A Bayesian approach for estimating calibration curves and unknown concentrations in immunoassays

Feng Feng[1,*], Ana Paula Sales[1] and Thomas B. Kepler[1,2,3]

[1]Center for Computational Immunology, Department of Biostatistics and Bioinformatics, Duke University Medical Center, 2424 Erwin Road, Hock Plaza Suite 1102, [2]Department of Immunology and [3]Department of Statistical Science, Duke University, Durham, NC 27705, USA

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Immunoassays are primary diagnostic and research tools throughout the medical and life sciences. The common approach to the processing of immunoassay data involves estimation of the calibration curve followed by inversion of the calibration function to read off the concentration estimates. This approach, however, does not lend itself easily to acceptable estimation of confidence limits on the estimated concentrations. Such estimates must account for uncertainty in the calibration curve as well as uncertainty in the target measurement. Even point estimates can be problematic: because of the non-linearity of calibration curves and error heteroscedasticity, the neglect of components of measurement error can produce significant bias.

Methods: We have developed a Bayesian approach for the estimation of concentrations from immunoassay data that treats the propagation of measurement error appropriately. The method uses Markov Chain Monte Carlo (MCMC) to approximate the posterior distribution of the target concentrations and numerically compute the relevant summary statistics. Software implementing the method is freely available for public use.

Results: The new method was tested on both simulated and experimental datasets with different measurement error models. The method outperformed the common inverse method on samples with large measurement errors. Even in cases with extreme measurements where the common inverse method failed, our approach always generated reasonable estimates for the target concentrations.

Availability: Project name: Baecs; Project home page: www.computationalimmunology.org/utilities/; Operating systems: Linux, MacOS X and Windows; Programming language: C++; License: Free for Academic Use.

Contact: feng.feng@duke.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Immunoassays are measurement techniques based on the interaction between a selective antibody and its target antigen. The basic principles of immunoassays were introduced in 1945 (Landsteiner, 1990), and the first application of immunologically based technology (radioimmunoassay) was reported in the late 1950s (Yalow and Berson, 1959, 1960). Since then, immunoassays have become a primary diagnostic and research tool within the medicine and life sciences. Compared to other analytical methods, immunoassays provide a rapid, simple, economic, highly specific and sensitive analysis with high throughput and adaptability to field use [reviewed in (Harris *et al.*, 1998; Hennion and Barcelo, 1998; Oellerich, 1980)]. Immunoassays do, however, have limitations. These include interference from matrix compounds used to stabilize the samples, cross reactivity to structural analogs and poor scalability for many multi-analyte applications. Several different types of labels have been used in immunoassays, including radiolabels, enzymes, fluorescent, luminescent and phosphorescent labels. The most common labels for clinical and research studies are enzymes and colorimetric substrates.

The approach used most frequently for the analysis of immunoassay data is performed in two steps on two sets of samples, the standards and the target samples. In the first step, standard analyte samples are prepared at known concentrations, and used to estimate a calibration curve, typically by non-linear least squares, relating concentrations to measurements in the system. In the second step, the unknown concentrations are determined by inverting the calibration curve—finding the concentration on the standard curve that produces the same response as that obtained from the unknown sample (Dudley *et al.*, 1985).

The calibration curve itself is only estimated with error: a limited number of samples can be used in an assay, and the true curve must be inferred from a limited number of noisy responses. This task is accomplished by selecting a family of mathematical functions, parameterized by a small number of parameters that is capable of fitting the variety of calibration curves encountered. Previous work has shown that the five-parameter logistic family provides a very good fit under most practical conditions (Gottschalk and Dunn, 2005). The parameters of the function are fitted to assay data, assuming some model for the variance of the measurement error, using non-linear least squares. Contemporary computers handle such computationally intensive algorithms very easily, although issues of convergence to false minima and failure of convergence altogether cannot be ignored. Most fitting algorithms will fail with injudicious initial guesses or in cases where the variance of the measurement error is large or modeled improperly.

Upon reading off unknown concentrations by inverting the calibration model, however, serious challenges can arise.

*To whom correspondence should be addressed.

