1}{4} : \frac{1}{16} : \dots : \frac{1}{4^9}$ . Each of the 10 samples is measured at dilutions of 1, 1/3, 1/9, and 1/27, and the entire experiment is replicated on a second plate (with the same unknown samples used for both plates). Each plate is aligned as pictured in Table 1, with 2 columns of standards and 10 columns of unknowns. However, for this study we are only using the top half of each column of unknowns. (The bottom half of each plate was used for a different study involving assays of contaminated samples.)

*4.2 Results of the Model Fit*

We perform our evaluation on each of the two plates separately. For each plate, we use the classical method to esti-

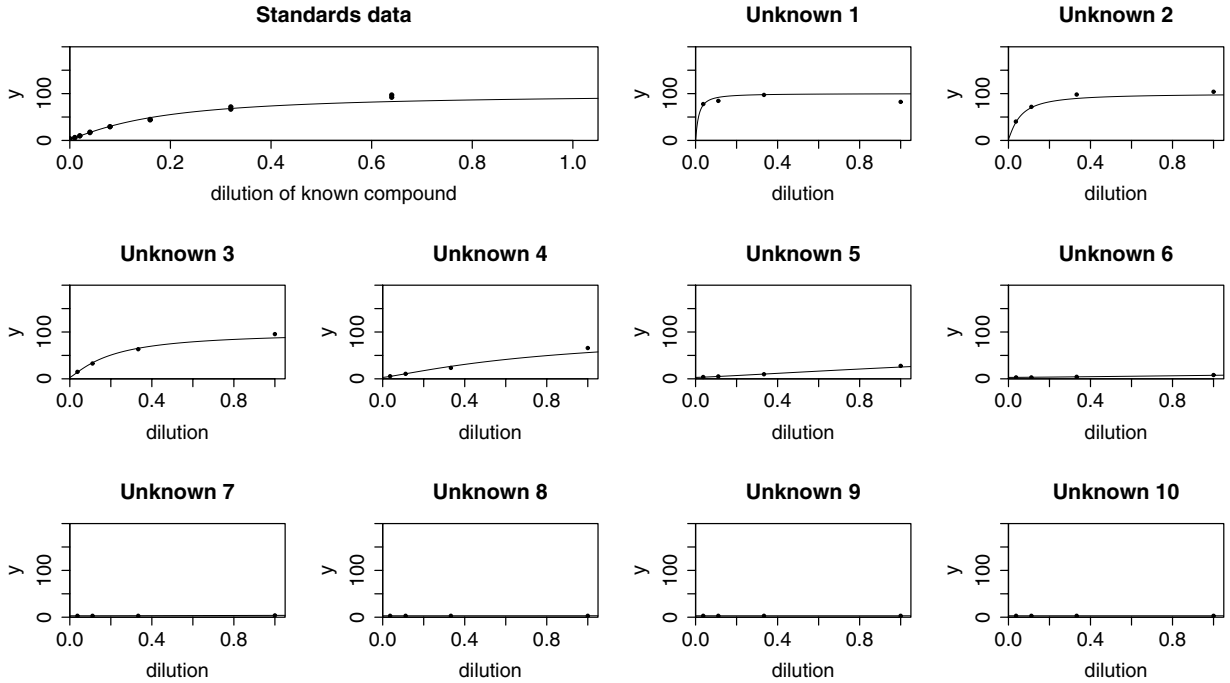
mate the calibration curve and each of the 10 unknown  $\theta_j$ 's, and then we fit the model using Bayesian inference. Figure 4 shows the data and estimated curve for one of the plates. The data from the second plate look similar.

