Environ Ecol Stat (2015) 22:161–177
DOI 10.1007/s10651-014-0292-5

Combining and comparing multiple serial dilution assays of particles in solution: application to brucellosis in elk of the Greater Yellowstone Ecosystem

Jarrett J. Barber · Pritam Gupta · William Edwards · Kiona Ogle · Lance A. Waller

Received: 30 September 2013 / Revised: 21 April 2014 / Published online: 14 May 2014

Abstract The concentration detection threshold (CDT) is the concentration of particles in solution beyond which a (serial dilution) assay detects particle presence. By our account, CDTs typically are not estimated but are fixed at some value. Setting a CDT to zero ($d = 0$) implies perfect detection, a common assumption, and setting $d > 0$ gives results that are "denominated" in units of d, i.e., are relative to the choice of d. Using multiple, different serial dilution assays, each with its own CDT, we choose a "reference assay," to which we assign a fixed CDT value, to obtain relative estimates of the remaining assays' CDTs and the underlying particle concentration. We present the CDTs as a novel way to account for or to compare different serial dilution assays, "sensitivities". We apply our methodology to data from four assays of the presence of bacterial (*B. abortus*) antibodies in the serum of elk in the Greater Yellowstone Ecosystem, where transmission of brucellosis—the disease ensuing from infection—to commercial livestock is managed by the Wyoming Game and Fish Department to

Handling Editor: Pierre Dutilleul.

J. J. Barber (✉)
School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, USA
e-mail: Jarrett.Barber@asu.edu

P. Gupta
Novartis Healthcare Pvt. Ltd., Hyderabad, Andhra Pradesh, India

W. Edwards
Wyoming Game and Fish Department, Laramie, WY, USA

K. Ogle
School of Life Sciences, Arizona State University, Tempe, AZ, USA

L. A. Waller
Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA

avoid the primary symptom of abnormal fetal abortion. Results agree qualitatively with the more traditional notion of sensitivity as the true positive rate.

Keywords Brucellosis · CDT · Concentration detection threshold · Particle concentration · Sensitivity · Serial dilution assay

1 Introduction

1.1 Motivating problem: brucellosis

Brucellosis is an infectious bacterial disease caused by *Brucella abortus* and causes reproductive failure primarily in cattle, elk, and bison as well as undulant fever in humans. The U.S. government initiated an aggressive eradication campaign in the 1930s to eliminate this disease from the nation's cattle population (Ragan 2002). Although it took over 70 years for this program to be successful in cattle, this disease remains endemic in the elk (*Cervus elaphus*) and bison (*Bison bison*) of the Greater Yellowstone Ecosystem (GYE), encompassing portions of Wyoming, Idaho, and Montana, USA (Thorne et al. 1997; Cheville et al. 1998). Transmission most commonly occurs by oral contact with aborted fetuses and associated birthing tissues and fluids (Thorne 2001).

In an effort to limit the spread of brucellosis from elk and bison to domestic cattle, the Wyoming Game and Fish Department (WGFD) operates 22 elk winter feedgrounds in western Wyoming, USA, while the US Fish and Wildlife Service oversees the National Elk Refuge, which provides supplemental feed to both elk and bison near Jackson, Wyoming, USA (Smith 2001). The main function of the feedground program is to provide spatial separation of elk and bison from cattle, and to prevent damage to agricultural resources. On the downside, feedgrounds perpetuate brucellosis by congregating animals during the period of peak brucellosis transmission from February through June (Roffe et al. 2004; Cross et al. 2007).

To determine if an animal has been exposed to *B. abortus*, WGFD uses a number of assays to test for the presence of *B. abortus* antibodies in blood serum. We use data from four federally approved serological assays of elk serum as implemented by the WGFD: card test (CARD), manual complement fixation test (CF), rivanol precipitation-plate agglutination (RIV) and standard plate agglutination test (SPT). The data used here were originally collected by the WGFD to assess assay sensitivities—in the traditional, true positive sense—(Morton et al. 1981); all infections were confirmed by *B. abortus* cultures from lymph node or blood extractions using methods described by Thorne et al. (1978).

Our longer-term goal, beyond the scope of this paper, is to combine other, similar assay results, as they become available from the ongoing WGFD disease management efforts, to predict underlying relative concentrations of *B. abortus* antibodies in serum as a function of spatial, temporal and other risk factors while accounting for different assay sensitivities—such accounting being our current focus. Ultimately, we envision mapping "hot spots" of relative *B. abortus* antibody concentration to serve

WGFD management and vaccination efforts and to inform the ecology of brucellosis in elk.

Our current data allow partial progress toward this goal. Moreover, our data motivate us to develop a model for relative concentration of particles in solution (e.g., serum antibodies) that has general applicability to multiple serial dilution assays of particles in solution. In particular, our model combines data from multiple serial dilution assays to allow comparison of assay "sensitivities" via particle concentration detection thresholds (CDTs) as we discuss below, a different notion than the traditional "true positive" sensitivity. Henceforth, we use "threshold sensitivity" to distinguish, when necessary, these two notions of sensitivity, and we note that a smaller CDT indicates a more (threshold) sensitive assay, which is opposite to interpreting a larger proportion of true positives to indicate a more sensitive assay.

1.2 Statistical background

The assumption that particles follow a homogeneous Poisson process(HPP)—equivalently, that any given number of particles is distributed uniformly in solution—has a relatively long history of application to serial dilution assays. Early accounts include McCrady (1915), Fisher (1922) and Cochran (1950). See also McCullagh and Nelder (1989, Sec.1.2.4), who use Fisher's (1922) account to introduce generalized linear models.

In addition to the HPP assumption, particles are often assumed to be detected perfectly (Finney 1978; Ridout 2005)—the presence of a single particle (or more) results in a positive response. This so-called "one-hit" Poisson model (Mehrabi and Matthews 1998) may be a tenable assumption in cases where, for example, a single bacterium divides to produce visible colonies over time; the "multi-hit" model of dose-response studies (Cornfield 1954; Cornfield and Mantel 1977) embodies the notion that more than one particle may be required for detection to occur. Below, we develop a model in terms of the concentration above which particles in solution are detected in a serial dilution assay. We refer to this concentration as the concentration detection threshold (CDT), $d > 0$, which serves as our basis for comparing and accounting for different threshold sensitivity of assays to detect particles in solution.

