Epidemiol. Infect. (2011), **139**, 636–643. © Cambridge University Press 2010
doi:10.1017/S0950268810000750

Usefulness of seroconversion rates for comparing infection pressures between countries

J. SIMONSEN[1]*, P. TEUNIS[2,3], W. VAN PELT[2], Y. VAN DUYNHOVEN[4], K. A. KROGFELT[5], M. SADKOWSKA-TODYS[6] AND K. MØLBAK[1]

[1] *Division of Epidemiology, Statens Serum Institut, Copenhagen, Denmark*
[2] *Epidemiology and Surveillance Unit, Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands*
[3] *Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta GA, USA*
[4] *Laboratory for Zoonoses and Environmental Microbiology, Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands*
[5] *Department of Bacteriology, Mycology and Parasitology, Statens Serum Institut, Copenhagen, Denmark*
[6] *Department of Epidemiology, National Institute of Public Health – National Institute of Hygiene, Warszawa, Poland*

(*Accepted 16 March 2010; first published online 12 April 2010*)

SUMMARY

Salmonella is a frequent cause of foodborne illness. However, since most symptomatic cases are not diagnosed, the true infection pressure is unknown. Furthermore, national surveillance systems have different sensitivities that limit inter-country comparisons. We have used recently developed methods for translating measurements of *Salmonella* antibodies into estimates of seroincidence: the frequency of infections including asymptomatic cases. This methodology was applied to cross-sectional collections of serum samples obtained from the general healthy population in three European countries. Denmark and The Netherlands had the lowest seroincidence (84169 infections/1000 person-years), whereas Poland had the highest seroincidence (547/1000 person-years). A Bayesian method for obtaining incidence rate ratios was developed; this showed a 6·3 (95% credibility interval 3·3–12·5) higher incidence in Poland than in Denmark which demonstrates that this methodology has a wider applicability for studies of surveillance systems and evaluation of control programmes.

Key words: ELISA, mathematical modelling, salmonellosis, serology.

INTRODUCTION

A general problem with infectious disease surveillance systems is that due to under-ascertainment, estimation of incidences based on numbers of reported laboratory-confirmed cases will often lead to considerable underestimation. This is especially true for self-limiting diseases with symptoms such as diarrhoea caused by foodborne bacteria [1]. Several barriers have to be crossed before an infected person is reported in a surveillance system, e.g. the person has to seek healthcare, a stool sample has to be taken, the bacteria should be identified by a laboratory [2], and the laboratory should report the positive finding.

* Author for correspondence: J. Simonsen, M.Sc., Division of Epidemiology, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark.
(Email: cob@ssi.dk)

A Dutch study has estimated that only one out of 14·3 *Salmonella* cases in the general population was reported to the laboratory surveillance system [3]. Since both the organization of healthcare and surveillance systems differ strongly between countries, the numbers of reported cases cannot be used to compare incidences as a measure for the disease burden across countries. Clearly we are in need of alternative, less biased methods to evaluate the incidence of salmonellosis and other diseases since official systems only capture a small and unknown fraction of case patients.

Community studies have been used in an attempt to estimate the true burden of gastroenteritis by use of questionnaires and the microbiological analysis of systematically collected stool samples from prospectively followed population cohorts [3, 4]. However, such community studies are very expensive and may therefore be difficult to apply in most countries. In the current study we use measurements of antibodies in representative serum samples randomly selected from the population to estimate the frequency of seroconversions against a certain pathogen. We show that these can be used to obtain a valid comparison of infection pressure between countries. This approach also benefits from being more cost-effective than traditional studies involving the collection of faecal samples, since it can use sera from well-defined cohorts collected for other purposes.

Blood samples from cross-sectional population studies in three European countries were gathered as part of a European collaborative project [5]. After measuring antibodies against *Salmonella* we were able to estimate the incidence of *Salmonella* seroconversions in the participating countries and make comparisons. The method presented here has the advantage of being completely independent of the healthcare and surveillance systems and is therefore well suited for comparing the incidence of *Salmonella* exposures between countries and periods. Moreover, this methodology can provide a basis for estimating the degree of underreporting, e.g. evaluating the ratio (sometimes referred to as the *multiplier estimate* [6]) between the seroincidence and the incidence of culture-confirmed cases.

