Seroincidence package methodology

European Centre for Disease Prevention and Control (ECDC)

2018-06-05

Contents

Introduction **2**

1. A simple model for the seroresponse **2**

2. Backward recurrence time **2**

3. Incidence of seroconversions **3**

4. True seronegative subjects **5**

5. Censored observations **6**

References **8**

Introduction

The revised seroincidence calculator package provides three refinements to the method for calculating seroincidence published earlier (Teunis et al. 2012) and implemented in R package **seroincidence**, versions 1.x: (1) inclusion of infection episode with rising antibody levels, (2) non–exponential decay of serum antibodies after infection, and (3) age–dependent correction for subjects who have never seroconverted. It is important to note that, although the implemented methods use a specific parametric model, as proposed in (de Graaf et al. 2014) and augmented in (Teunis et al. 2016), the methods used to calculate the likelihood function allow seroresponses of arbitrary shape.

1. A simple model for the seroresponse

The current version of the seroincidence package uses the model of (Teunis et al. 2016) for the shape of the seroresponse:

Infection/colonization episode	Waning immunity episode
$b'(t) = \mu_0 b(t) - cy(t)$	$b(t) = 0$
$y'(t) = \mu y(t)$	$y'(t) = -\nu y(t)^r$

With baseline antibody concentration $y(0) = y_0$ and initial pathogen concentration $b(0) = b_0$. The serum antibody response $y(t)$ can be written as

$$y(t) = y_+(t) + y_-(t)$$

where

$$y_+(t) = y_0 \mathrm{e}^{\mu t}[0 \le t < t_1]$$
$$y_-(t) = y_1 \left(1 + (r-1) y_1^{r-1} \nu (t - t_1)\right)^{-\frac{1}{r-1}} [t_1 \le t < \infty]$$

Since the peak level is $y_1 = y_0 \mathrm{e}^{\mu t_1}$ the (unobservable) growth rate μ can be written as

$$\mu = \frac{1}{t_1} \log\left(\frac{y_1}{y_0}\right)$$

Antibody decay is different from exponential (log–linear) decay. When the shape parameter $r > 1$, log concentrations decrease rapidly after infection has terminated, but decay then slows down and low antibody concentrations are maintained for a long period. When r approaches 1, exponential decay is restored.

2. Backward recurrence time

Considering the (fundamental) uniform distribution u_f of sampling within a given interval, the interval length distribution $p(\Delta t)$ and the distribution of (cross–sectionally) sampled interval length (Teunis et al. 2012)

$$q(\Delta t) = \frac{p(\Delta t) \cdot \Delta t}{\overline{\Delta t_p}}$$

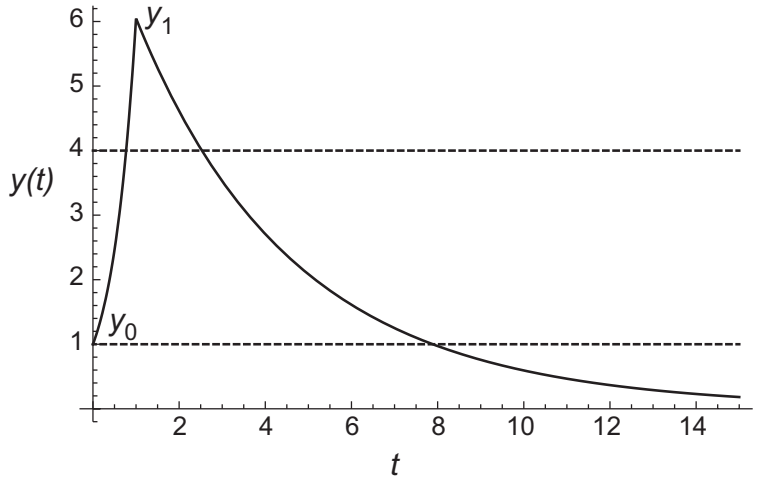

Figure 1: The antibody level at $t = 0$ is y_0; the rising branch ends at $t = t_1$ where the peak antibody level y_1 is reached. Any antibody level $y_0 \leq y(t) < y_1$ occurs twice.

the joint distribution of backward recurrence time and cross–sectional interval length is the product $u_f \cdot q$ because these probabilities are independent.