To establish concepts and notation, consider the relatively simple problem of detecting particles using a single assay that results in an indicator of the presence of particles in a "batch" of solution, e.g., antibodies in serum, with concentration $\lambda_0 > 0$. If Z is the number of particles in a volume, v, of solution then $Z \sim \text{Poisson}(\lambda_0 v)$, assuming particles are completely spatially random (CSR)—i.e., particles follow an HPP. If Y is a binary indicator of the presence of one or more particles, then a "positive" event, $\{Y = 1\}$, corresponds to $\{Z > 0\}$, and the probability of particle presence follows as $[Y = 1] = [Z > 0] \equiv Pr(Z > 0)$—the probability of (perfect) detection. While we refer to Y as a "binary" variable here, it may be better for subsequent development to emphasize Y as a categorical variable whose values indicate the highest dilution level—trivially one in this example—at which particles are detected, zero indicating that particles are not detected at any dilution level.

More generally, we may allow for imperfect detection with $\{Y = 1\} = \{Z > c\}$, for some particle count $c \geq 0$. Reparameterizing with $c = dv$, we express particle detection as $\{Y = 1\} = \{Z > dv\}$, where $d \geq 0$ is the aforementioned CDT, so that $[Y = 1] = [Z > dv] \equiv Pr(Z > dv)$ is the probability of possibly imperfect detection, and $[Y = 0] = [Z \leq dv]$ is probability of non-detection, $1 - [Y = 1]$. Below, we develop the estimation of the concentration, d, while accounting for measurement uncertainty in volume, v.

As suggested by previous work (Lee and Whitmore 1999) using a single assay, wherein d has been fixed, we cannot identify d from λ_0. To see this, choose values $d = d_1$ and $\lambda_0 = \lambda_{01}$ (with fixed v) so that $p = Pr(Z > d_1 v)$. Then, for a different value, $d = d_2$, we can generally choose a different value, $\lambda_0 = \lambda_{02}$, to obtain the same result, $p = Pr(Z > d_2 v)$. Thus, different combinations of d and λ_0 result in the same probability of detection, hence they cannot be identified from one another. With data from multiple assays, however, one assay serves as a reference assay, with fixed CDT, which allows inference about remaining assay CDTs and the common concentration, λ_0.

In their presentation of statistical summaries for grouped serial dilution assay data, Hamilton and Rinaldi (1988) discuss "threshold concentration" as the "...concentration of the test material just sufficient to cause a response," where "response" again indicates presence or absence of a substance in solution. However, they do not appear to account for imperfect detection of particles. Lee and Whitmore (1999) discuss detection thresholds greater than one, but they ultimately fix $d = 1$ in their analysis so that their concentrations are "denominated in units of d." In other words, by fixing $d = 1$, they effectively assume perfect detection of a "d unit" of concentration. We follow a similar route by fixing one of our four assay CDTs to 1, creating our aforementioned reference assay CDT.

If the CDT of an assay is known, which is implied by, for example, the assumption of perfect detection, or if we could calibrate measurements against a known standard concentration, then, in principle, we could estimate true concentration. For example, continuing the ideas of Giltinan and Davidian (1994), Dellaportas and Stephens (1995) and Davidian and Giltinan (1995, Chap. 10), Gelman et al. (2004) use the concentration of a known standard in a Bayesian framework to calibrate concentrations obtained from serial dilution assays. In addition, Block and Chavance (1998) use count data arising at each dilution level from repeated application of the same assay, along with the assumption of perfect detection, to estimate concentration in the context of comparing treatment effects on viral load. Without the assumption of perfect detection or known detection concentration, without a known, calibrating concentration, or without other assumptions, estimates of particle concentration are relative. Ridout (2005) gives an overview of serial dilution assays, including point estimation, bias correction, interval estimation, and design.

Features of our approach include the following: (i) assay procedures are compared via explicitly specified assay-specific CDTs, or threshold sensitivities; (ii) standards of known concentration are not required to compare assay sensitivities; (iii) information from multiple observations from multiple (different) assays are used simultaneously to inform the comparison of assay-specific threshold sensitivities; (iv) the models are a conceptually straightforward extension of statistical notions underlying existing

studies on serial dilution assays; and (v) the models are easily implemented in a Bayesian framework using freely available software.

Section 2 continues the basic statistical ideas presented here to develop models that combine and compare observations from multiple, different serial dilution assays of the presence of particles in solution. We illustrate our models using simulated data in Sect. 3, then apply the models to data from multiple assays of seropositivity of the *B. abortus* antibody in elk of the Greater Yellowstone Ecosystem (GYE) of Wyoming. We conclude with a discussion of our methodology and application in Sect. 4.

2 Models

Generally, we consider J serial dilution assay procedures where assay $j \in \{1, 2, \ldots, J\}$ has K_j dilution levels. It seems natural to envision a procedure whereby particle detection is assessed and recorded for each dilution level of an assay. However, it is important to note, for subsequent model development, that the recorded data available to us indicate only the last dilution level at which particles were detected. Thus, we are assured that particles were not detected at subsequent dilution levels but cannot say whether or not particles were detected at previous dilution levels. For our data, $J = 4$ and, using the assay abbreviations introduced in Sect. 1.1, $\{1, 2, 3, 4\}$ corresponds to {CARD, CF, RIV, SPT}.

Following our remarks on Y and Z in Sect. 1.2, we let Y_{ij} indicate the highest dilution number k at which particles are detected by assay j for sample $i \in \{1, 2, \ldots, N\}$ obtained from a batch of solution of concentration λ_0; $Y_{ij} = 0$ indicates non-detection at all dilution levels. Thus, Y_{ij} is a categorical random variable with support $\{0, 1, \ldots, K_j\}$, $j = 1, \ldots, J$—$\{0, 1\}$ $(j = 1)$, $\{0, 1, \ldots, 7\}$ $(j = 2)$, $\{0, 1, 2, 3, 4\}$ $(j = 3)$ and $\{0, 1, 2, 3, 4\}$ $(j = 4)$.

Continuing with our simple model in Sect. 1.2, we associate Y_{ij} with $Z_{ijk} \sim \text{Poisson}(\mu_{ijk})$, where $\mu_{ijk} = m_{jk}\lambda_0 v_{ijk}$ is the mean, and m_{jk} is a known dilution fraction for dilution k of assay j, relative to undiluted solution; thus, m_{jk} is an additive offset in the (log) linear predictor of the mean that adjusts initial concentration, λ_0, to reflect dilution. Often $m_{jk} = m_j^{-k}$, where m_j is a known, constant dilution ratio for assay j. For example, if $m_j = 2$, then the solution concentration is halved from one dilution to the next. In our experience, a dilution procedure typically does not vary by sample, and we omit the i subscript from m. For us, nominal solution volumes, v^*_{jk}, do not change across i, but, to account for uncertain volume measurements, we use a Berkson specification (Dellaportas and Stephens 1995) so that "true" volume, v_{ijk}, varies about nominal volume; see below. Table 1 summarizes the number of dilutions, K_j, the dilution fractions, m_{ij}, and nominal volumes, v^*_{jk}, for our data.