We applied a mathematical model from a previously published study aimed at estimating the frequency of seroconversions [7], a measure termed as the seroincidence. Our paper will show the results of applying this method to estimate the seroincidence in cohorts from different countries and periods. We further extended the model such that the ratio (with corresponding credible intervals) between pairs of incidences could be estimated.

METHODS

Materials

Longitudinal study

A follow-up study of 302 persons infected with either *Salmonella* Enteritidis or *Salmonella* Typhimurium was performed [8]. Blood samples were taken three or four times in a period of 18 months after onset of infection. Anti-*Salmonella* IgA, IgM and IgG concentrations were measured in arbitrary units of optical density (OD) values in an in-house mixed ELISA, using lipopolysaccharides of *S.* Enteritidis + *S.* Typhimurium as capture antigens [8].

Cross-sectional studies

Population-representative sera were collected in three countries: Denmark, Poland and The Netherlands. The Danish sera were obtained from the Helbred 2006 cohort and consisted of 1780 blood samples collected in 2006 and 2007. The Polish sera came from Bank Surowic Zakladu Wirusologii PZH and consisted of 500 samples collected in 2004 [9]. In The Netherlands we included sera from two cohorts: the first was the Regenboog cohort for which 1053 blood samples were collected in 1998–2002 while the second, the Pienter II cohort [10], consisted of 1065 blood samples collected in 2006 and 2007.

Antibody titres in each blood sample from the cross-sectional samples were measured in the same units by the same assay as in the longitudinal study.

Estimation of seroprevalence

In order to define the threshold for being seropositive, the 95% percentile was calculated for each of the antibody classes (IgG, IgM, IgA) in the Danish cross-sectional cohort and these values were used as cut-offs. The antibody levels observed in the three other countries were then compared with this cut-off value. The seroprevalence in each of the four cohorts was estimated by the fractions of seropositive samples. This was done separately for each of the three antibody classes (IgG, IgM, IgA).

Estimation of seroincidence

The seroincidence is here defined as the number of seroconversions per (1000) number of person-years. Under the assumption of a rapid rise in antibody levels following infection and a subsequent slow decay, we designed a mathematical model. In order to handle individual variation of peak level, decay rate and baseline level, the model had a two-level hierarchical structure where the parameters describing the individual response curves were allowed to be random components with global mean and variance parameters. Inference of these parameters was obtained from the longitudinal study [7]. Using the estimated parameters it was then possible to predict the antibody level for any number of days since a person experienced their last infection. Conversely, given a set of antibody measurements (IgG, IgM, IgA), it is then possible to back-calculate the time elapsed since a last infection. This was done by Monte Carlo simulation, producing sets of time since infection, $\{T_i\}_j$, where j refers to any individual simulation and i to an individual in any of the cross-sectional samples.

A Bayesian approach was used for the estimation of seroincidence. This means that we considered the seroincidence as a random variable of which we aimed to find the conditional distribution (commonly called the posterior distribution) given the observed data. A requirement for doing so is definition of a distribution of the model parameter(s) prior to any observations (the prior distribution). In a situation with no prior information (which most often is the case) it is often possible to define a non-informative prior distribution reflecting complete ignorance about the model parameters. Then the posterior distribution depends (entirely) on the information contained in the observed data.

The approach used here is explained in detail in Simonsen *et al*. [7]. Inference of the incidence is based on the posterior distribution of the incidence given the observed antibody values. Since an incidence is treated as a scale parameter, the non-informative prior distribution should be flat on the log-scale. Therefore, the non-informative prior distribution of the incidence is given by the improper probability density function, $\pi(\gamma) \propto 1/\gamma$, where γ is the unknown incidence.

