Web Material

Effects of Accounting for Interval-Censored Antibody Titer Decay on Seroincidence in a Longitudinal Cohort Study of Leptospirosis

Katharine A. Owers Bonner, Jaqueline S. Cruz, Gielson A. Sacramento, Daiana de Oliveira, Nivison Nery, Mayara Carvalho, Federico Costa, James E. Childs, Albert I. Ko, and Peter J. Diggle

Table of Contents

Web Appendix: Statistical Model Formulation and Inference

To model titer decay, we postulated that the observed titer dilution K is an interval-censored version of a latent (unobservable) real-valued variable Z (response intensity) that is proportional to the underlying antibody level, W. The relationship between Z and K depends on the dilutions used. Here, $K = k$ corresponds to $d_k < Z < d_{k+1}$ for $k = 0, 1, \ldots, 7$ and $d_k = 0$, 50, 100, 320, 1000, 3200, 10,000, and 32,000, these being the dilutions used in Lupidi *et al.* (1).

Biological considerations suggest that, following an infection and absent re-infection, the mean of W should show an initial rapid increase followed by a slower decay. A simple model that meets these requirements is

$$m(t) = at^{\gamma}exp(-\beta t), \qquad (1)$$

where t denotes time since infection. To convert from a multiplicative to an additive scale, we define $Y = log(Z)$ and fit a model to Y. Since Z is proportional to W, it follows that $Y = logW + e$, where e is an unknown constant. Writing $\mu(x,t)$ to denote the mean of Y at time t, and taking logarithms in (1) gives

$$\mu(x, t) = x'\alpha - \beta t + \gamma \log t. \qquad (2)$$

In (2) we have assumed that the non-negative quantity a in (1) has a log-linear dependence on a vector of covariates x whose elements include terms for serovar (a factor on three levels) and individual (a factor on 18 levels), with corresponding regression parameters α. The value of the unknown constant e is absorbed into the intercept term of (2). Finally, we assume that Y is Normally distributed with mean $\mu(x,t)$ and variance σ^2, and that repeated measurements of Y on the same individual are stochastically independent.

Under the defined model, the probability of observing dilution $K = k$ at time t from an individual with covariates x_i is

$$p_k(x_i, t) = \int_{\log d_k}^{\log d_{k+1}} f(u; \mu(x_i, t), \sigma^2)du \qquad (3)$$

where $f(u;\ \mu, \sigma^2)$ is the probability density function of N(μ, σ^2) and, for dilution $K = 0$, the lower bound of the integral is $-\infty$. The log-likelihood contribution from each individual is the sum of log $p_k(t_j)$ over the follow-up times t_j = 1, 9, 18, 36, and 54 months. The overall log-likelihood is therefore

$$L(\theta) = \sum_{i=1}^{18}\sum_{j=1}^{5} \log(p_k(x_i, t_j)) \qquad (4)$$

where $\theta = (\alpha, \beta, \gamma, \sigma^2)$.

We estimated the model parameters by maximizing the log-likelihood (4) and tested for effects of serovar and individual using generalized likelihood ratio (deviance) tests. Table 1 summarizes the results of the generalized likelihood ratio tests, which confirm the need to include both serovar and individual effects in the model. The coefficient of log t is not significant, suggesting that any initial rise in T following infection occurred sufficiently rapidly that it cannot be

detected with a first follow-up time of t=1 month. Table 2 therefore shows parameter estimates and standard errors for the fitted model with γ=0. The R value is calculated as $exp(-\beta) = exp(-0.077) = 0.926$.

To check the model fit, we constructed residuals as follows. According to the model, an observed titer dilution, k_i has fitted value $f_i = \sum_{k=0}^{7} k p_k(x_i, t_i)$ and variance $v_i = \sum_{k=0}^{7} k^2 p_k(x_i, t_i) - f_i^2$. We therefore define standardized residuals, $r_i = (k_i - f_i)/\sqrt{v_i}$.

Appendix Figure 1 shows two diagnostic residual plots. The first is a scatterplot of the standardized residuals r_i against the fitted values f_i. It shows no systematic structure other than the banding that is characteristic of residuals when the response variable is discrete-valued, indicating an acceptable fit. The second plot shows the scatter of the standardized residuals within each individual, distinguishing between serovars. This shows no systematic difference between individuals or between serovars within individuals.