If the Z_{ijk} are obtained from disjoint volumes, as for our data, then the Z_{ijk} are conditionally independent given their means, hence the Y_{ij} are conditionally independent—given, additionally, the CDTs and volumes—and the probabilities $[Y_{ij} = k]$ follow from the product rule applied to the Poisson probabilities of the following events:

Table 1 Summary of assay procedures

Assay j	Number of dilutions K_j	Dilution fraction relative to undiluted solution m_{jk}	Nominal volume (μl) v^*_{jk}
1	1	1/2	60
2	7	1/20, 1/40, 1/80, 1/160, 1/320, 1/640, 1/1280	50, 50, 50, 50, 50, 50, 50
3	4	8/11, 4/7, 2/5, 1/4	110, 70, 50, 40
4	4	4/11, 2/7, 1/5, 1/8	110, 70, 50, 40

$$\begin{aligned}
&\bigcap_{k=1}^{K_j}\{Z_{ijk} \le d_j v_{ijk}\} && \equiv \{Y_{ij} = 0\}\\
&\left\{Z_{ijk'} > d_j v_{ijk'}\right\}\bigcap\left\{\bigcap_{k=k'+1}^{K_j}\{Z_{ijk} \le d_j v_{ijk}\}\right\} && \equiv \{Y_{ij} = k'\} \quad k' \in \{1,\ldots,K_j - 1\}.\\
&\{Z_{ijK_j} > d_j v_{ijK_j}\} && \equiv \{Y_{ij} = K_j\}
\end{aligned} \tag{1}$$

The first line in (1) depicts the event that none of the particle counts exceed their corresponding threshold counts at any dilution, the last line depicts the event that the last dilution count exceeds its threshold, and the middle line indicates the event that dilution k' is the last dilution at which a count exceeds its threshold, with counts of all subsequent dilutions falling at or below their thresholds but with counts of previous dilutions unconstrained. Note that expression (1) embodies the description of our data given at the beginning of this section. That is, our data tell us that no exceedances occur beyond the last observed exceedance, of course, but, moreover, our data do not tell us about exceedances or (likely rare) non-exceedances at previous dilutions. Also, it explains our remark, in Sect. 1.2, that refers to Y as "categorical" rather than "binary".

To draw a connection to the common assumption of perfect detection, mentioned above but not used in this article, the probability of not (perfectly) detecting particles at any dilution is $[Y_{ij} = 0] = \exp(-\sum_{k=1}^{K_j}\mu_{ijk})$, which implies that particles do not exist in solution volumes at or beyond the first measured dilution level, $k = 1$ (which may or may not be diluted). The probability of (perfectly) detecting particles at all dilution levels is $[Y_{ij} = K_j] = \prod_{k=1}^{K_j}(1 - \exp(-\mu_{ijk}))$, and the probability that dilution k' is the last dilution at which particles are (perfectly) detected is $[Y_{ij} = k'] = (1-\exp(-\mu_{ijk'}))\exp(-\sum_{k=k'+1}^{K_j}\mu_{ijk})$. Of course, if some assays, in actuality, do not have perfect detections or, more generally, if assays have different detection thresholds, then the assumption of perfect detection gives incorrect probabilities, hence would lead to incorrect inference of particle concentration. Intuitively, for example, if we assume perfect detection, but, in actuality, detection is imperfect, then a perfect detection model would tend to underestimate concentration in a downward adjustment from the true, higher concentration required to generate particle counts above non-zero thresholds.

Finally, collecting the Y_{ij} and d_j into vectors $\mathbf{Y}$ and $\mathbf{d}$, respectively, we write the likelihood as

$$[\mathbf{Y} \mid \mathbf{d}, \lambda_0] \equiv \prod_{ij} [Y_{ij} \mid d_j, \lambda_0].$$

Introducing priors, $[\mathbf{d}]$ and $[\lambda_0]$, we write the posterior distribution as

$$[\mathbf{d}, \lambda_0 \mid \mathbf{Y}] \propto [\mathbf{Y} \mid \mathbf{d}, \lambda_0][\mathbf{d}][\lambda_0]. \tag{2}$$

We refer to model (2) as the "constant effect model" for the constant concentration, λ_0. We notationally suppress the volumes, v_{ijk}, here and in much of the remainder of our presentation because the volume measurement error is not informed by our data—posteriors (not shown) are indistinguishable from priors—and this error appears to have practically no discernible impact on inferences, but we do want a more complete accounting of uncertainty.

In our application (Sect. 3), we do not have a well-defined batch, but, instead, have a serum sample from each of N elk, which lived among a population of elk in the GYE, and we view λ_0 as a population-level serum antibody concentration. And, we extend the model to allow concentration to vary by subject (elk) i. More precisely, we specify independent subject random effects

$$\lambda_i \sim [\lambda_i \mid \sigma_1] \equiv \text{log-normal}(0, \sigma_1^2) \tag{3}$$

given λ_0 and the (log scale) variance component, σ_1^2; we use the 1 subscript to distinguish from additional effects and their variance component in a subsequent model. Thus, $\mu_{ijk} = \lambda_0 \lambda_i m_{jk} v_{ijk}$, in which we see multiplicative elk effects modifying the overall effect, λ_0. Letting $\boldsymbol{\lambda}_1$ be the vector with ith component λ_i, we write $[\boldsymbol{\lambda}_1 \mid \sigma_1] \equiv \prod_i [\lambda_i \mid \sigma_1]$, and re-write the likelihood and posterior, respectively, as

$$[\mathbf{Y} \mid \mathbf{d}, \lambda_0, \boldsymbol{\lambda}_1] \equiv \prod_{ij} [Y_{ij} \mid \lambda_0, \lambda_i, d_j],$$

and

$$[\mathbf{d}, \lambda_0, \boldsymbol{\lambda}_1, \sigma_1 \mid \mathbf{Y}] \propto [\mathbf{Y} \mid \mathbf{d}, \lambda_0, \boldsymbol{\lambda}_1][\mathbf{d}][\lambda_0][\boldsymbol{\lambda}_1 \mid \sigma_1][\sigma_1]. \tag{4}$$

We refer to model (4) as the "elk effects model".

Results (Sect. 3; Table 4) from fitting (4) to our data suggest that our observed Y_{ij} values exhibit more variability than can be accounted for by this model, so we extend the model to include decomposition of variability into further effects, λ_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, J$, which we may interpret as (additive) "elk–assay" interaction effects (on the log scale). We speculate on the sources of these effects and offer related discussion in Sect. 4. In this article, we are mainly interested in comparing assays via the d_j CDTs, and to a lesser extent, in λ_0, λ_i, and λ_{ij}, and seek to separate these effects from the aforementioned effects, regardless of their source. With additional data, e.g., spatial location, accounting for such heterogeneity may help to better infer λ_i as a function of unidentified spatial risk factors (Elliott et al. 2000), a brief point of discussion in Sect. 4.