We assumed first that the times since infection are known (T_i for individual i) for each individual in the cross-sectional cohorts. Due to the fact that the antibody levels for IgM and IgA appeared to reach steady state after 60 days, we chose to censor estimated time since infection at 60 days [7]. The conditional distribution of incidence given the time values is then given by

$$\begin{aligned} p(\gamma|T_1, \cdots, T_N) &\propto \pi(\gamma) \prod_{i=1}^{N} p(T_i|\gamma) \\ &= \pi(\gamma) \prod_{i=1}^{N} \gamma^{1_{T_i<60}} \mathrm{e}^{-\gamma \min(T_i, 60)} \\ &= \gamma^{\left(\sum_{i=1}^{N} 1_{T_i<60}\right)-1} \mathrm{e}^{-\gamma \sum_{i=1}^{N} \min(T_i, 60)} \end{aligned} \tag{1}$$

where $1_{T_i<60}$ takes the value 1 if individual i was infected within the last 60 days and zero otherwise. Note that the last term shows that γ is Gamma-distributed (after conditioning with time since infection). This implies that the mean value (which can be used as an estimate) of γ is

$$\frac{\sum_{i=1}^{N} 1_{T_i<60}}{\sum_{i=1}^{N} \min(T_i, 60)},$$

which intuitively is correct; the number of observed events divided by the total observed time between infection/censoring event and observation event is the commonly used estimator of an incidence.

However, only the antibody values are known – not the actual time since infection. To overcome this problem, we can simulate sets of the time since infection from their conditional distribution given the observed antibody values averaging over the distributions where we condition with the simulated time values. These simulations were performed by construction of a Markov Chain [11]. The posterior distribution of the incidence is therefore

$$\begin{aligned} p(\gamma|\text{data}) &= E_{\{T_i\}_N|\text{data}}\, p(\gamma|\{T_i\}, \text{data}) \\ &= E_{\{T_i\}|\text{data}}\, p(\gamma|\{T_i\}) \\ &\approx \frac{1}{M} \sum_{j=1}^{M} p(\gamma|\{T_i\}_j) \end{aligned} \tag{2}$$

where $\{T_i\}$ (with no subscript j) refers to the unknown random variable time since infection while $\{T_i\}_j$ refers to the jth simulated set of values of time since infection and M is the number of simulated sets of time since infection. The second part of equation (2) follows from the fact that for given values of time since infection, antibody levels do not provide any further information about the incidence. Convergence of the Markov chains was tested by verifying that there were no significant differences between different parts

https://doi.org/10.1017/S0950268810000750 Published online by Cambridge University Press

(of various sizes) of the chains (Geweke's test) [11, 12]. We especially verified if the selected burn-in period was long enough by testing for significant difference between samples in the initial part of Markov chains directly after the burn-in period (the burn-in period was the initial 5000 iterations in the chain, that were discarded) and samples in the final part of the chain. Further, the Markov chains were thinned in order to approach independent samples. This was done by verifying that all the autocorrelations in the thinned chains were insignificantly different from zero.

By using equation (1) it can be seen that the posterior distribution of the incidence can be estimated by a mixture of Gamma distributions. Note that the observed antibody values only affect the posterior distribution through the simulated time values which were simulated in the conditional distribution given the antibody levels.

The distribution given by equation (2) allows estimation of the median incidence as well as 95% credibility intervals† as the 2·5% and 97·5% percentiles.

Incidence comparison

Pairwise comparisons of the seroincidences between the four cohorts were made by constructing the posterior distribution of the incidence rate ratios. The posterior distribution of the ratio of two incidences could be constructed analytically. After calculating the posterior distribution of the incidence in the two countries under consideration, we simply need to find the distribution of the ratio of two random variables which are distributed as the posterior distributions of the pairs of incidences from the two countries.

We assumed that X and Y are two independent stochastic variables which are both Gamma-distributed with shape and scale parameters (α_X, λ_X) and (α_Y, λ_Y), respectively.