*4.3 Evaluating and Comparing the Classical and Bayes Estimates*

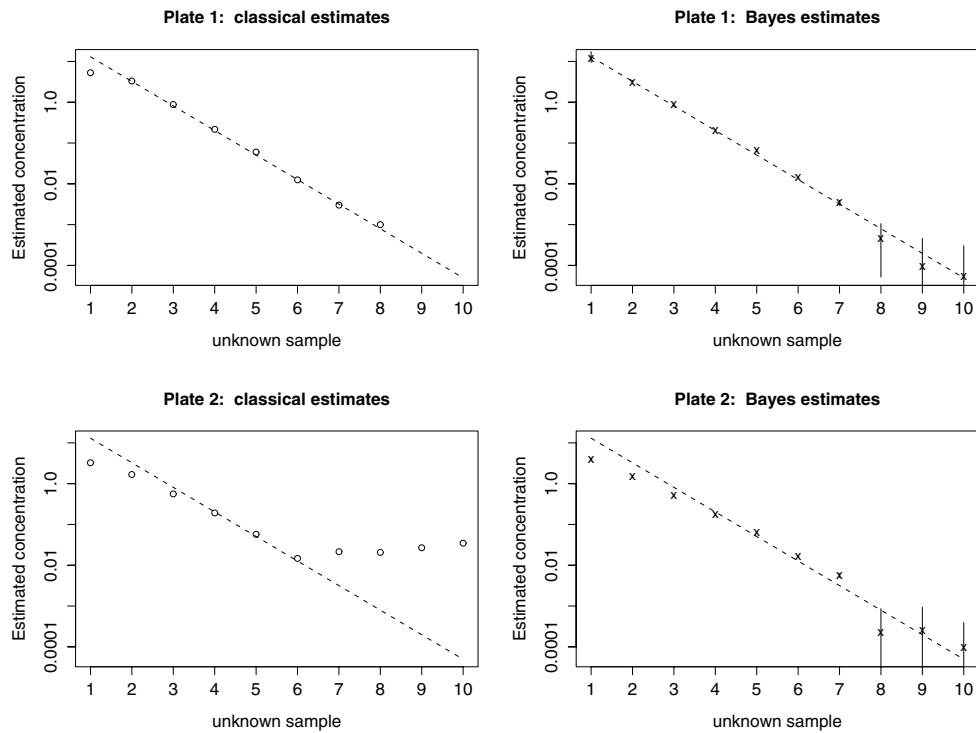
Our experiment was designed to allow simple evaluation of the inferences. If an estimation method is performing perfectly, it should obtain estimates for all 10 unknowns  $\theta_j$ , and the estimates should be in the ratio 1, 1/4, 1/16, etc. That is, a plot of  $\log \hat{\theta}_j$  versus  $j$  should have a slope of  $-\log(4)$ .

Figure 5 displays the results for the two plates in the validation study. In each graph, the dotted line indicates where we would expect the estimates should be—the line has a slope of  $-\log(4)$ , and its intercept is set by fitting it to the classical estimates for unknown samples 3–5 (chosen because their data are in the middle of the calibration curve). We use the classical estimates to set the baseline as a form of conservatism in our comparison: We find the Bayesian inferences to fall closer to the dotted line, even with that line defined based on the classical procedure. To summarize, the slope of the dotted line, indicating the true concentrations of the 10 samples, is known from the design of the experiment but not “known” to the estimation procedures.

Comparing the estimates to the lines, we see that the Bayesian inference does better in each plate. In plate 1, the classical procedure gives no estimate (“below detection limit”) for the last two samples, whereas the Bayesian inferences are reasonable (and, appropriately, have large uncertainties). In plate 2, the classical estimates make no sense for the last four plates (also essentially a detection limit problem), whereas the



**Figure 4.** Data and estimated curve (based on posterior medians of the parameters) from one of the two plates of the validation study. The 10 unknowns start with data that are mostly “above detection limit” and gradually decrease in concentration until all measurements are “below detection limit.”



**Figure 5.** Classical estimates and Bayesian posterior medians for the 10 unknown concentrations from each of the two plates of the validation study. The Bayesian estimates also show 50% error bars. The graph is on a logarithmic scale, and the dotted lines, with slope  $-\log(4)$ , display the pattern that the true concentrations are in the ratio  $1 : \frac{1}{4} : \frac{1}{16} : \dots : \frac{1}{4^9}$ . The Bayesian estimates are closer to the line for a much wider range of concentrations for both plates, with the classical method failing by giving no estimate (as in the last two samples of plate 1) or a highly inaccurate estimate (as in the first and last four samples of plate 2).

Bayesian posterior medians again do a good job of tracking the dotted line.

In addition, plate 2 shows a problem with *both* methods of estimation, in that points fall systematically off the dotted line with slope  $-\log(4)$ , even for the middle samples where the data are strong. We suspect this to be a problem with the model itself, but it is not directly relevant to our goal here of comparing the Bayesian and classical inferences. Related to this problem, the Bayesian 50% intervals for the first seven samples are clearly too narrow.