Thus, we specify independently

$$\lambda_{ij} \sim [\lambda_{ij} \mid \sigma_2] \equiv \text{log-normal}(0, \sigma_2^2), \tag{5}$$

where σ_2^2 is a (log scale) variance component. Thus, $\mu_{ijk} = \lambda_0 \lambda_i \lambda_{ij} m_{jk} v_{ijk}$. To revise the likelihood, let $\boldsymbol{\lambda}_2$ be the vector that collects the λ_{ij} and define $[\boldsymbol{\lambda}_2 \mid \sigma_2] \equiv \prod_{ij} [\lambda_{ij} \mid \sigma_2]$. Now, we denote the likelihood as

$$[\mathbf{Y} \mid \mathbf{d}, \lambda_0, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2] \equiv \prod_{ij} [Y_{ij} \mid \mathbf{d}, \lambda_0, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2],$$

and posterior as

$$[\mathbf{d}, \lambda_0, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \sigma_1, \sigma_2 \mid \mathbf{Y}] \propto [\mathbf{Y} \mid \mathbf{d}, \lambda_0, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2][\mathbf{d}][\lambda_0] \times [\boldsymbol{\lambda}_1 \mid \sigma_1][\boldsymbol{\lambda}_2 \mid \sigma_2][\sigma_1][\sigma_2]. \tag{6}$$

We refer to model (6) as the "elk–assay effects model". We do not include assay effects, say λ_j, in the linear predictor of the mean, μ_{ijk}, because we cannot identify these effects from the d_j.

Finally, for lack of information, we specify independently, relatively vague priors, $\lambda_0 \sim$ log-normal(0, 1000) and $\sigma_p \sim$ uniform(0, 20), $p = 1, 2$. For the CDTs in our example, we set $d_2 = 1$, making the CF assay the reference assay, which our experience and results indicate to be the smallest CDT, hence CF is the most threshold sensitive assay. Without this constraint or other constraint, e.g., sum to zero on the log scale, we cannot identify the overall mean level of the d_j from λ_0, which is presumably the reason for fixing d in previous work (Lee and Whitmore 1999), and which is more generally akin to placing constraints on linear predictor effects to achieve estimability or identifiability. The remaining d_j are independently distributed vaguely as log-normal(0, 1000). Finally, we specify independently, $v_{ijk} \sim$ log-normal($\ln(v^*_{jk})$, $\ln(0.005^2 + 1)$) to incorporate volume measurement error about nominal values, v^*_{jk} (Table 1), with 0.5 % coefficient of variation, according to WGFD experience.

3 Applications to simulated and observed data

We explored our models and implementations using simulated data before applying them to our observed data. For Markov chain Monte Carlo (MCMC) posterior samples, we used JAGS (Plummer 2003), and for ML, we used the algorithm by Nelder and Mead (1965) as implemented in the `optim` function in `R` (R Core Team 2012). More details of the MCMC procedure accompany the presentation of the analysis of the observed data, after that of the simulated data. Preliminary runs, as well as results from simulated data (Sect. 3.1) suggested no discernable differences in results arising from the use of the Anscombe normal approximation (Anscombe 1948) to the Poisson, and the approximation required

roughly half the model run time. All reported results are based on this approximation.

3.1 Simulated data

We used the elk–assay effects model (6) to simulate $N = 72$ responses from each of $J = 4$ fictitious assays having the same characteristics as those used to obtain our observed data (Table 1). Then, we used these simulated data to obtain posterior distributions to assess the model before moving on to our real data. We chose values of $\ln(d_1) = 4$ (fictitious "CARD" assay), $\ln(d_2) = 0$ ("CF"), $\ln(d_3) = 4$ ("RIV"), $\ln(d_4) = 5$ ("SPT"), $\lambda_0 = 6$, $\sigma_1^2 = 0.8$ and $\sigma_2^2 = 0.8$. (We chose these to be comparable to values obtained from preliminary, exploratory analyses (not shown) of our observed data; see Sect. 3.2, below.) Thus, for our simulation, assay $j = 2$ is the most threshold sensitive because it has the lowest CDT value, and assay $j = 4$ is least sensitive. Given $\sigma_1^2 = 0.8$, we generated N random effects, λ_i, according to (3). Similarly, with $\sigma_2^2 = 0.8$, we generated $N \times J$ random effects, λ_{ij}, according to (5). The Y_{ij} were then generated according to Poisson—not the Anscombe normal—probabilities that follow from equivalent events in expression (1), where the v_{ijk} were generated according to our volume error model (Sect. 2).

As noted above, we expect non-identifiability between the overall effect, λ_0, and the overall level of the CDTs, d_j. Preliminary attempts to fit the elk–assay effects model (6) to its simulated data indicated lack of convergence consistent with this lack of identifiability. Thus, we fixed $d_2 = 1$, so that assay 2 (CF) is our "reference" assay, and proceeded to infer remaining model parameters now relative to this fixed reference.

The values used to simulate random effects and data are captured within 95 % credible intervals (Table 2). Also, the coverage rate for random effects is comparable to the nominal 95 % intervals shown in Fig. 1: λ_i, $65/72 = 0.90$; λ_{i1}, $69/72 = 0.96$; λ_{i2}, $64/72 = 0.89$; λ_{i3}, $64/72 = 0.89$; λ_{i4}, $69/72 = 0.96$; all random effects, $331/(5*72) = 0.92$. Differences between posterior medians and corresponding means are too small to appreciate visibly in Fig. 1, but seem to add clutter, so we omit medians from the figure.

Encouraged by these results, we next applied our models to our observed data.