The distribution of the ratio of two stochastic variables, X and Y can be calculated as

$$f_{X/Y}(z) = \int y f_X(zy) f_Y(y)\,\mathrm{d}y,$$

where the f_X and f_Y are probability mass functions for the stochastic variables X and Y. When both X and Y are Gamma-distributed the following distribution is obtained for their ratio:

$$f_{X/Y}(z) = \int_0^\infty y \frac{\lambda_X^{\alpha_X}(zy)^{\alpha_X-1}\mathrm{e}^{-\lambda_X zy}}{\Gamma(\alpha_X)} \frac{\lambda_Y^{\alpha_Y} y^{\alpha_Y-1}\mathrm{e}^{-\lambda_Y y}}{\Gamma(\alpha_Y)}\mathrm{d}y$$
$$= \frac{\Gamma(\alpha_X+\alpha_Y)}{\Gamma(\alpha_X)\Gamma(\alpha_Y)} \lambda_X^{\alpha_X}\lambda_Y^{\alpha_Y} z^{\alpha_X-1}(\lambda_Y+\lambda_X z)^{-\alpha_X-\alpha_Y}. \quad (3)$$

In the present case, X and Y, represent the incidences in two different countries that are not Gamma-distributed but rather mixtures of Gamma-distributed variables:

$$X = \frac{1}{M}\sum_{i=1}^{M} X_i \quad \text{and} \quad Y = \frac{1}{M}\sum_{i=1}^{M} Y_i,$$

where X_i and Y_i are, respectively, $\Gamma(\alpha_{Xi}, \lambda_{Xi})$ and $\Gamma(\alpha_{Yi}, \lambda_{Yi})$ distributed.

The distribution of $f_{X/Y}(z)$ is therefore given by probability mass function of the form

$$\frac{1}{M^2}\sum_{j=1}^{M}\sum_{i=1}^{M} f_{Xj/Yi}(z),$$

where $f_{Xi/Yi}(z)$ is given by equation (3).

Using this distribution, median values and 95% credibility intervals can be calculated for the incidence rate ratios.

Applied software

The Markov chain used for estimating the longitudinal parameters was produced in WinBugs [13]. We wrote our own procedures for constructing the Markov chains for estimating the time since infection, the posterior distributions of the incidence rates, and the posterior distributions of the incidence rate ratios as we were unable to find software capable of performing these tasks. This was done in SAS language [14].

RESULTS

Seroprevalences

The distributions of the observed OD values are shown in Figure 1. For all three antibody types the samples from Poland had the largest fraction of high OD values ($\geqslant$0·4) while Denmark had the lowest fraction of high OD values. Crude seroprevalence estimates in each of the three cross-sectional samples are shown in Table 1. The seroprevalence was lowest

† In a Bayesian framework, 'credibility intervals' are the commonly used term for intervals showing the accuracy of the estimates. Basically, these are equivalent to confidence intervals.

https://doi.org/10.1017/S0950268810000750 Published online by Cambridge University Press

Table 1. *Estimated seroprevalences and seroincidences in the participating countries*

Country	Study name	Sample size	Sample year	Seroprevalence			Seroincidence per 1000 person-years (95% credibility interval)
				IgG	IgM	IgA	
Denmark	Helbred 2006	1780	2006–2007	5%	5%	5%	84 (41–141)
The Netherlands	Pienter [10]	1053	2006–2007	12·2%	5·3%	6·1%	149 (78–245)
The Netherlands	Regenboog	1065	1998–2002	19·1%	7·1%	7·4%	169 (91–271)
Poland	Bank Surowic Zakladu Wirusologii PZH [9]	500	2004	39·8%	20·6%	17·0%	547 (343–813)