4.4 Estimating Ratios of Unknowns

Another way to compare the two estimates is to examine their accuracy at estimating the ratio of two concentrations, which is automatically known from the design of the experiment to be a given power of 4. The most challenging assignment is to compare the concentrations of samples 1 and 10, which have a ratio of  $4^9$ . For convenience we examine the logarithms base 4 of the estimated ratio, which should thus equal 9. The classical estimate is undefined from plate 1 (where all four measurements of unknown 10 are below detection limit) and 3.3 from plate 2. The Bayesian posterior medians are 9.1 (with a 50% posterior interval of [7.8, 11.3]) from plate 1 and 7.7 [6.7, 9.2] from plate 2. These intervals are wide (e.g., on the unlogged scale, the two ends of 50% interval estimated from plate 1 differ by a factor of 130) but are still preferable to estimates that are nonexistent or off by more than a factor of 2000.

To look more systematically, the 10 unknowns allow 45 comparisons for each plate, all of which can be compared to the appropriate power of 4. Of this total of 90 comparisons for the two plates, 17 cannot be made because one or both of the classical estimates are undefined, in 47 cases the Bayesian estimate is better (in the sense of the posterior median being closer to the true value on the logarithmic scale), and in 26 cases the classical estimate is better. In those cases where the Bayes estimate was closer, it was by an average of 2.2 on the log scale. In those cases where the classical estimate was closer, the difference was 0.2 on the log scale. This is consistent with the pattern in the lower right of Figure 3, that the Bayesian estimates typically are more reliable than the classical when compared head to head.

4.5 Summary

In the two plates of the validation study, we found the Bayesian posterior medians to be nearly identical to the classical estimates in the range at which the data are strongest (in this particular experiment, samples 2–6 on each plate). When the two estimates differ, the Bayesian estimate is almost always better, sometimes far better. This includes a setting such as in plate 2 where the model is imperfect. Finally, when the classical procedure gives no estimate at all, the Bayesian estimates have wide posterior uncertainties but are still reasonable. This is potentially important for public health studies of allergens and childhood asthma, for which even very low exposures can be dangerous.

5. Information Provided by Each Data Point

An intriguing feature of the Bayesian approach is that it uses all the data, including those previously discarded as “outside

detection limits.” However, all data are not equally informative. Here we describe a method for quantifying this information, which should be helpful for data analysis, comparison of methods, and design of future assays.

5.1 Weighted Averages

The key idea here is the weighted average. Consider several independent measurements  $y_1, \dots, y_n$  of a parameter  $\theta$ , of the form,

$$y_i \sim N(a_i + b_i\theta, \sigma_i^2).$$

These can be linearly transformed to be direct estimates of  $\theta$ :

$$\hat{\theta}_i = \frac{y_i - a_i}{b_i} \sim N(\theta, \sigma_i^2/b_i^2).$$

The least squares estimate of  $y$  (or the Bayes estimate under a uniform prior distribution; see, e.g., Gelman et al., 1995) is then a weighted average of the direct estimates,  $\hat{\theta}_i$ , where the observations have weights,

$$w_i \propto \frac{b_i^2}{\sigma_i^2}.$$

5.2 Equivalent Weights for Nonlinear Models

With a nonlinear model,  $y_i \sim N[f_i(\theta), \sigma_i^2]$ , we can apply the same idea by linearizing  $f$  at the estimated value of  $\theta$ , thus writing  $f_i(\theta) = a_i + b_i(\theta)$ , where  $b_i = f'_i(\theta)$ . The weight for data point  $i$  is then,

$$w_i \propto \frac{(f'_i)^2}{\sigma_i^2}.$$

5.3 Application to Serial Dilution

For the data from a dilution  $d_i$  of an unknown sample  $\theta$ , we can express model (2) as,

$$\begin{aligned} y_i &\sim N[g(x_i, \beta), \sigma_i^2], \\ &= N[g(d_i \cdot \theta, \beta), \sigma_i^2]. \end{aligned} \tag{5}$$

From (5), we can linearize  $g$  as a function of  $\theta$  to obtain weights,

$$w_i \propto \frac{[d_i \cdot g'(x_i)]^2}{\sigma_i^2}, \tag{6}$$

where  $g'(x_i)$  is the derivative of  $g(x, \beta)$  with respect to  $x$ , evaluated at the current estimate of  $x_i$  (i.e.,  $d_i$  multiplied by the estimated  $\theta$ ) and the current estimate of  $\beta$ .

We can express the weights in (6) in a slightly more usable form by expanding the variance  $\sigma_i^2$  from (2). Also, we can ignore factors that are the same for all measurements for a given sample and thus can be absorbed into the proportionality constant. The weights for measurements  $i$  on dilutions  $d_i$  of a single sample with estimated concentration  $\theta$  are then,

$$w_i \propto \left[ \frac{d_i \cdot g'(x_i)}{g(x_i)^\alpha} \right]^2, \tag{7}$$

where the differentiation of  $g$  is taken with respect to  $x$ , and the entire expression is evaluated at the estimated parameter values.