Table 2 Posterior summary for the elk–assay effects model (6) fit to simulated data

Parameter	Mean	2.5 %	25 %	50 %	75 %	97.5 %
$\ln(\lambda_0)$	5.98	5.72	5.89	5.98	6.07	6.23
$\ln(d_1)$	4.12	3.70	3.99	4.13	4.26	4.52
$\ln(d_3)$	3.70	3.38	3.60	3.70	3.81	4.00
$\ln(d_4)$	4.97	4.70	4.88	4.97	5.06	5.27
σ_1	0.82	0.62	0.74	0.81	0.89	1.06
σ_2	0.76	0.65	0.72	0.76	0.80	0.90

$d_2 = 1$ is the reference CDT. Corresponding true values: $\ln(d_1) = 4$ ("CARD"), $\ln(d_2) = 0$ ("CF"), $\ln(d_3) = 4$ ("RIV"), $\ln(d_4) = 5$ ("SPT"), $\ln(\lambda_0) = 6$, $\sigma_1 = 0.8$, $\sigma_2 = 0.8$. See also Fig. 1

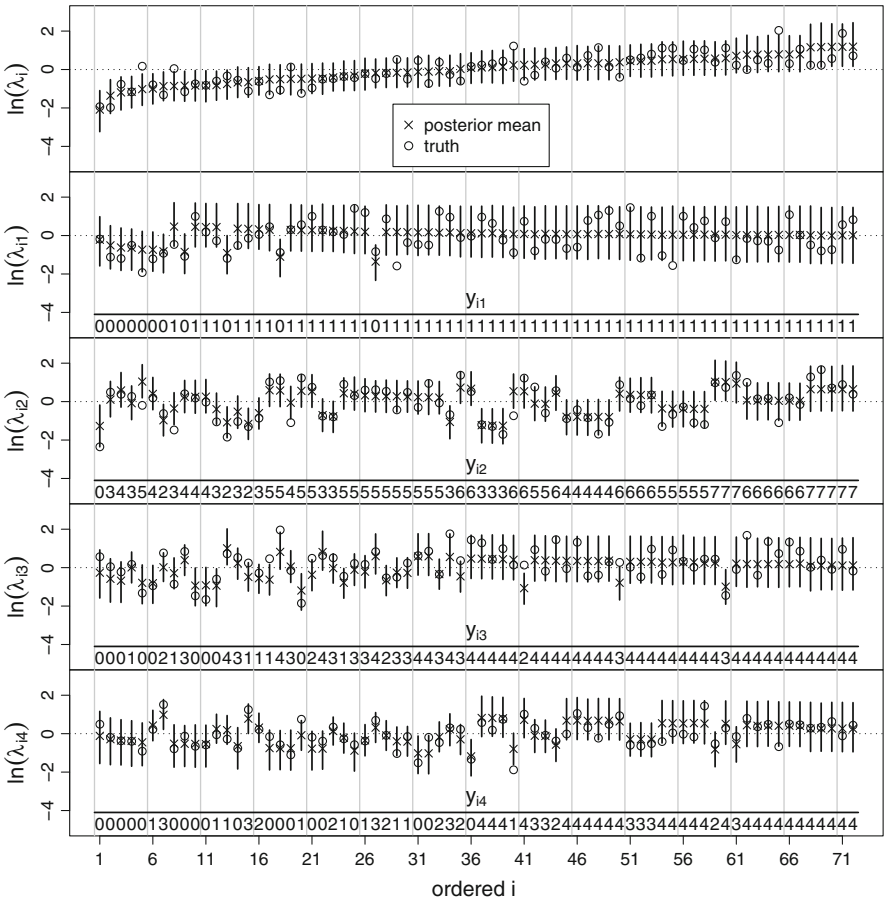

Fig. 1 Posterior summary of effects, $\ln(\lambda_i)$ and $\ln(\lambda_{ij})$, for the elk–assay effects model (6) fit to simulated data, Y_{ij}, shown for comparison. Observation i is made to follow the ascending order of corresponding posterior means of $\ln(\lambda_i)$ (*top*). See also Table 2

3.2 Brucellosis in GYE elk

While MCMC mixing and convergence for the elk effects model (4) fit to our observed data was good (with the reference constraint, $d_2 = 1$), the model was not flexible enough to replicate the data well, as we explain more below. So, we implemented the elk–assay effects model (6) to account for the apparent heterogeneity beyond that captured by the elk effects, λ_i, alone.

Maximum likelihood (ML) values from the constant effect model (2) were used to start five MCMC chains for the elk–assay effects model (6) with reference constraint $d_2 = 1$; starting values were dispersed from the ML values. We dispersed starting values for σ_1 and σ_2 near 1, and used JAGS' default procedure to randomly generate the starting values for the random effects, λ_i and λ_{ij}. Dilution volumes, v_{ijk}, were started at their nominal values, v^*_{jk} (Table 1).

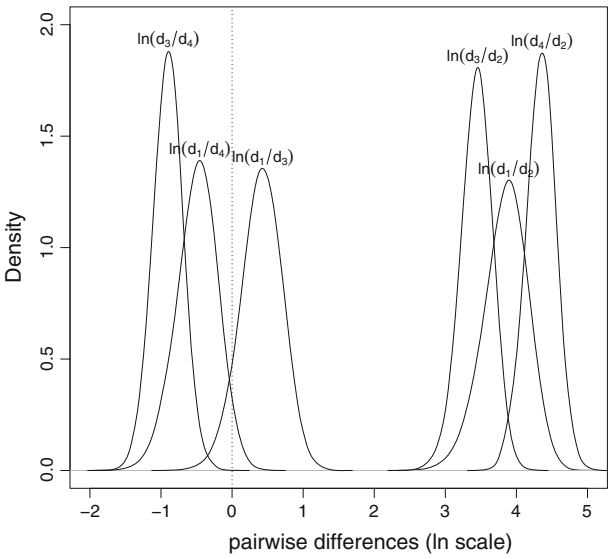

Fig. 2 Marginal posteriors of all pairwise differences of CDTs for the elk–assay effects model (6) fit to observed data

After using JAGS' default adaption procedure for 10,000 iterations and after omitting several hundred thousand subsequent iterations before visually assessing convergence (not shown), we retained a subsequent 200,000 iterations for each of the five chains, each thinned by 40 (for storage reduction and speed of summary procedures), for a total MCMC sample size of 25,000. All of the upper 95 % credible bounds for the potential scale reduction factors (psrf) of all stochastic quantities in model (6) were less than 1.02, and the multivariate psrf was 1.07 (Brooks and Gelman 1998).

A summary of marginal posteriors is given in Table 3. The reference assay CDT ($d_2 = 1$) is clearly the smallest, hence the CF assay is the most threshold sensitive assay. Relative threshold sensitivities among assays may be more easily seen by the posteriors of the pairwise differences of CDTs (ln scale) in Fig. 2, which reveals, roughly, $d_2 < d_3 < d_1 < d_4$ ("CF < RIV < CARD < SPT"), so that SPT is the least threshold sensitive assay, and RIV and CARD are intermediate.