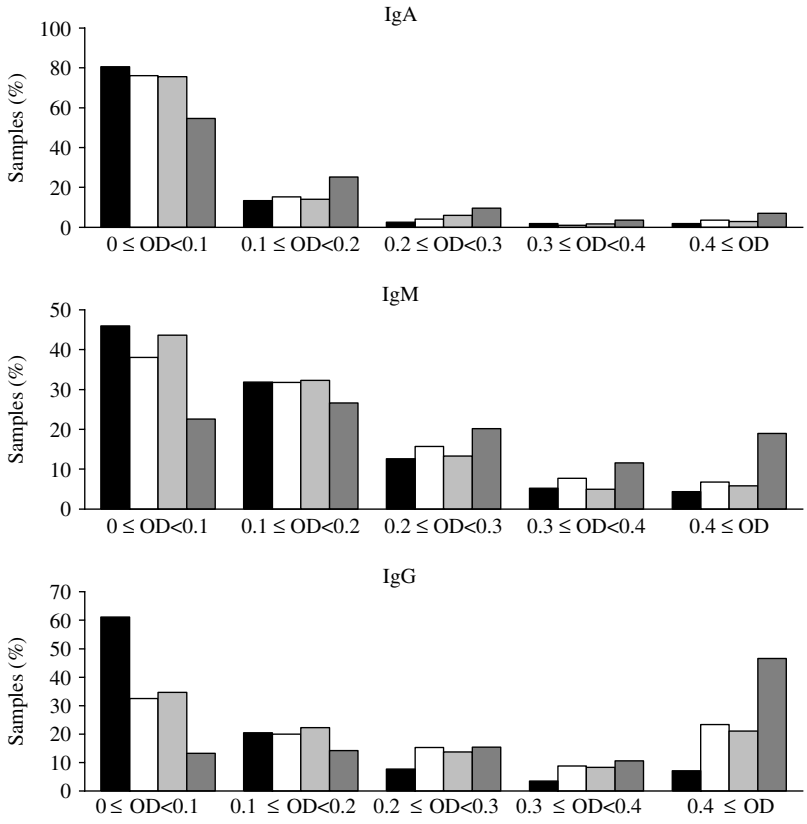

Fig. 1. Distribution of optical density (OD) values for each of the four of serum sample collections. ■, Denmark (2006–2007); □, The Netherlands (2006–2007); ▫, The Netherlands (1998–2002); ▪, Poland (2004).

in Denmark (5% per definition) while the highest prevalence was found in the Polish cohort (39·8%, 20·6%, 17·0% positive samples for IgG, IgM and IgA, respectively); The Netherlands had intermediate values.

Seroincidences

Trace plots and Geweke's tests did not give any indication that convergence of the Markov chains were not attained (data available as Supplementary online material). Further, plots showing goodness-of-fit are shown in Simonsen *et al.* [7]. The posterior distribution of seroincidence is shown in Figure 2 and summarized by median (as a point estimate) and 95% credibility intervals (95% CI) in Table 1. The ordering followed the same pattern as the seroprevalences. The Polish incidence estimate was 547 infections/1000 person-years (95% CI 343–813), corresponding approximately to one seroconversion every second year, while the lowest incidence was in the Danish cohort where it was estimated that on average 84 (95% CI 41–141) seroconversions took place per 1000 person-years, i.e. one seroconversion every 12 years.

https://doi.org/10.1017/S0950268810000750 Published online by Cambridge University Press

Table 2. *Estimated incidence rate ratios with 95% credibility intervals between the four cohorts. These are found by taking medians, 2·5% and 97·5% percentiles from the posterior distributions of the incidence rate ratios*

	Target country			
Reference country	Poland	The Netherlands, Regenboog	The Netherlands, Pienter II	Denmark
Denmark	6·34 (3·31–12·53)	2·01 (0·92–4·20)	1·74 (0·85–3·72)	1 (ref.)
The Netherlands, Pienter II	3·65 (1·84–7·06)	1·16 (0·52–2·38)	1 (ref.)	
The Netherlands, Regenboog	3·15 (1·64–6·57)	1 (ref.)		
Poland	1 (ref.)			