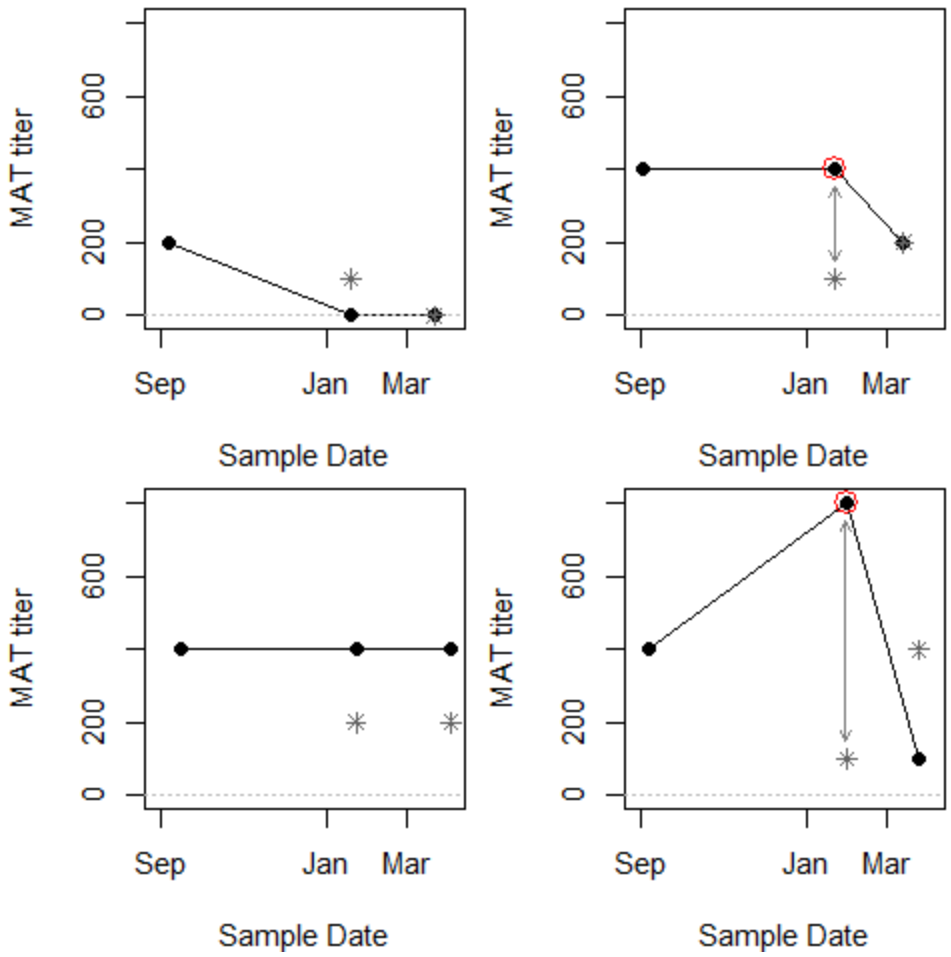
The data from Lupidi *et al.*(1) and code for these analyses can be found at https://github.com/kowers/titer-decay-modeling.

Web Table 1. Generalized likelihood ratio tests for inclusion of covariates. The deviance statistic to compare two nested models is $D = 2(L_1 - L_0)$ where L_0 and L_1 are the maximized log-likelihoods under the null and alternative hypotheses. The null distribution of D is chi-squared, with degrees of freedom (df) equal to the difference between the numbers of parameters in the two models.