The distribution of backward recurrence time is the marginal distribution

$$\begin{aligned} u(\tau) &= \int_{\Delta t=0}^{\infty} u_f(\tau; \Delta t) \cdot q(\Delta t) \mathrm{d}\Delta t \\ &= \int_0^{\infty} \frac{[0 \leq \tau \leq \Delta t]}{\Delta t} \cdot \frac{p(\Delta t) \cdot \Delta t}{\overline{\Delta t_p}} \mathrm{d}\Delta t \\ &= \frac{1}{\overline{\Delta t_p}} \int_{\tau}^{\infty} p(\Delta t) \mathrm{d}\Delta t \end{aligned}$$

3. Incidence of seroconversions

To calculate the incidence of seroconversions, as in (Teunis et al. 2012), the distribution $p(\Delta t)$ of intervals Δt between seroconversions, is important. Assuming any subject is sampled completely at random during their intervals between seroconversions, and accounting for interval length bias (Satten et al. 2004; Zelen 2004), the distribution of backward recurrence times τ can be written as (Teunis et al. 2012)

$$u(\tau) = \frac{1}{\overline{\Delta t}} \int_{\tau=0}^{\infty} p(\Delta t) \mathrm{d}\Delta t = \frac{1 - P(\Delta t)}{\overline{\Delta t}}$$

where $\overline{\Delta t}$ is the p–distribution expected value of Δt.

This density is employed to estimate seroconversion rates. The antibody concentration y is the observable quantity, and we need to express the probability (density) of observed y in terms of the density of backward recurrence time.

First, the backward recurrence time can τ be expressed as a function of the serum antibody concentration y

$$\tau(y) = \tau_+(y) + \tau_-(y)$$

where

$$\tau_+(y) = \frac{1}{\mu}\log\left(\frac{y_+}{y_0}\right)[0 \le \tau < t_1]$$
$$\tau_-(y) = \left(t_1 + \frac{y_-^{-(r-1)} - y_1^{-(r-1)}}{\nu(r-1)}\right)[t_1 \le \tau < \infty]$$

with corresponding derivatives

$$\frac{\mathrm{d}\tau_+}{\mathrm{d}y} = \frac{1}{\mu y_+} \quad \text{and} \quad \frac{\mathrm{d}\tau_-}{\mathrm{d}y} = -\frac{1}{\nu y_-^r}$$

Now, consider the probability that an antibody level y', corresponding to a time since infection τ', is less than or equal to y (see Figure 1)

$$\begin{aligned} P(y' \le y) &= P\left(y_0 \le y_+(\tau') \le y_+(\tau) \vee y_-(\tau') \le y_-(\tau) \le t_1\right) + [y_1 < y] \\ &= P\left(0 \le \tau' \le \frac{1}{\mu}\log\left(\frac{y_+}{y_0}\right)\right) \\ &\quad + P\left(t_1 + \frac{y_-^{-(r-1)} - y_1^{-(r-1)}}{\nu(r-1)} \le \tau' < \infty\right) + [y_1 < y] \end{aligned}$$

The probability density for y then is

$$\begin{aligned} \rho(y) &= \frac{\mathrm{d}}{\mathrm{d}y}P(y' \le y) \\ &= \frac{\mathrm{d}}{\mathrm{d}\tau_+}\frac{\mathrm{d}\tau_+}{\mathrm{d}y_+}P\left(0 \le \tau' \le \tau_+(y)\right) + \frac{\mathrm{d}}{\mathrm{d}\tau_-}\frac{\mathrm{d}\tau_-}{\mathrm{d}y_-}P\left(t_1 + \tau_-(y) \le \tau' < \infty\right) \\ &= \rho_+(y_+) + \rho_-(y_-) \end{aligned}$$

So that

$$\begin{aligned} \rho_+(y_+) &= \frac{1}{\mu y_+}u\left(\frac{1}{\mu}\log\left(\frac{y_+}{y_0}\right)\right) \\ \rho_-(y_-) &= \frac{1}{\nu y_-^r}u\left(t_1 + \frac{y_-^{-(r-1)} - y_1^{-(r-1)}}{\nu(r-1)}\right) \end{aligned} \tag{1}$$

when $[y_0 \le y < y_1]$ there are two contributions to the density, one from the rising and one from the decaying branch of the antibody response.

If, as assumed before (Teunis et al. 2012), intervals between incident infections are generated by a process with Gamma probability density, $\overline{\Delta t} = (m+1)/\lambda$. The cumulative distribution function for τ is

$$P_m(\tau) = 1 - \frac{\Gamma(m+1, \lambda\tau)}{m!} \tag{2}$$

and the density of backward recurrence times is

$$u_m(\tau) = \frac{1 - P_m(\tau)}{\overline{\Delta t}} = \frac{\lambda\Gamma(m+1, \lambda\tau)}{(m+1)!} = \frac{\lambda \mathrm{e}^{-\lambda\tau}}{m+1}\sum_{j=0}^{m}\frac{(\lambda\tau)^j}{j!} \tag{3}$$