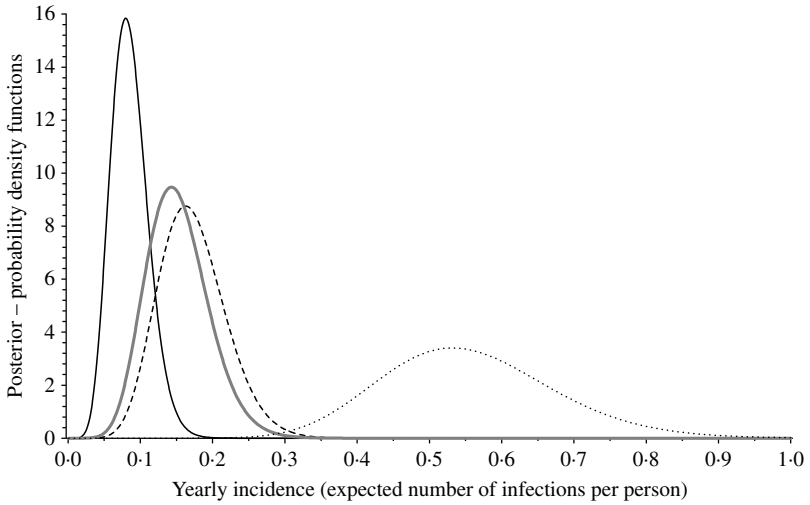

Fig. 2. Posterior density functions of the seroincidence in the participating countries. ——, Denmark; ——, The Netherlands (2006–2007); – – –, The Netherlands (1998–2002); ·····, Poland.

Seroincidence estimates for the two Dutch serum collections were very similar.

Incidence comparisons

Incidence rate ratios with 95% credibility intervals are shown in Table 2. The seroincidence in Poland was significantly higher than in the two Dutch cohorts and the Danish cohort. The largest difference was found between the Polish and the Danish cohort: 6·34 (95% CI 3·31–12·53) infections in Poland per infection in Denmark. Incidence estimates calculated on base of the two Dutch cohorts, which were sampled 6–7 years apart, were both higher than the Danish incidence (ratio 2·01 for Regenboog, 1·74 for Pienter) but this was not statistically significant.

DISCUSSION

This paper shows how seroincidence can be estimated and how this measure can be used to compare S*almonella* infection rates between countries. The use of seroincidence has several benefits compared to the more commonly used seroprevalence. First, the magnitudes of the seroprevalences are dependent on the arbitrary choice of cut-off for being defined as seropositive.

Further, since the persistence of antibodies varies between individuals, any given seroprevalence cannot be converted into a frequency of infection by simple means. Finally, the three types of antibodies (IgG, IgM, IgA) each produce an estimate of seroprevalence, and these three estimates are difficult to combine. In contrast the seroincidence does not suffer from such problems with arbitrary cut-off levels for different antibody types. Further, as seroincidence is based on serum antibody titres measured with the same ELISA in all sera, it does not suffer from the biases present in the numbers of laboratory-reported cases from different nations, and can therefore be used to compare the sensitivity of surveillance systems between countries or periods.

https://doi.org/10.1017/S0950268810000750 Published online by Cambridge University Press

It is important to emphasize that both symptomatic as well as asymptomatic infection can lead to seroconversion; therefore, the seroincidence shown in this paper should not be seen as a measure of frequency of disease, but rather a measure of frequency of *Salmonella* exposures leading to seroconversion. However, as this is likely to reflect the exposure at the population level to this foodborne agent, it might be applicable to evaluate the effect of control programmes between countries.

In general, it seems that the higher the seroincidence, the more the IgG-based seroprevalence exceeds the two others (IgM- and IgA-based seroprevalence). Most extreme was the Polish seroprevalence which varied from 17·0% for IgA to 39·8 for IgG. IgG antibodies decay more slowly than IgM and IgA antibodies and therefore, if incidence is high, IgG antibodies tend to persist while IgA and IgM antibodies may decrease between subsequent infections.