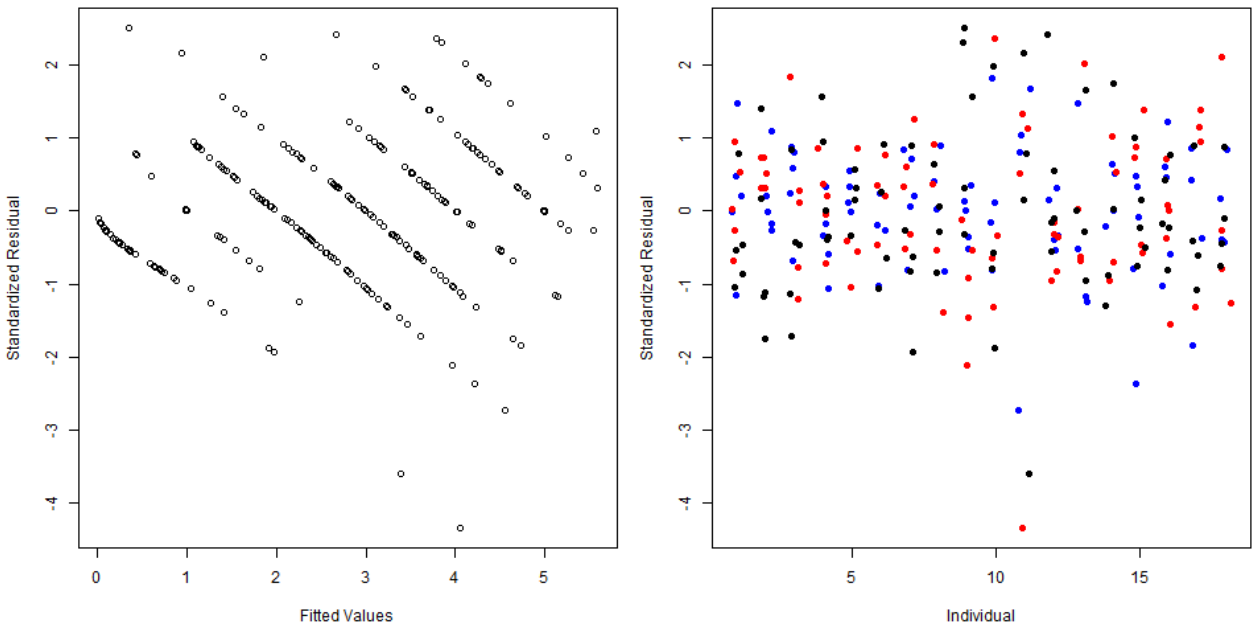
Regression model	log L(θ)	Deviance	df	p-value
Time	-390.1199			
Time + log(Time)	-389.6177	1.00444	1	0.316
Time + Serovar	-373.3673	33.5053	2	< 0.001
Time + Serovar + Individual	-303.7657	139.203	17	< 0.001

Web Table 2. Parameter estimates and standard errors for the fitted model. Model parameters are the 18 individual effects (parameters 1-18), serovar effects as contrasts of serovars bratislava and lora against serovar australis, the rate of exponential decay (β), and the error variance (σ^2).