Table 1. Example experimental data of interferon-β (IFN-β) concentration measured using enzyme-linked immunosorbent assay (ELISA) and analyzed by the common non-linear least squares approach implemented widely by commercial software

Standards data		Unknown sample data			
Concentration (pg/ml)	OD	ID	Dilution	OD	Estimated concentration
1000.0	3.279	1	1	0.340	32.84
1000.0	3.024		1	0.337	32.31
500.0	2.128		1/10	0.160	ND*
500.0	2.279		1/10	0.172	2.58
250.0	1.319				
250.0	1.364	2	1	0.279	22.09
125.0	0.775		1	0.279	22.09
125.0	0.797		1/10	0.148	ND*
62.5	0.506		1/10	0.146	ND*
62.5	0.506				
31.25	0.345	3	1	0.159	ND*
31.25	0.334		1	0.171	2.31
0.00	0.130		1/5	0.138	ND*
0.00	0.152		1/5	0.133	ND*

The estimated five-parameter logistic function is $OD_{450} = 0.16 + \frac{11.3 - 0.16}{(1 + (\frac{conc}{858.8})^{-0.60})^{1.97}}$, which is then used to read off unknown concentrations of samples. The samples containing unknown concentration of IFN-β are measured undiluted and 1:10 diluted, and the optical density values at 450 nm (OD) values are obtained for estimation of concentrations of IFN-β. 'ND*' indicates that OD is below the detection limits, and the concentration is non-detectable by common non-linear least squares and inverse method done using the R software (R Development Core Team, 2006).

The common methods widely implemented by many commercial software products do not account for calibration uncertainty. This neglect can lead to significant underestimation of error in the estimated target concentrations, and to bias in the point estimates, particularly at concentrations at the ends of the calibration range.

Response data in immunoassays are generally heteroscedastic: their variances are not constant across the range of concentrations measured. It is common for the variances of standards at the upper end of the response curve to be 3 or 4 order of magnitude larger than those at the lower end. Heteroscedasticity is typically incorporated into the statistical model by assuming that the variance is proportional to some power of the mean response. Denoting the variance by σ^2 and the mean by μ, we have

$$\sigma^2 = \delta\mu^\theta \tag{1}$$

where δ is a scale parameter and a function of magnitudes of the response, the exponent θ is usually a number between 1.0 and 2.0 (Finney and Phillips, 1977).

Variances in response measurements at the extreme ends of the calibration curve can place sample measurements above or below detection limits; these measurements are typically discarded. Table 1 (and Supplementary Fig. S1) illustrates some of the challenges encountered when using the common inverse method for estimating concentrations of unknown samples. It shows data on interferon β (IFN-β) concentrations measured by enzyme-linked immunosorbent assay (ELISA). The first part is standards data, which were used to fit a five-parameter logistic calibration curve by the non-linear least squares regression. Then the sample measurement data were used to estimate unknown concentrations by inverting the calibration function. The unknown sample data contain three samples of unknown concentration of interferon-β (IFN-β). Each sample was measured either undiluted or 1:10 diluted in duplicates. Undiluted sample 1 and 2 contains about 32 and 22 pg/ml IFN-β, respectively. The 10-time diluted sample 1 and 2 were recorded as 'ND*' (non-detectable) and no estimation of concentration was given, even though samples generated optical density (ODs) larger than the lowest value in the system. In case of sample 3, one replicate of the sample was measured to contain about 2 pg/ml IFN-β, while all other three measurements gave no estimation of concentrations, even though OD values consistently showed decline from undiluted samples to 1/10 diluted ones and were higher than the lowest OD values in the system. As can be seen in our experimental data, sample 3 is not an extreme case, but rather is typical of the data for some cytokines, especially those of important functions such as tumor necrosis factor-α (TNF-α), interferon-γ (IFN-γ), etc. This is also true for other types of analytes. For example, in measurements of allergens or carcinogens, even low concentrations can be important (Gelman *et al.*, 2004). Furthermore, some of 'the low extreme cases' are not biologically low, but rather low relative to the machine or kit reagent detection limits. Moreover, the extreme cases with high or low concentration of target analytes are usually associated with abnormal states of the system, such as diseases or activation/inactivation of cells, and are of great research interests. For example, during a septic shock, the concentration of the pro-inflammatory cytokine interleukin-6 (IL-6) can go up to more than 12 000 pg/ml in the blood (Ng *et al.*, 2003), but the Bio-Plex Pro 23-plex Assay [low photomultiplier tube (PMT) setting] (BioRad) can only detect up to 2100 pm/ml of this cytokine, which is about 6-fold lower (Supplementary Fig. S1). Clearly, a better strategy is necessary to gain more information from those extreme values of samples.

In Bayesian statistical modeling, inference about unknown parameters model is based on the fact that their uncertainty of them is described by the posterior density conditional on observed data. Before data collection, a prior distribution is used to describe parameter uncertainty. After obtaining data, this prior information is combined with information present in the data through the likelihood function and Bayes' rule to obtain the posterior distribution, from which parameter estimate summary statistics are computed.