Combining equations (1) and (3) the marginal density of y can be found

$$\rho_+(y_+) = [y_0 \le y_+ < y_1] \frac{\lambda y_0}{\mu(m+1)} \left(\frac{y_+}{y_0}\right)^{-(1+\lambda/\mu)} \sum_{j=0}^{m} \frac{\left(\frac{\lambda}{\mu}\log(y_+/y_0)\right)^j}{j!} \tag{4}$$

and

$$\begin{aligned}\rho_-(y_-) = [0 < y_- \le y_1] &\frac{\lambda}{\nu y_-^r (m+1)} \mathrm{e}^{-\lambda\left(t_1 + \frac{y_-^{-(r-1)} - y_1^{-(r-1)}}{\nu(r-1)}\right)} \\ &\times \sum_{j=0}^{m} \frac{\lambda^j}{j!} \left(t_1 + \frac{y_-^{-(r-1)} - y_1^{-(r-1)}}{\nu(r-1)}\right)^j \end{aligned} \tag{5}$$

where ρ_+ and ρ_- refer to the contributions of the rising and decaying part of the seroresponse to the density of y.

The within–host parameters $\boldsymbol{\theta} = (y_0, y_1, t_1, \nu, r)$ vary among responses of individual subjects. Heterogeneity in these parameters may be described by their joint distribution, which can be used to calculate the marginal distribution $\rho(y)$ (Teunis et al. 2012). Since a Monte Carlo sample of the posterior joint distribution is available from the longitudinal model (Teunis et al. 2016) the marginal distribution of $\rho(y)$ may be approximated by the sum

$$\rho(y) = \frac{1}{N} \sum_{n=1}^{N} \rho_+(y|\lambda, m, \boldsymbol{\theta}_n) + \rho_-(y|\lambda, m, \boldsymbol{\theta}_n) \tag{6}$$

for a Monte Carlo sample $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N)$. A cross–sectional sample of antibody concentrations $(Y_1, Y_2, \ldots, Y_K)$ can now be used to calculate a likelihood

$$\ell(\lambda, m) = \prod_{k=1}^{K} \rho(Y_k|\lambda, m)$$

that can be used to estimate (λ, m).

4. True seronegative subjects

At the time that a cross–sectional serum sample is collected, the subject whose blood is drawn may have never been infected in their lifetime. The antibody concentration y in that sample then represents a true seronegative. For such a sample, the backward recurrence time does not exist. For a given longitudinal response, the backward recurrence time τ corresponding with measured antibody concentration Y can be calculated. If that backward recurrence time is longer than the age of the person (at time of sampling), their antibody concentration Y cannot have resulted from prior seroconversion.

If, in the summation in equation (6), terms not satisfying the condition $\tau(y) < a$ are discarded, the resulting partial sum

$$\rho^*(y, a) = \frac{1}{N} \sum_{n=1}^{N} [\tau(y, \boldsymbol{\theta}_n) < a](\rho_+(y|\lambda, m, \boldsymbol{\theta}_n) + (\rho_-(y|\lambda, m, \boldsymbol{\theta}_n)$$

counts only those seroconversions that can have occurred during the lifespan of the person whose serum was sampled.

Serum from a true seronegative subject is expected to have a low antibody concentration, representative of a "true" negative sample (de Greeff et al. 2012). The antibody concentrations y in sera from such truly negative subjects are not expected to decay over time: the baseline distribution $\rho_0(y)$ may be assumed fixed and independent of m and λ. Note that also when y corresponds to a backward recurrence time within the lifespan of a person, that same antibody concentration y could still result from the baseline distribution $\rho_0(y)$.

Given the interval distribution for incident infections, the probability that a sampled subject has never seroconverted depends on their age. For the gamma process assumed above, the survival function $P_m(a|m,\lambda)$ gives the probability of a subject having not seroconverted before age a (equation (2)).

Thus, for a serum sample with antibody concentration y from a subject of age a the probability density is

$$\psi(\lambda, m|y, a, \boldsymbol{\theta}, \boldsymbol{\theta}_0) = \rho^*(y, a) + P_m(a|m, \lambda)\rho_0(y|\boldsymbol{\theta}_0)$$

When the seroconversion rate is low, or a subject is young, or both, the probability of a true negative may be considerable.