For comparison, Fig. 3 shows traditional sensitivities—probabilities of a true positive; recall that our data are from elk that are known to have brucellosis by *B. abortus* culture methods (Thorne et al. 1978). Typically, the WGFD computes such traditional sensitivities directly from the data as the proportion of an assay's "positive" results out of N. In particular, they compute $\sum_{i=1}^{N} I(Y_{i1} \geq k_j)/N$, where $k_1 = 1$, $k_2 = 2$, $k_3 = 1$ and $k_4 = 3$ are their chosen dilution levels at or beyond which antibody presence must be observed to declare a "positive" result for assay j of serum sample i. Thus, the WGFD would obtain empirical estimates of $Pr(Y_{ij} = 1)$: $66/72 = 0.92$ ($j = 1$); $52/53 = 0.98$ ($j = 2$, omitting 19 missing observations); $68/72 = 0.94$ ($j = 3$); $57/72 = 0.79$ ($j = 4$). (see Fig. 4 for the data used to compute these values.) Figure 3 shows the posterior predictive distributions of $\sum_{i=1}^{N} I(Y_{i1}^{rep} \geq k_j)/N$, $j = 1, \ldots, J$, where the Y_{ij}^{rep} are replicated values of the observations, Y_{ij}, along

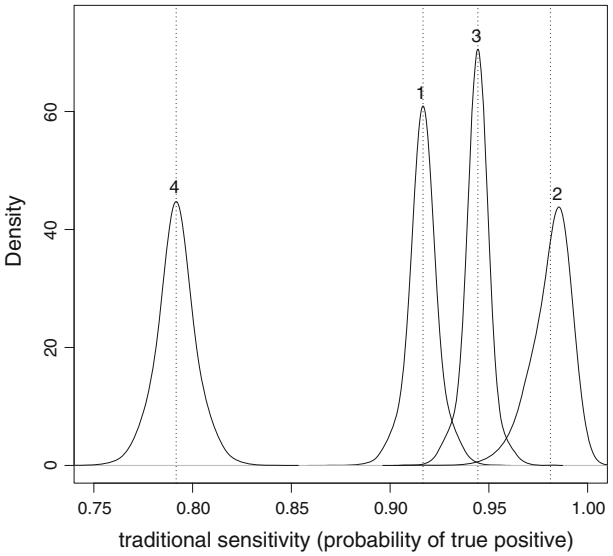

Fig. 3 Marginal posteriors of the probability of a true positive for the elk–assay effects model (6) fit to observed data. Values atop density estimates indicate assay j. *Dotted vertical lines* indicate corresponding empirical estimates, each of which occurs closest to its corresponding marginal posterior, as expected. See text for "positive" criteria

with the corresponding empirical estimates indicated by vertical dotted lines. We see that the posterior and empirical estimates are in good agreement. And, we see that, from the traditional perspective, sensitivities are ordered in the same way as our threshold sensitivities, above (Table 3; Fig. 2): CF ($j = 2$) is most sensitive, SPT ($j = 4$) is least sensitive, and RIV ($j = 1$) and CARD ($j = 3$) sensitivities are intermediate.

The WGFD also computes an overall, or composite, traditional sensitivity as the proportion of observations for which $(\{Y_{i1} = 1\} \cap \{Y_{i3} \geq 1\} \cap \{Y_{i4} \geq 3\}) \cup \{Y_{i2} \geq 2\}$ is true: 71/71; despite 19 missing CF ($j = 2$) values out of $N = 72$, we can still evaluate this empirical expression for all but one of the missing cases because a "positive" can be declared for 18 of the cases that give (positive $\cup$ missing) $=$ positive, while one case gives (negative $\cup$ missing) $=$ missing. The posterior predictive mean of $\sum_{i=1}^{N} I((\{Y_{i1}^{rep} = 1\} \cap \{Y_{i3}^{rep} \geq 1\} \cap \{Y_{i4}^{rep} \geq 3\}) \cup \{Y_{i2}^{rep} \geq 2\})/N$ is 0.9959, and the posterior predictive standard deviation is 0.00016, practically perfect sensitivity and too precise to display well with the individual assay sensitivities shown in Fig. 3.

The summaries of random effect standard deviations (Table 3) indicate that heterogeneity due to elk (σ_1) and to elk–assay "interaction" (σ_2) are comparable. Posterior summaries of the effects, λ_i and λ_{ij}, associated with these standard deviations, are shown in Fig. 4 along with the data, Y_{ij}, and maximum a posteriori (MAP) values of 19 missing values of the CF ($j = 2$) assay. Table 4 shows the clearly improved predictive performance of model (6) over model (4) for observed assay $j = 3$; results for the remaining assays revealed similarly improved performance (not shown).

Looking at the posterior means of the λ_i in Fig. 4, we see that the relative concentrations for elk range roughly from $\exp(-2) = 0.14$ to $\exp(1) = 2.7$ times the overall

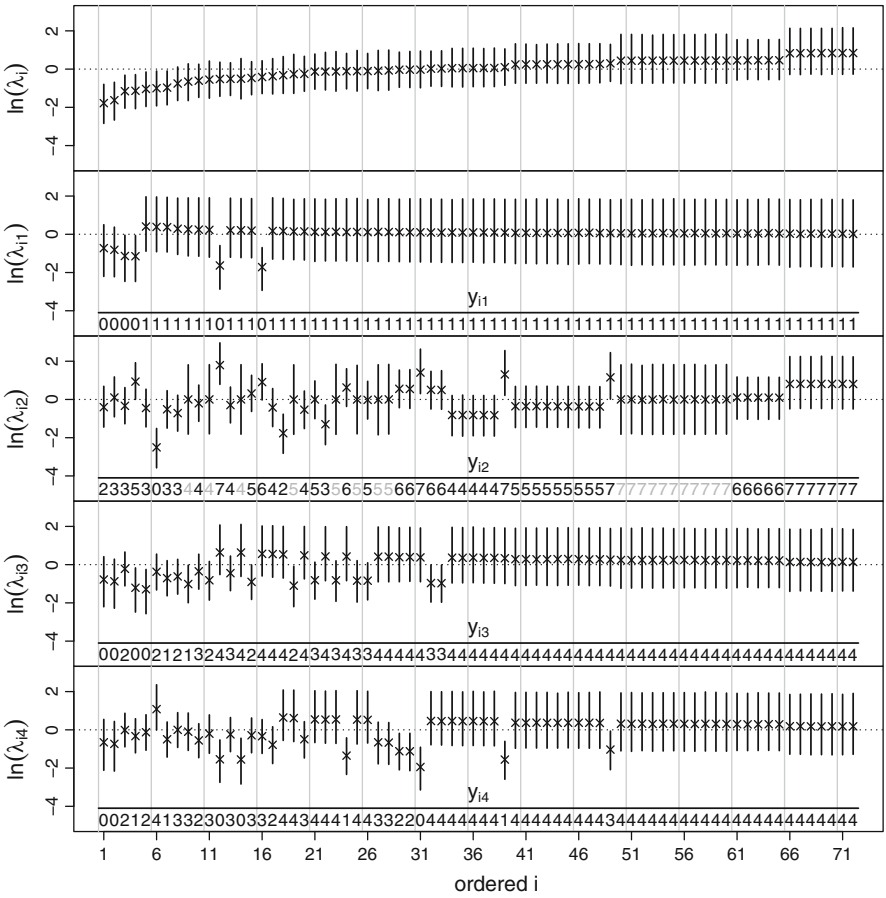

Fig. 4 Posterior summary of effects, λ_{ij}, in the elk–assay effects model (6) fit to observed data. Observation i is made to follow the ascending order of corresponding posterior means of $\ln(\lambda_i)$ (*top*). Observed data shown for comparison; *light grey values* indicate MAP values of missing CF data