Bayesian methods have been widely used in many biomedical applications and have proven useful for accounting for multiple sources of uncertainties arising from experimental variation, instrumental noise, etc. In addition, they do not require point estimation or linearization of variations, either of which can underestimate the uncertainties in non-linear models (Davidian and Giltinan, 1995; Dellaportas and Stephens, 1995). They have been employed successfully for data analysis in microarray studies, PCR amplification curves, serial dilution assays and fluorescence chemistry (Conlon *et al.*, 2007; Frigessi *et al.*, 2005; Gelman *et al.*, 2004b; Sivaganesan *et al.*, 2008). In the case of immunoassay data analysis, the likelihood function and posterior distribution automatically incorporate suitable variations arising from the use of multiple hybridization probes of different lengths and composition. For serial dilution assays, the Bayesian approach has been shown to generate reasonable estimates at very low and very high concentrations of the calibration curve, without discarding extreme data as being below or above the detection limits.

In thisarticle, we report the use of a Bayesian approach to generate calibration curves and estimate unknown concentrations in immunoassays such as ELISA and Luminex assays. The Markov Chain Monte Carlo (MCMC) method is used to generate samples from the posterior distribution for the parameters and unknown concentrations jointly. Compared to the common approach, our proposed approach allows (i) an acceptable estimation of uncertainty in the concentration estimates via credible intervals; (ii) a better use of data points with large variances (especially the very low and very high concentrations) without discarding those as being 'beyond detection limits'. The proposed Bayesian approach is validated using both simulated and experimental data. We have also implemented the method in simple-to-use software, which is freely available to the research community.

2 METHODS

2.1 The model

The model relating the expectation of measured values to the underlying concentrations is the five-parameter logistic function family (Gottschalk and Dunn, 2005; Higgins *et al.*, 1998),

$$E(y|x;\boldsymbol{\beta}) = f(x,\boldsymbol{\beta})$$
$$\boldsymbol{\beta} = [A,B,C,D,E]$$
$$f(x,\boldsymbol{\beta}) = B + \frac{A-B}{\left(1+\left(\frac{x}{C}\right)^D\right)^E} \tag{2}$$

where x denotes the unknown sample concentration, $\boldsymbol{\beta}$ is the vector of parameters with components A, B, C, D and E. This model has been shown to improve performance on assays with asymmetric calibration curves, over the simpler and widely used symmetric four-parameter logistic function (Gottschalk and Dunn, 2005).

For the error model, we assume that the log transformation regularizes the variance. This assumption is borne out by analysis of both Luminex and ELISA data for 22 murine cytokines; a constant variance model provides a good approximation with the transformed data. (Supplementary Fig. S2). The error model, therefore, is

$$\log(y_{ij}) \sim N(\log(f(x_i,\beta)),\sigma^2),$$
$$i=1,2,\dots n; \quad j=1,2,\dots k; \tag{3}$$

where y_{ij} is the j-th observed measurement of the i-th sample, x_i is the i-th concentration, $\boldsymbol{\beta}$ is the parameter vector and $f(x_i, \boldsymbol{\beta})$ is the five-parameter logistic function as in Equation (2), σ^2 is the constant random variance, n is the number of calibration points and k is the number of replicates in each calibration point.

The following diffuse prior distributions are used:

$$A,B,C \sim \text{Log-}N(0,10^2),$$
$$-D \sim \text{Log-}N(0,10),$$
$$E \sim \text{Log-}N(0,10),$$
$$\sigma \sim \text{Log-}N(0,1), \tag{4}$$

the log-normal distribution (Log-N) is used here to keep parameters A, B and C positive. D is a negative number usually larger than -1. When E is 1, the five-parameter logistic function becomes the four-parameter function, which is a good model for many symmetric calibration curves. The variances for the log-normal are set so to make the prior distribution diffuse/non-informative. Other diffuse distributions (such uniform or inverse gamma distribution) can be used as well. Given the amount of standards data in a typical immunoassay, the posterior distributions of these parameters can be well identified.

2.2 Estimation by MCMC

We use MCMC to draw samples approximating those drawn from the true posterior distribution, and in particular use the Metropolis–Hastings algorithm implemented as follows.

(1) Choose starting values $\boldsymbol{\beta}_0 = [A_0, B_0, C_0, D_0, E_0, \sigma_0]$ and x_0.
Starting values are set by rough estimation from the standards data. A_0 is set to be the maximum observed measurement, B_0 to be the minimum one and C_0 to be the mean concentration value from sample data (Gottschalk and Dunn, 2005), D_0 is a negative small number (-1.0) and E_0 and σ_0 are positive small numbers (1.0 and 0.1, respectively). x_0 is selected using crude estimates based on observed measurements and standards data.