Expression (7) makes sense:

- Smaller dilutions have smaller weights (because the variance is magnified when the low-dilution estimates are scaled back up).
- Measurements at the steeper part of the curve (where  $g'$  is higher) have higher weights.
- Weights are lower for measurements with higher variance.

The weights depend on the unknown parameters  $\theta$ ,  $\beta$ ,  $\alpha$ , and so when we are fitting the model, we compute the set of weights for each of the unknown samples and normalize each set to sum to 1, for each posterior simulation draw. When the simulation is done, we report the set of weights for each sample, averaging over the simulation draws.

5.4 Using the Weights to Understand the Information in Existing Data

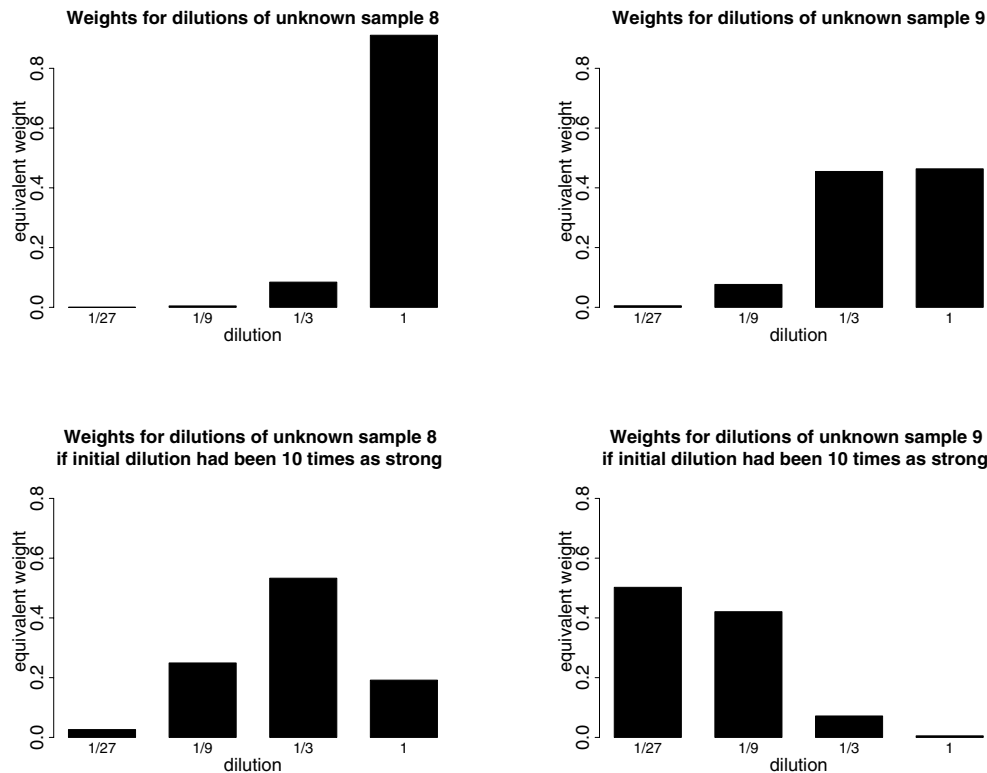
For any given unknown sample, we now have a weight for each measurement, and these can be normalized to sum to 1. The weights depend only on  $x_i$ , not on the measurement  $y_i$ , and so multiple measurements at the same dilution have identical weights. Thus, the measurement array in Table 1 yields four weights for each unknown, corresponding to the dilutions 1, 1/3, 1/9, 1/27. The weights give a sense of the

relative importance of the data at each dilution, and can be compared to the classical procedure, which implicitly assigns equal weights to all measurements within detection limits and zero weights to the others.

The top row in Figure 6 shows the weights for the four dilutions of each of the two unknown samples in Figure 1 whose raw data are given in the right panel of Table 2. For each of these, the initial dilutions have the dominant weighting because the later dilutions are near the low end of the curve. Thus, there is some sense to the classical claim that the lower dilutions of these data are “below detection limit.” However, there is no need to make a yes/no declaration on detection, since the Bayesian method automatically downweights these low-end observations. In comparison, the classical approach inefficiently assigns equal weights to all measurements within detection limits.