5. Censored observations

In case observations are censored at y_c such that an observed $Y = \max(Y, y_c)$, then for $y_c < Y$ the density $\rho(y)$ as in equations (4) and (5) holds, but the likelihood of any $Y \leq y_c$ can be calculated

$$\ell(\lambda, m|y \leq y_c) = \int_{z=0}^{y_c} \rho(z, \lambda, m)\mathrm{d}z = R(y_c|\lambda, m)$$

We need the cumulative distribution $R(y)$ for the backward recurrence time, from equation (3).

$$U_m(\tau) = \frac{1}{m+1}\sum_{j=0}^{m} P_j(\tau|\lambda)$$

while the cumulative distribution of antibody levels y is

$$\begin{aligned} P(y' \leq y) = U\left(\frac{1}{\mu}\log\left(\frac{y}{y_0}\right)\right)[y_0 \leq y < y_1] + \\ \left(1 - U\left(t_1 + \frac{y^{-(r-1)} - y_1^{-(r-1)}}{\nu(r-1)}\right)\right)[y \leq y_1] + [y_1 < y] \end{aligned}$$

so that

$$\begin{aligned} R(y) = \frac{1}{m+1}\sum_{j=0}^{m} P_j\left(\frac{1}{\mu}\log\left(\frac{y}{y_0}\right)|\lambda\right)[y_0 \leq y < y_1] \\ + \left(1 - \frac{1}{m+1}\sum_{j=0}^{m} P_j\left(t_1 + \frac{y^{-(r-1)} - y_1^{-(r-1)}}{\nu(r-1)}|\lambda\right)\right)[y \leq y_1] \\ + [y_1 < y] \end{aligned}$$

Using equation (2) this can be written in terms of incomplete gamma functions

$$
\begin{aligned}
R(y) = \frac{1}{m+1}\sum_{j=0}^{m}\left(1-\frac{\Gamma\left(j+1,\frac{\lambda}{\mu}\log\left(\frac{y}{y_0}\right)\right)}{j!}\right)[y_0 \le y < y_1] \\
+\left(1-\frac{1}{m+1}\sum_{j=0}^{m}\left(1-\frac{\Gamma\left(j+1,\lambda\left(t_1+\frac{y^{-(r-1)}-y_1^{-(r-1)}}{\nu(r-1)}\right)\right)}{j!}\right)\right)[y < y_1] \\
+[y_1 < y]
\end{aligned}
$$

or

$$
\begin{aligned}
R(y) = \left(1-\frac{1}{m+1}\sum_{j=0}^{m}\frac{\Gamma\left(j+1,\frac{\lambda}{\mu}\log\left(\frac{y}{y_0}\right)\right)}{j!}\right)[y_0 \le y < y_1] \\
+\frac{1}{m+1}\sum_{j=0}^{m}\frac{\Gamma\left(j+1,\lambda\left(t_1+\frac{y^{-(r-1)}-y_1^{-(r-1)}}{\nu(r-1)}\right)\right)}{j!}[y < y_1] \\
+[y_1 < y]
\end{aligned}
$$

References

de Graaf, W. F., M. E. E. Kretzschmar, P. F. M. Teunis, and O. Diekmann. 2014. "A Two-Phase Within-Host Model for Immune Response and Its Application to Serological Profiles of Pertussis." *Epidemics* 9 (December): 1–7. https://doi.org/10.1016/j.epidem.2014.08.002.

de Greeff, S. C., P. F. M. Teunis, H. E. de Melker, F. R. Mooi, D. W. Notermans, B. Elvers, and J. F. P. Schellekens. 2012. "Two-Component Cluster Analysis of a Large Serodiagnostic Database for Specificity of Increases of IgG Antibodies Against Pertussis Toxin in Paired Serum Samples and of Absolute Values in Single Serum Samples." *Clinical and Vaccine Immunology* 19 (9). https://doi.org/10.1128/CVI.00229-12.

Satten, G. A., F. Kong, D. J. Wright, S. A. Glynn, and G. B. Schreiber. 2004. "How Special Is a 'Special' Interval: Modeling Departure from Length-biased Sampling in Renewal Processes." *Biostatistics* 5 (1). https://doi.org/10.1093/biostatistics/5.1.145.

Teunis, P. F. M., J. C. van Eijkeren, C. W. Ang, Y. T. van Duynhoven, J. B. Simonsen, M. A. Strid, and W. van Pelt. 2012. "Biomarker Dynamics: Estimating Infection Rates from Serological Data." *Statistics in Medicine* 31 (20): 2240–8. https://doi.org/10.1002/sim.5322.

Teunis, P. F. M., J. C. van Eijkeren, W. F. de Graaf, A. Bonačić Marinović, and M. E. E. Kretzschmar. 2016. "Linking the Seroresponse to Infection to Within-Host Heterogeneity in Antibody Production." *Epidemics* 16 (September). https://doi.org/10.1016/j.epidem.2016.04.001.

Zelen, M. 2004. "Forward and Backward Recurrence Times and Length Biased Sampling: Age Specific Models." *Lifetime Data Analysis* 10 (4).