(2) Based on current parameters from previous steps, generate new candidate parameters, $\boldsymbol{\beta}^* = [A^*, B^*, C^*, D^*, E^*, \sigma^*]$ and x^*, from appropriate proposal distributions.
A simple random-walk algorithm is used to generate new candidate parameters from the normal distribution centered on the current value of β as the proposal distribution. To generate new candidate parameters,

$$\boldsymbol{\beta}* = \boldsymbol{\beta} + \Delta\boldsymbol{\beta} \text{ and } \Delta\boldsymbol{\beta} \sim N(0,\varepsilon_\beta^2)$$
$$x* = x + \Delta x \text{ and } \Delta x \sim N(0,\varepsilon_x^2) \tag{5}$$

where ε_β^2 and ε_x^2 are proposal variances. An important practical issue in this step is the scaling of the random walk density. If the proposal variance of the algorithm is too large, most steps will move into tails of the distribution and will be rejected. On the other hand, if the variance is too small, the autocorrelation in the sequence of accepted draws will be high and convergence to the target distribution will be slow. So in our implementation, small MCMC trials with different values of ε are applied to search for an appropriate proposal variance until the acceptance rate is approximately 15~50% (Roberts and Rosenthal, 2001).

(3) Given the candidate parameters $\boldsymbol{\beta}^*$ and x^*, calculate the ratio of the likelihoods of the proposed parameters and current ones.

$$\alpha = p(\boldsymbol{\beta}^*, x^*|y, Xs, Ys)/p(\boldsymbol{\beta}, x|y, Xs, Ys)$$
$$p(\beta, x|y, Xs, Ys) \propto K\sigma^{-2n+k} e^{-\frac{\sum_{Ys,y}(\log y_i - \log \hat{y}_i)^2}{2\sigma^2}}$$
$$\hat{y}_i = f(x_i,\beta) = B + \frac{A-B}{\left(1+\left(\frac{xi}{C}\right)^D\right)^E},$$
$$i=1,2,\dots n+k. \tag{6}$$

where Xs and Ys are calibration data, y is the response measurement whose concentration is to be estimated, K is a constant, n is the number of calibration data points, k is the number of replicates for the response measurement.

(4) If $\alpha > 1$, accept $\boldsymbol{\beta}^*$ and x^*; if $\alpha < 1$, accept $\boldsymbol{\beta}^*$ and x^* with probability α, else reject it.

(5) Return to step 2.

Following a sufficient burn-in period, the chain approaches its stationary distribution and samples from the MCMC closely approximate ones from the posterior distribution of parameters and unknown concentrations. Finally, the posterior means and independent 95% credible intervals from the posterior marginal densities are computed.

2.3 Implementation

The software implementation of the proposed approach was developed in C++ for linux and MacOS X and C# for Windows. The source code and executables for Windows, Linux and MacOS X are available for free
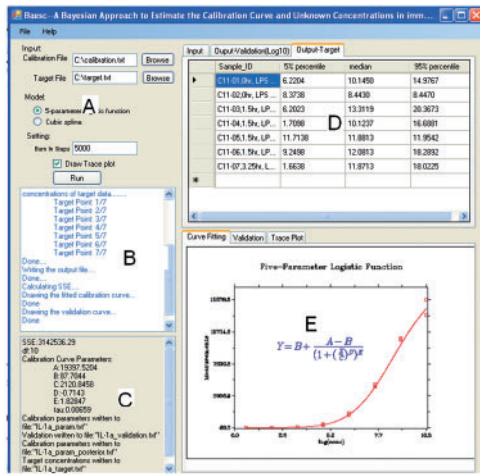
Fig. 1. Screen Shot of Baesc software. (**A**) The input control. The input data are taken in as tab-delimited text files. The calibration file contains standards data and the target file contains sample measurement data used for estimation of unknown concentrations. (**B**) The program log. The log shows current activities of the program. (**C**) The output log. It shows the output results of the curve fitting and estimation of unknown concentrations. (**D**) The data display table. It shows the input and output data in a data table format. (**E**) The graph panel. It displays the calibration curve fitted using the Bayesian approach, the validation curve with 95% Bayesian confidence interval (BCI) and trace plot for the parameter (see more in Supplementary Fig. S6).

distribution. The Windows version of the program has a simple and intuitive graphical interface. Figure 1 shows a screenshot of the software, Baesc, and illustrates its various components. The input of the program is through data files. The fitted calibration curve as well as the diagnostic trace plot and validation plot are displayed on the interface. The results are written to files with parameters and estimated concentrations. The Linux and Macintosh version of the program are run in command line and generate output files allowing further analysis such as MCMC diagnostics, hypothesis testing and the fitting of models for biological processes that regulate the concentration of the assayed analytes.