5.5 Using the Weights to Consider Alternative Designs

We can use the weights as a tool to understand the model and consider alternative designs. For example, suppose that the original samples 8 and 9 had been 10 times as concentrated. In that case, the data would have come further up on the calibration curve, and the data at lower dilutions would become more informative. The bottom row in Figure 6 shows the new weights that would apply for these more concentrated



**Figure 6.** Top row: Equivalent weights for the data at each dilution, for two of the unknown samples with data displayed in Figure 1 and Table 2. For each of these two samples, about 70% of the information is supplied by the initial dilution, with most of the rest given by the 1/3 dilution. In comparison, the classical method gives equal weight for all data within detection limits; it thus gives no estimate for sample 8 and weights of 0, 0, 0.5, 0.5, for the four dilutions of sample 9 (see the right column of data in Table 2). Bottom row: Equivalent weights for the data at each dilution for unknowns 8 and 9, under the hypothetical scenario that the initial concentration for each sample had been 10 times as strong.

samples. In any case, the weights are for helping us understand the estimate, not for the actual estimation of the concentration  $\theta$ , which is obtained using Bayesian inference as described in Section 3.2.

## 6. Discussion

### 6.1 Performance of Bayesian Inference Compared to the Existing Estimation Method

We have demonstrated that, for the purpose of estimating unknown concentrations with serial dilution, a Bayesian formulation of the standard four-parameter logistic model outperforms the currently standard approach based on inverting an estimated curve. The Bayesian estimates can be performed using the free software packages R and Bugs. We have also programmed the inference using the Metropolis algorithm (see, e.g., Gilks, Richardson, and Spiegelhalter, 1996) directly in R.

Section 3.2 shows how Bayesian methods allow estimation of unknown concentrations with reasonable accuracy using much less data than needed by conventional methods, even with measurements that at first appear to be outside “detection limits.” In Section 4, we show that it is possible to obtain accurate estimates of concentrations varying by a factor of  $4^6$  or  $4^7$  (that is, ranging from sample 1 to sample 7 or 8 in Figure 5) and reasonable estimates over the entire range of  $4^9$ , or 260,000.

It is hardly surprising that a Bayesian or likelihood approach works better than an inversion procedure that ignores estimation uncertainty. However, previous statistical treatments of serial dilution assay (see Hamilton and Rinaldi, 1988; Racine-Poon et al., 1991; Higgins et al., 1998; Lee and Whitmore, 1999) have focused on the estimation of the calibration curve or inference from discrete data rather than the problem considered here, of inference for several unknown concentrations from continuous assays. Dellaportas and Stephens (1995) consider a fully Bayesian approach but with a slightly simpler model in a nonhierarchical setting.

### 6.2 Information Provided by Each Data Point

In a serial dilution assay, the amount of information given by each measurement depends upon its position along the calibration curve and on its dilution (see [6]), as derived in Section 5. Figure 6 shows how the weights for a given set of measurements can be displayed to give insight into the estimation of each unknown sample, and how these weights can inform considerations of alternative designs. By comparison, the existing procedure gives zero weights to observations outside “detection limits” and equal weights to the remaining measurements.

Displays of the Bayesian weights, as in Figure 6, may help users in understanding the information present in serial dilution data and deciding what dilutions to use in successive assays of similar items (e.g., dust samples from several different homes).

### 6.3 Further Work

The model can potentially be improved in various ways, most notably by generalizing the function (1) of expected measurements. It is perhaps most important that the model be accurate at the extremes of very low measurements, in order to

get reasonable estimates for samples “below detection limit.” Inferences for these low measurements are sensitive to the assumed power-law variance relation in (2).

A more serious problem occurs if the function relating measurements to concentration is different for standards and unknowns. We have seen some of this in the systematically biased estimates for the unknown concentrations in plate 2 in the validation study (see Figure 5). Unfortunately, impurities in the sample can affect the assay so that the curve estimated from the standards is not appropriate for the unknowns, which calls into question the whole structure of the calibration process. This represents a problem with both classical and Bayesian estimates, and we suspect it is the reason why estimates from dilution assays are in practice much more variable than would be suggested by even the classical estimates in Figure 3. An important direction of future research is to study which aspects of the curves vary between samples and which are stable, to allow the possibility of more accurate calibration.