When comparing the estimated seroincidences in the two Dutch cohorts we saw a decline over the period from 169 to 149 cases/1000 person-years. The same pattern was found in the incidence of reported cases which had a decline from 0·18 to 0·14 cases/1000 person-years. However, the study also shows that the annual number of reported cases can lead to a very skewed interpretation of the actual infection pressure. While we observed a ratio of seroincidence rates between Poland and Denmark of 6·34 (Table 2), the incidence of reported cases in the two countries was 0·42 cases and 0·29 cases/1000 person-years, respectively, which gives a rate ratio of 1·44. This underlines that inter-country comparisons based on officially reported surveillance figures and seroincidence ratios have a completely different interpretation. Our methodology – which in the current study has been extended to include formal comparisons of seroincidence between countries – offers an attractive alternative approach that can be used to evaluate surveillance systems and control programmes.

NOTE

Supplementary material accompanies this paper on the Journal's website (http://journals.cambridge.org/hyg).

ACKNOWLEDGEMENTS

The study was partly financially supported by MED-VET-NET, an EU Network of Excellence for research on the prevention and control of zoonoses (EU contract no. FOOD-CT-2004-506122), the Danish Graduate School of Biostatistics and the Faculty of Health Science, Copenhagen University.

The work by P.T. was funded by POLYMOD, EU-FP6 contract no. SSP22-CT-2004-502084. We are very grateful to Dinna Krüger and Tina Hansen, Serological Laboratory, SSI for their careful work and to Gerhard Falkenhorst, Division of Epidemiology, SSI, for his help with administration of the project.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Herikstad H, *et al.*** A population-based estimate of the burden of diarrhoeal illness in the United States: FoodNet, 1996–1997. *Epidemiology and Infection* 2002; **129**: 9–17.
2. **Scallan E.** Activities, achievements, and lessons learned during the first 10 years of the Foodborne Diseases Active Surveillance Network: 1996–2005. *Clinical Infectious Diseases* 2007; **44**: 718–725.
3. **van Pelt W, *et al.*** Laboratory surveillance of bacterial gastroenteric pathogens in the Netherlands, 1991–2001. *Epidemiology and Infection* 2003; **130**: 431–441.
4. **Wheeler J, *et al.*** Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. *British Medical Journal* 1999; **318**: 1046–1050.
5. **MedVetNet Workpackage 32.** 26 September 2006 (http://www.medvetnet.org/cms/templates/doc.php?id=92). Accessed 23 November 2009.
6. **Voetsch A, *et al.*** FoodNet estimate of the burden of illness caused by nontyphoidal Salmonella infections in the United States. *Clinical Infectious Diseases* 2004; **38**: S127–S134.
7. **Simonsen J, *et al.*** Estimation of incidences of infectious diseases based on antibody measurements. *Statistics in Medicine* 2009; **28**: 1882–1895.
8. **Strid MA, *et al.*** Kinetics of the human antibody response against *Salmonella* enterica serovars Enteritidis and Typhimurium determined by lipopolysaccharide enzyme-linked immunosorbent assay. *Clinical and Vaccine Immunology* 2007; **14**: 741–747.
9. **Smith JS, *et al.*** Type specific seroprevalence of HSV-1 and HSV-2 in four geographical regions of Poland. *Sexually Transmitted Infections* 2006; **82**: 159–163.
10. **van der Klis FR, *et al.*** Second national serum bank for population-based seroprevalence studies in the Netherlands. *The Netherlands Journal of Medicine* 2009; **67**: 301–308.
11. **Gilks WR, Richardson S, Spiegelhalter DJ.** *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall/CRC, 1996, pp. 1–15.

12. **Dodds MG, Vicini P.** Assessing convergence of Markov chain Monte Carlo simulations in hierarchical Bayesian models for population pharmacokinetics. *Annals of Biomedical Engineering* 2004; **32**: 1300–1313
13. **Lunn DJ.** WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**: 325–337.
14. **SAS.** SAS version 9.1 for Windows. SAS Institute, Cary, NC, USA, 2002–2003.

https://doi.org/10.1017/S0950268810000750 Published online by Cambridge University Press