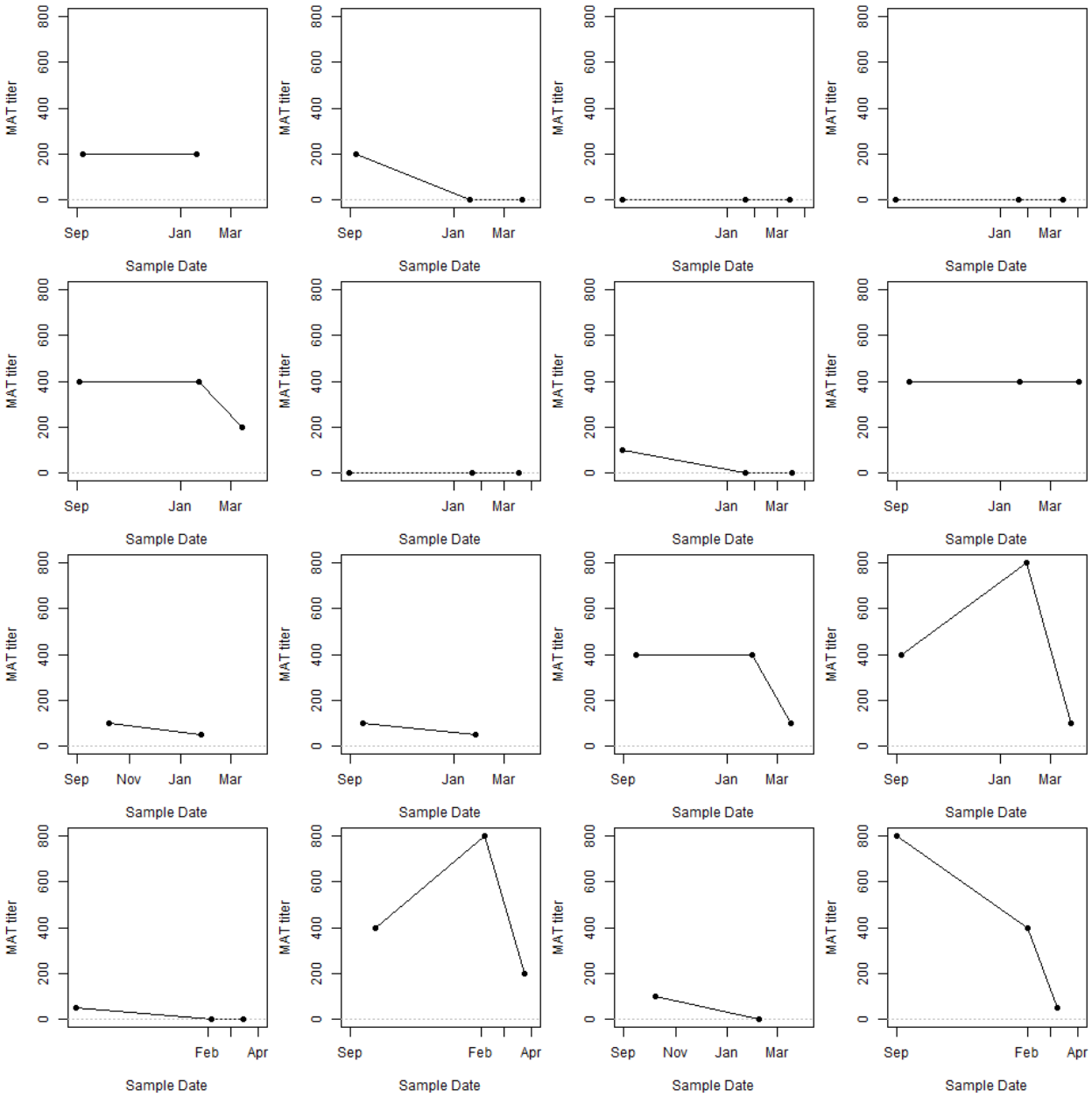
Parameter	Estimate	Standard error
1	3.630	0.313
2	5.020	0.314
3	3.208	0.320
4	3.109	0.322
5	3.484	0.336
6	3.827	0.336
7	2.692	0.329
8	1.340	0.362
9	3.473	0.315
10	2.639	0.340
11	2.957	0.334
12	2.132	0.335
13	3.022	0.333
14	4.079	0.335
15	2.553	0.326
16	2.266	0.326
17	3.155	0.335
18	0.487	0.393
Serovar bratislava	0.767	0.177
Serovar lora	1.351	0.176
β	0.077	0.005
σ^2	1.027	0.056



Web Figure 1: Decayed MAT titer values and resulting infections defined for participants in a longitudinal cohort study in Pau da Lima, Brazil. These panels represent titer trajectories for four participants from the Pau da Lima cohort subset sampled with intervals of approximately 4 then 2 months and are used to demonstrate our method of allowing for titer decay. Points in black are the measured titer values. Grey asterisks represent a decayed version of the previous titer (a single iteration is shown for clarity). No infections were defined among participants using the conventional definition, but two infections (marked with red circles) were identified using definitions allowing for titer decay. The arrows indicate the comparison: a decayed version of the previous titer is compared to the observed titer at the end of the inter-sample interval.



Web Figure 2. Diagnostic residual plots. The left-hand panel is a scatterplot of standardized residuals against fitted values. The right-hand plot shows the distributions of the sets of 15 standardized residuals from each of the 18 individuals, color-coded according to serovar (black: australis; red: bratislava; blue: lora).



Web Figure 3. MAT titer values for Pau da Lima cohort participants. A subset of participants in the biannual cohort were sampled with intervals of approximately 4 then 2 months.

Reference

1. Lupidi R, Cinco M, Balanzin D, *et al.* Serological follow-up of patients involved in a localized outbreak of leptospirosis. J Clin Microbiol. 1991;29(4):805-9.