Another research direction is in the design of the assays, improving from the existing design shown in Table 1, with standards diluted from 1 to 1/64 and unknowns from 1 to 1/27. Would it be better to have the two sets of standards at slightly different values (e.g., with initial dilutions of 1/2 and 1/3, rather than both 1/2) instead of pure replications? Would it be better to have standards below 1/64 to better capture the behavior at very low levels? Should the initial dilution be set at a higher concentration so that the upper limit of the curve is estimated more accurately? Similarly, how many dilutions are recommended for each unknown sample, and how many samples per plate, balancing the goals of efficiency in estimation with that of measuring more items?

By yielding more accurate estimates and quantifying inferential uncertainties (especially in the cases previously deemed outside detection limits), the Bayesian approach sets the stage for more systematic studies of model and design innovations, which we hope will lead to an even broader extension of the range of concentrations to which assays can be applied. This is very much in the original spirit of serial dilution assays, which can generically be considered as an approach to extending the useful dynamic range of a measurement process.

## ACKNOWLEDGEMENTS

We thank Wenbin Lu for help in developing the model, Mark Hansen and the reviewers for helpful comments, and the National Institutes of Health and National Science Foundation for financial support through grants 1 R01 ES10922-01A1 and SES-0084368.

## RÉSUMÉ

Dans les essais de dilution en cascade la concentration d'un composé est estimée en combinant des mesures faites sur plusieurs dilutions d'une solution mère de titre inconnu. La relation entre le titre et les mesures est non linéaire et hétéroscédastique, on ne doit donc pas accorder des poids égaux aux différentes mesures. Pour analyser de telles données on rejette, dans l'approche standard actuelle, une proportion élevée d'observations car au dessus ou au dessous des limites de détection. Nous présentons une méthode bayésienne pour estimer conjointement une courbe de réponse et le titre inconnu en utilisant la totalité des données. Nos estimateurs

ont des erreurs standards beaucoup plus faibles que ceux de la méthode existante et nous obtenons des estimations même quand toutes les observations sont à l'extérieur des "limites de détection." Nous évaluons empiriquement les performances de notre méthode sur des données d'allergènes de blatte mesurés dans des échantillons de poussière de maison. Nos estimations sont bien plus précises que celles de la méthode usuelle. De plus nous développons une méthode pour déterminer le poids "effectif" attaché à chaque observation, à l'aide d'une linéarisation locale du modèle estimé. Les poids effectifs donnent une indication sur l'information apportée par chaque observation et suggèrent d'éventuelles améliorations du dispositif de l'essai.

## REFERENCES

- Davidian, M. and Giltinan, D. (1995). *Nonlinear Models for Repeated Measurement Data*, Chapter 10. London: Chapman and Hall.
- Dellaportas, P. and Stephens, D. A. (1995). Bayesian analysis of errors-in-variables regression models. *Biometrics* **51**, 1085–1095.
- Finney, D. J. (1976). Radioligand assay. *Biometrics* **32**, 721–740.
- Gelman, A. (2002). *Bugs.R.: Functions for running Bugs from R*. Available at [www.stat.columbia.edu/~gelman/bugsR/](http://www.stat.columbia.edu/~gelman/bugsR/).
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457–511.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.
- Gelman, A., Goegebeur, Y., Tuerlinckx, F., and Van Mechelen, I. (2000). Diagnostic checks for discrete-data regression models using posterior predictive simulations. *Applied Statistics* **49**, 247–268.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D., eds. (1996). *Practical Markov Chain Monte Carlo*. London: Chapman and Hall.
- Giltinan, D. and Davidian, M. (1994). Assays for recombinant proteins: A problem in nonlinear calibration. *Statistics in Medicine* **13**, 1165–1179.
- Hamilton, M. A. and Rinaldi, M. G. (1988). Descriptive statistical analyses of serial dilution assay 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**, 19–32.
- Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics*, Volume 4. New York: Wiley.
- Lee, M. L. T. and Whitmore, G. A. (1999). Statistical inference for serial dilution assay data. *Biometrics* **55**, 1215–1220.
- Molecular Devices. (2002). *Softmax Pro 4.3*. Sunnyvale, California.
- R Project. (2000). *The R project for statistical computing*. Available at [www.r-project.org](http://www.r-project.org).
- Racine-Poon, A., Weihs, C., and Smith, A. F. M. (1991). Estimation of relative potency with sequential dilution errors in radioimmunoassay. *Biometrics* **47**, 1235–1246.
- Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., and Lunn, D. (1994, 2003). *BUGS*. MRC Biostatistics Unit, Cambridge, U.K. Available at [www.mrc-bsu.cam.ac.uk/bugs/](http://www.mrc-bsu.cam.ac.uk/bugs/).

Received April 2003. Revised September 2003.

Accepted November 2003.