Table 3 Posterior summary of elk–assay effects model (6) fit to observed data

Parameter	Mean	2.5 %	25 %	50 %	75 %	97.5 %
$\ln(\lambda_0)$	6.25	5.90	6.13	6.25	6.37	6.60
$\ln(d_1)$	3.85	3.19	3.65	3.87	4.06	4.41
$\ln(d_3)$	3.44	2.97	3.29	3.45	3.59	3.86
$\ln(d_4)$	4.34	3.91	4.20	4.35	4.48	4.75
σ_1	0.78	0.54	0.69	0.77	0.87	1.08
σ_2	0.92	0.75	0.85	0.91	0.98	1.13

$d_2 = 1$ is the reference CDT. See also Fig. 4

relative concentration, λ_0. And, λ_0 is modified further by the λ_{ij}, whose magnitudes are comparable to the elk effects, λ_i. We offer possible explanations of the source of variability of the λ_{ij} in the Discussion (Sect. 4).

Table 4 Summary of predicted data values, Y_{i3}^{rep}, versus observed data values, Y_{i3}, for assay $j = 3$

Predicted Y_{i3}^{rep}	Observed Y_{i3}				
Model (4)	0	1	2	3	4
0	53.7	6.2	1.4	0.1	1.8
1	21.8	22.8	7.8	1.1	3.1
2	16.4	37.9	27.6	8.5	6.1
3	6.0	23.5	38.0	30.6	10.7
4	2.1	9.6	25.2	59.8	78.3
Model (6)	0	1	2	3	4
0	98.8	3.5	0.0	0.0	0.0
1	1.2	90.6	2.9	0.0	0.0
2	0.0	5.9	92.5	2.4	0.0
3	0.0	0.0	4.6	93.4	0.6
4	0.0	0.0	0.0	4.2	99.4

Observed column k shows the posterior distribution (as percentages) of the average, $\sum_{i=1}^{N}[Y_{i3}^{rep} \mid Y_{i3} = k]/N, k = 0, 1, 2, 3, 4$

4 Discussion

We have presented models to combine data from multiple serial dilution assays for detecting the presence of particles in solution. We have introduced a novel notion of assay sensitivity—threshold sensitivity—in the form of the concentration detection threshold (CDT), the particle concentration above which an assay detects particle presence; the smaller an assay's CDT, the more sensitive we say the assay is. In our application, the CDTs agree qualitatively with the more traditional notion of sensitivity as the probability of a true positive. We suggest that our model may be more generally useful where multiple serial dilution assays are available and the intent is to compare or to account for assay sensitivities while inferring relative concentration.

We applied our models to data from four serial dilution assays of the presence of *B. abortus* antibodies in the blood serum of $N = 72$ elk having lived in the Greater Yellowstone Ecosystem (GYE) wherein the Wyoming Game and Fish Department (WGFD) has collected such data for the purpose of managing the transmission of the disease to commercial livestock. Sensitivities, as measured by CDTs or the proportion of true positives, show the CF assay to be most sensitive, the SPT assay to be least sensitive, and the RIV and CARD assays to be intermediate, which is consistent with the experience of the WGFD and previously reported results (Morton et al. 1981).

Elk–assay effects, λ_{ij}, accounted for a level of heterogeneity comparable to that of elk effects, λ_i, and we speculate that the former effects may reflect the processing of elk serum samples and assay procedures, may reflect the clumping into antibody–antigen complexes where the degree of clumping differs from elk to elk and depends on the assay, or may reflect errors in reading the dilution levels at which a positive response is observed.

Though our data provided little information for volume measurement error, we attempt a more full accounting of error through our prior distribution on volume error. In principle, we could have chosen to account for error, instead, in the dilution fractions, but this seems to us to be a less natural specification, especially given that our information presents itself via volume measurement error. Also, we can see by μ_{ijk}, after expression (5), that volume error can be seen as a multiplicative adjustment to the dilution fractions. To specify both errors seems redundant given that error in volume measurements induces errors in dilution fractions by definition. For more on modeling volume and dilution errors in serial dilution assays, see Chase and Hoel (1975), who derive an approximate specification for dilution errors based on a multiplicative volume error model.

We anticipate the availability of future data, similar to those used here, to inform model extensions. We expect future data to include, in particular, spatial and temporal coordinates, among other potential risk factors of interest to the WGFD brucellosis management effort and to wildlife disease ecologists. These forthcoming data may include assay results for which the true status of *B. abortus* infection and antibody presence is not known as for the data used here. For such results, we will explore zero-inflation as a way to accommodate elk with no *B. abortus* antibodies, perhaps modeling the probability of having no serum antibodies as a function of available risk factors. Also, our model will account for differing assay sensitivities, as treated here, while exploring further modeling of elk effects, λ_i, as a function of other risk factors. In particular, we envision a disease map (Elliott et al. 2000) to help identify "hot spots" of unmeasured spatial or temporal risk factors of brucellosis in aid of future brucellosis management of in the GYE.

Finally, the results shown here are based on the Anscombe normal approximation to the Poisson (Anscombe 1948), which showed no discernable differences with the Poisson results (not shown), with roughly half the computation time. The appropriateness of the approximation in our application is illustrated by its ability to capture known values in our simulated data presentation of Sect. 3.1. (Data were simulated using the Poisson model described herein.) While the reduction in computation time may be modest here, computation may become more burdensome with anticipated, future model extensions, and justification, beyond verification with simulated data, may be appropriate.

Acknowledgments The authors wish to thank Jessica Jennings-Gaines and Hally Killionof for compiling the data. The first two authors received partial support from U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service (CSREES) Grant USDACSRE45232BA.