3 RESULTS

3.1 Simulated data

In order to validate the approach and test the software implementation, we first ran tests on simulated data. Two types of simulated data following either the equal variance or unequal variance model were generated.

To simulate data with equal variance, we followed the model below,

$$E(y_i) = B + \frac{A - B}{\left(1 + \left(\frac{x_i}{c}\right)^D\right)^E} \quad i = 1, 2, \dots n;$$

$$\log(y_{ij}) \sim N(\log(E(y_i)), \sigma^2) \quad j = 1, 2, \dots k. \tag{7}$$

where σ was a constant, $[A, B, C, D, E] = [70000, 20, 113459.8, -1.79, 0.476]$, $n = 8$, and $k = 4$. The parameters were chosen based on experimental data of IL-1α cytokine measured in a Luminex assay in our lab (see Supplementary Table S1). Ten concentrations were selected starting at 1.0 to represent levels typical for the assay
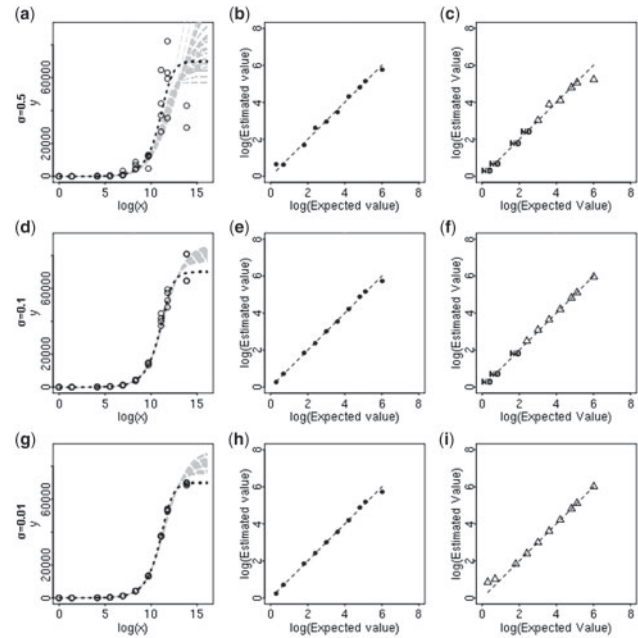

Fig. 2. Validation of the Bayesian approach and common inverse approach on simulated data with equal variances. The simulated data were generated following the five-parameter logistic model $y = 20 + \frac{70000 - 20}{\left(1 + \left(\frac{x}{113459}\right)^{-1.79}\right)^{0.476}}$ with equal variances. Eight concentrations starting at 1.0 were used and four replicated measurements were simulated for each concentration. Different variances (σ^2) were used to compare the performance of two approaches. (**a**), (**b**) and (**c**), $\sigma = 0.5$; (**d**), (**e**) and (**f**), $\sigma = 0.1$; (**g**), (**h**) and (**i**), $\sigma = 0.01$. (a), (d) and (g), the curve fitting using the Bayesian approach. The black dotted line is the real logistic function used to simulate the data. The gray dashed lines are the fitted calibration curves by the implemented software using the Bayesian approach. The calibration curves were fitted 25 times and plotted on the same graph. The open circles are the simulated standard data points used for the calibration curve estimation. (b), (e) and (h), the validation of unknown concentration estimation using the Bayesian approach. The graph is on a logarithmic scale, and the dotted line displays the true concentrations. The circle points are the estimated unknown concentrations with 95% error. (c), (f) and (i), the validation of unknown concentration estimation using the common inverse approach. The triangle points are the estimated unknown concentrations by the common inverse approach. ('NA' means no estimation because below/above detection limits.). The Bayesian estimate were closer to the line as the data variance went bigger and also gave reasonable estimates at the extreme low measurements, with the common inverse method failing by giving no estimate. (See Supplementary Fig. S3 for results on simulated data with σ of 0.95 and 0.001). The non-linear least squares regression was done in R software (R Development Core Team, 2006).

data. For each concentration, there were four replicates of simulated measurements. The constant variance σ^2 with different values was used to test the robustness of the proposed algorithm.