References

Anscombe FJ (1948) The transformation of poisson, binomial and negative-binomial data. Biometrika 35(3/4):246–254. http://www.jstor.org/stable/2332343

Block J, Chavance M (1998) A mixed model for repeated dilution assays. Biometrics 54(2):482–492. http://www.jstor.org/stable/3109757

Brooks S, Gelman A (1998) General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7(4):434–455

Chase GR, Hoel DG (1975) Serial dilutions: error effects and optimal designs. Biometrika 62(2):329–334. http://www.jstor.org/stable/2335368
Cheville NF, McCullough DR, Paulson R (1998) Brucellosis in the Greater Yellowstone Area. National Academy Press
Cochran WG (1950) Estimation of bacterial densities by means of the "most probable number". Biometrics 6(2):105–116. http://www.jstor.org/stable/3001491
Cornfield J (1954) Measurement and comparison of toxicites: the quantal response. In: Kempthorne O, Bancroft TA, Gowen JW, Lush JL (eds) Statistics and mathematics in biology. Iowa State College Press, Ames
Cornfield J, Mantel M (1977) A discussion of "Estimation of safe doses in carcinogenic experiments" by Hartley and Sielken. Biometrics 33(1):21–24
Cross PC, Edwards WH, Scurlock BM, Maichak EJ, Rogerson JD (2007) Effects of management and climate on elk brucellosis in the Greater Yellowstone Ecosystem. Ecol Appl 17:957–964
Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data. Chapman & Hall/CRC, Boca Raton
Dellaportas P, Stephens DA (1995) Bayesian analysis of errors-in-variables regression models. Biometrics 51(3):1085–1095. http://www.jstor.org/stable/2533007
Elliott P, Wakefield J, Best N, Briggs D (2000) Spatial epidemiology. Oxford University Press, New York. iSBN 0-19-262941-7
Finney DJ (1978) Statistical method in biological assays. MacMillan, New York
Fisher R (1922) On the mathematical foundations of theoretical statistics. Philos Trans R Soc A 222:309–368
Gelman A, Chew GL, Shnaidman M (2004) Bayesian analysis of serial dilution assays. Biometrics 60(2):407–417. http://www.jstor.org/stable/3695768
Giltinan D, Davidian M (1994) Assays for recombinant proteins: a problem in non-linear calibration. Stat Med 13:1165–1179
Hamilton MA, Rinaldi MG (1988) Descriptive statistical analyses of serial dilution data. Stat Med 7:535–544
Lee MLT, Whitmore GA (1999) Statistical inference for serial dilution assay data. Biometrics 55(4):1215–1220
McCrady MH (1915) The numerical interpretation of fermentation-tube results. J Infect Dis 17(1):183–212. http://www.jstor.org/stable/30083495
McCullagh P, Nelder J (1989) Generalized linear models, 2nd edn. Monographs on statistics and applied probability, vol 37. Chapman & Hall, London
Mehrabi Y, Matthews JNS (1998) Implementable Bayesian designs for limiting dilution assays. Biometrics 54(4):1398–1406. http://www.jstor.org/stable/2533666
Morton JK, Thorne ET, Thomas GM (1981) Brucellosis in elk III: serologic evaluation. J Wildl Dis 17(1):23–31
Nelder J, Mead R (1965) A simplex algorithm for function minimization. Comput J 7:308–313
Plummer M (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003)
R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/, ISBN 3-900051-07-0
Ragan VE (2002) The Animal and Plant Health Inspection Service (APHIS) brucellosis eradication program in the united states. Veterinary Microbiol 90(1–4):11–18. doi:10.1016/S0378-1135(02)00240-7. http://www.sciencedirect.com/science/article/pii/S0378113502002407
Ridout M (2005) Serial dilution assay. Encycl Biostat 4079–4084. doi:10.1002/0470011815.b2a06022
Roffe TJ, Jones LC, Coffin K, Drew ML, Sweeney SJ, Hagius SD, Elzer PH, Davis D (2004) Efficacy of single calfhood vaccination of elk with Brucella abortus strain 19. J Wildl Manag 68:830–836
Smith BL (2001) Winter feeding of elk in western North America. J Wildl Manag 65:173–190
Thorne ET (2001) Brucellosis. In: Williams ES, Barker IK (eds) Infectious diseases of wild mammals. Iowa State Press, Ames, pp 372–395
Thorne E, Morton J, Blunt F, Dawson H (1978) Brucellosis in elk ii. Clinical effects and means of transmission as determined through artificial infections. J Wildl Dis 14:280–291
Thorne ET, Smith SG, Aune K, Hunter D, Roffe TJ (1997) Brucellosis: the disease in elk. In: Thorne ET, Boyce MS, Nicolletti P, Kreeger TJ (eds) Brucellosis, bison, elk, and cattle in the Greater Yellowstone Area: defining the problem, exploring solution. Wyoming Game and Fish Department, Cheyenne, pp 33–44

Jarrett J. Barber is an Assistant Professor of Statistics in the School of Mathematical and Statistical Sciences at Arizona State University. He received his B.S. in Forestry (1990), M.S. Forestry (1992), and M.S. Mathematics (1997) from Northern Arizona University. He received is Ph.D in Statistics from North Carolina State University (2002). He was a Visiting Assistant Professor of Statistics at the Institute of Statistics and Decision Sciences at Duke University and at the National Center for Atmospheric Research (2002–2004). His interests include spatial statistics and modeling of environmental and ecological phenomena.

Pritam Gupta is a Senior Biostatistician at Novartis Healthcare Pvt. Ltd at Hyderabad, India, supporting Exjade projects as a GMA Trial Statistician. He holds a B.S. in Statistics from the University of Calcutta, India, an M.S. in Statistics from the University of Calcutta, India, and a Ph.D. in Statistics from the University of Wyoming, USA.

William Edwards is a Wildlife Disease Specialist with the Wyoming Game and Fish Department in Laramie Wyoming. He received a B.S. in Molecular Biology in 1991, and an M.S. in Pathobiology in 1995 from the University of Wyoming. He works with many diseases in wildlife, but concentrates on brucellosis in elk and bison as well as pneumonia in bighorn sheep.

Kiona Ogle is an Associate Professor in the School of Life Sciences at Arizona State University. She received a B.S. in Biology and Mathematics (dual major) from Northern Arizona university in 1992 and an M.S. in Statistics and Ph.D. in Biology from Duke University in 2003. She did a PostDoc at Princeton University during 2003–2006, followed by an Assistant Professor position at the University of Wyoming (2006–2010). Her current research interests include quantifying plant and ecosystem responses to environmental perturbations, with an emphasis on applying Bayesian methods to synthesizing field, lab, and literature data in the context of process-based models.

Lance A. Waller is Rollins Professor and Chair of the Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University. He received his B.S. in Mathematics from New Mexico State University (1986), and his Ph.D. in Operations Research from Cornell University (1991). His research involves the development and application of statistical methods for spatially referenced data including applications in environmental justice, neurology, epidemiology, disease surveillance, and conservation biology. He has published in a variety of biostatistical and statistical journals and is co-author with Carol Gotway of the text Applied Spatial Statistics for Public Health Data (2004, Wiley).