As shown in Figure 2 and Supplementary Figure S3, different values of variance σ^2 were used for simulation, and data were used for analyses by our approach as well as the common inverse approach for comparison purpose. The variances used for simulation were 0.95^2, 0.5^2, 0.1^2, 0.01^2 and 0.001^2 with the real variance value at a level of $0.01(0.1^2)$ as determined from experimental data. The results showed that with low measurement errors (small σ^2)

Table 2. The SSEs for calibration curve fitting using the Bayesian approach and common non-linear least squares regression on simulated data

Variance level (σ)	SSE	
	Bayesian approach	Non-linear least squares regression
0.01	1.28×10^7	0.174×10^7
0.1	3.42×10^8	5.75×10^8
0.2	4.83×10^7	6.85×10^8
0.3	8.70×10^9	8.29×10^9
0.5	6.24×10^9	1.54×10^{10}

The standards data were simulated with equal variance models based on a five-parameter logistic function with different variance levels as in Figure 2. The Bayesian approach and the non-linear least squares regression were used to estimate the five-parameter calibration curve, and SSEs were calculated for comparison. The non-linear least squares regression was done using the R software (R Development Core Team, 2006).

both approaches worked well on estimation of the calibration curve and unknown concentrations. With a small variance, although the common inverse method generated a more precise calibration curve with smaller sum square errors (SSE; Table 2), it failed on estimation of extreme measurements. When the variance became bigger, our Bayesian approach had a better performance on estimation of the calibration curve (smaller SSEs) and unknown concentrations (closer to their true values and no 'below/above detection limit' values at ends of the curve with extreme concentrations).

The Bayesian approach was also applied to simulated data with an unequal variance model. The model to generate data was similar to Equation (7), but following an unequal variance function (Davidian *et al.*, 1988),

$$\sigma_i = \delta^2 E(y_i)^\theta \qquad (8)$$

where δ is a scale factor, θ is the power of the mean responses, $[A, B, C, D, E] = [30.5754, 1.9173, 1.7235, 0.9730]$, $n = 12$ and $k = 4$.

The parameters were based on the experiment data of an ELISA assay in Davidian *et al.* (1988). The simulated data contained 12 concentrations, each with 4 replicated measurements representing levels typical for assay data. δ is the average noise level. In the simulation, it was set to be 0.1 [the actual δ value in the example data (Davidian *et al.*, 1988)], 0.3 and 0.5. θ usually is very nearly 1 for most immunoassays (Dudley *et al.*, 1985). When $\theta = 0$, the variance model [Equation (8)] reduces to an equal-variance model as we tested in the above section; When $\theta = 1$, the variance is proportional to the response; When $\theta = 2$, the coefficient of variation of the response is constant (Dudley *et al.*, 1985). In the simulation, different values of θ (1.0, 1.5, 2.0 and 4.0) were tested to compare the proposed approach with the common inverse approach.

Figure 3 displays the results of the Bayesian analysis and the common inverse approach on simulated data generated by the unequal variance model. The two panels show results for data generated using δ of 0.3 and 0.5 (see Supplementary Fig. S4 for $\delta = 0.1$). For each δ, different θ's were used to test the performance of the two approaches. We found that the Bayesian procedure was able to produce very good estimates for the unknown concentrations even on the data with unequal variances, indicating that the equal-variance approximation in this context was reasonable. Furthermore, compared to the common inverse approach, it worked better
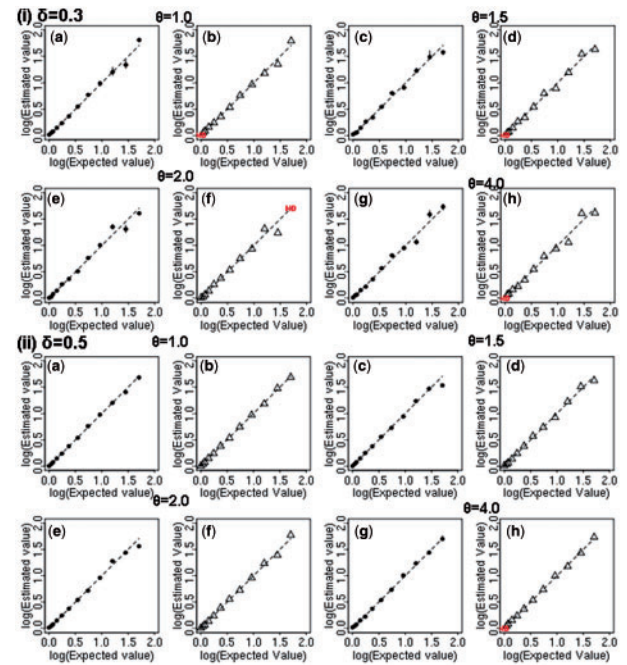

Fig. 3. Validation of the Bayesian approach and common inverse approach on simulated data with unequal variances. The simulated data were generated following the five-parameter logistic model $y = 1.91 + \frac{30.5 - 1.91}{\left(1 + \left(\frac{x}{5.60}\right)^{-0.972}\right)^{1.0}}$ with unequal variances. The variance follows $\sigma = \delta^2 y^\theta$. Twelve concentrations starting at 0.1025 were used, and four replicated measurements were simulated for each concentration. Different values of δ and θ were used to compare the performance of two approaches, with panel (i) $\delta = 0.3$ and panel (ii) $\delta = 0.5$ (see supplementary Fig. S4 for $\delta = 0.1$). In each panel, (**a**) and (**b**) $\theta = 1.0$; (**c**) and (**d**) $\theta = 1.5$; (**e**) and (**f**) $\theta = 2.0$; (**g**) and (**h**) $\theta = 4.0$. The dotted line in each graph is the true concentrations. The filled circles are estimations with 95% error bars by the Bayesian approach, and triangles are estimations by the common inverse method. The Bayesian approach could generate reasonable estimates with the common inverse method failing at the extreme measurements. ('NA', no estimate due to below/above detection limit). The non-linear least squares regression was done in R software (R Development Core Team, 2006).

for extreme measurements. On all simulated data we tested, the common procedure gave no estimates ('NA' as below detection limits) at either very high or very low responses, whereas the Bayesian estimates were reasonable and provided appropriately large uncertainties.

3.2 Experimental data

The Bayesian inference approach was also applied on experimental data generated from ELISA and Luminex assays. Figure 4 shows results for two cytokines, TNF-α by Luminex and IFN-β by ELISA. The results showed the curve fitting of estimated parameters and validation of estimated concentrations by the Bayesian approach. From the graph, we can see that our approach worked very well with fitted calibration curves and good estimations of unknowns. We also compared the Bayesian inference with the common inverse approach on 20 other cytokines measured by ELISA or Luminex assays (Supplementary Fig. S5). These results consistently indicated that the Bayesian estimates were more reliable than those of the
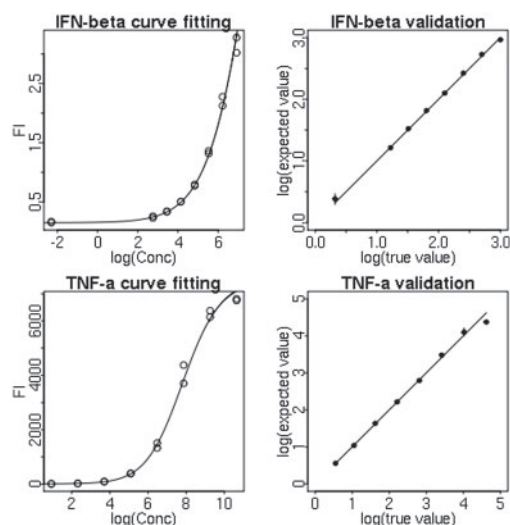
Fig. 4. Validation of the Bayesian approach on experimental data from ELISA and Luminex assays. The IFN-β data were from an ELISA assay, and TNF-α data were from a Luminex assays. The standards data contained eight concentrations with two replicated measurements, and data were then used to estimate the calibration curve by the software implementing the Bayesian approach. Lines in the curve fitting graphs (left figures) are the estimated calibration curve by our approach, and open circle points are standards data points. Lines in the validation graphs (right figures) are true concentrations and points are estimated concentrations with 95% error bar. (Please see Supplementary Fig. S5 for validation of the Bayesian approach and the common approach on data for other 20 cytokines.)

common inverse approach (Supplementary Fig. S5), especially for samples with very high and very low concentrations.

In summary, our Bayesian method produced better results than the common inverse approach on both equal and unequal variance simulated data and biological experimental data. The Bayesian method was especially good at processing noisy data. This is significant considering the prolific use of immunoassays and the fact that even realistic experimental concentrations can carry high variances. The experimentally relevant variance values for the logarithmic transformed data are 0.01 ($\sigma^2 = 0.1^2$) based on our observations as well as others (Bienvenu *et al.*, 1998; Fichorova *et al.*, 2008). This level of variance is equivalent to the coefficient of variation (CV) of 10% for the non-transformed data. The commonly accepted reliability criteria for immunoassays is set as an intraassay CV <20% (Bienvenu *et al.*, 1998; Fichorova *et al.*, 2008). A dataset with this CV value has a variance of about 0.2^2 ($\sigma^2 = 0.2^2$) after the natural logarithmic transformation. As shown in Table 2, at the level of $\sigma^2 = 0.2^2$ and above, our approach outperforms the non-linear least squares with a smaller SSE. Furthermore, depending on the type of assays and the manufacturer of kit reagents, the proportion of measurements with an intraassay CV >20% can go up to 25% (Fichorova *et al.*, 2008). Clearly methods such as our Bayesian approach are necessary to those data with high variances, where the normal approach could fail.

Furthermore, the proposed Bayesian approach can be easily extended for a wide range of applications. For example, the proposed

model by the Bayesian approach with its estimated uncertainty in calibration curves and target measurements can be used as a lower level in the hierarchical Bayesian models (Gelman *et al.*, 2004) that integrate a sampling likelihood to compare cytokine concentrations across treatment groups, measurements on different experiment days or even with different assays. Such a model would correctly incorporate the calibration uncertainties and benefit the following inference. Another possible application is the Bayesian non-parametrics (such as mixtures of Polya trees) (Christensen *et al.*, 2008) to further relax assumptions on the prior distributions of parameters of the model.

REFERENCES

Bienvenu,J.A. *et al.* (1998) Cytokine assays in human sera and tissues. *Toxicology*, **129**, 55–61.

Christensen,R. *et al.* (2008) Parametric nonparameteric statistics. *Am. Stat*, **62**, 296–306.

Conlon,E.M. *et al.* (2007) Bayesian meta-analysis models for microarray data: a comparative study. *BMC Bioinformatics*, **8**, 80.

Davidian,M. and Giltinan,D.M. (1995) *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall, London.

Davidian,M. *et al.* (1988) Variance functions and the minimum detectable concentration in assays. *Biometrika*, **75**, 549–556.

Dellaportas,P. and Stephens,D.A. (1995) Bayesian analysis of errors-in-variables regression models. *Biometrics*, **51**, 1085–1095.

Dudley,R.A. (1985) Guidelines for immunoassay data processing. *Clin. Chem.*, **31**, 1264–1271.

Fichorova,R.N. *et al.* (2008) Biological and technical variables affecting immunoassay recovery of cytokines from human serum and simulated vaginal fluid: a multicenter study. *Anal. Chem.*, **80**, 4741–4751.

Finney,D.J. and Phillips,P. (1977) The form and estimation of a variance function, with particular reference to radioimmunoassay. *Appl. Stat.*, **26**, 312–320.

Frigessi,A. *et al.* (2005) Genome-wide estimation of transcript concentrations from spotted cDNA microarray data. *Nucleic Acids Res.,* **33**, e143.

Gelman, A. *et al.* (2004a) *Bayesian Data Analysis*. Chapman&Hall/CRC, Boca Raton, FL, USA.

Gelman,A. *et al.* (2004b) Bayesian analysis of serial dilution assays. *Biometrics*, **60**, 407–417.

Gottschalk,P.G. and Dunn,J.R. (2005) The five-parameter logistic: a characterization and comparison with the four-parameter logistic. *Anal. Biochem.*, **343**, 54–65.

Harris,A.S. *et al.* (1998) Development and application of immunoassys for biological and environmental monitoring. In Cech, J.J. Jr *et al.* (eds) *Multiple Impacts on Ecosystems*. Lewis Chelsea, MI, USA, pp. P135–P153.

Hennion,M.-C. and Barcelo,D. (1998) Strengths and limitations of immunoassays for effective and efficient use for pesticide analysis in water samples: a review. *Anal. Chim. Acta*, **362**, 3–34.

Higgins,K.M. *et al.* (1998) The effect of serial dilution error on calibration inference in immunoassay. *Biometrics*, **54**, 19–32.

Landsteiner,K. (1990) *The Specificity of Serological Reactions*. Harvard University Press, Cambridge, MA, USA.

Ng,P.C. *et al.* (2003) Proinflammatory and anti-inflammatory cytokine responses in preterm infants with systemic infections. *Arch. Dis. Child Fetal Neonatal Ed.*, **88**, F209–F213.

Oellerich,M. (1980) Enzyme immunoassays in clinical chemistry: present status and trends. *J. Clin. Chem. Clin. Biochem.*, **18**, 197–208.

R Development Core Team (2006) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roberts,G.O. and Rosenthal,J.S. (2001) Optimal scaling for various Metropolis-Hastings algorithms. *Stat. Sci.*, **16**, 351–367.

Sivaganesan,M. *et al.* (2008) A Bayesian method for calculating real-time quantitative PCR calibration curves using absolute plasmid DNA standards. *BMC Bioinformatics*, **9**, 120.

Yalow,R.S. and Berson,S.A. (1959) Assay of plasma insulin in human subjects by immunological methods. *Nature*, **184** (Suppl. 21), 1648–1649.

Yalow,R.S. and Berson,S.A. (1960) Immunoassay of endogenous plasma insulin in man. *J. Clin. Invest.*, **39**, 1157–1175.