LA-UR- 08-6913

Approved for public release;
distribution is unlimited.

Title:	Distributions for Case Mortality Rate Based on Historic Pandemic Influenza Death Rates
Author(s):	Emily J. Powell Michael S. Hamada Leslie M. Moore Dennis R. Powell
Intended for:	~~Distribution~~ Proceedings of Joint Statistical Meetings, Washington DC, USA

Los Alamos
NATIONAL LABORATORY
EST. 1943

Form 836 (7/06)

Distributions for Case Mortality Rate Based on Historic Pandemic Influenza Death Rates

Emily J. Powell[1], Michael S. Hamada[2], Leslie M. Moore[2], Dennis R. Powell[3]

B-7[1], CCS-6[2], D-6[3]

Los Alamos National Laboratory

August 2008

Abstract:

Los Alamos National Laboratory has developed a simulation model to understand infrastructure impacts of a pandemic influenza (PI) outbreak. Inputs to the model include parameters such as PI case mortality rate and disease spread rate. To investigate PI consequences and impacts with this model a distribution of values for these PI parameters is needed. Our objective is to explore ways to incorporate limited historical information to characterize possible distributions for these inputs, primarily using Bayesian statistical methods and analysis tools. We model historical PI data using a nested binomial model with beta priors that incorporate expert opinion and are consistent with a simplified epidemiological model. A Bayesian inferential approach is used to obtain posterior distributions on case mortality rate and attack rate and we study the effect of choices of prior distributions on the results.

Key Words: attack rate, Bayes' Theorem, case mortality rate, nested binomial model, pandemic influenza, probability distribution

Introduction:

Pandemic influenzas (PI), or pandemics of any kind, pose a greater risk as the infrastructures of society become more complex and intertwined. Additionally, as travel between countries becomes easier and faster, pandemics have the potential to rapidly spread, limiting response options to combat the disease. Computer simulations that model disease progression and consequences are useful for analyzing possible impacts of an outbreak. In such simulations, statistics plays a potentially significant role in understanding useful and applicable data for inputs, as well as in summarizing outputs from the simulations. A computer program which models the propagation of PI and consequences in terms of multiple metrics, such as economic costs, deaths, and illness requires as inputs likely values for case mortality (the probability of death) and transmission rate (the probability of infection). The goal of this paper is to investigate methods for using available historical data and developing reasonable distributions for these input parameters, as required for further study of PI impact via simulation models.

There have been three flu pandemics in the past century. They occurred in 1918-1920, 1957-1960, and 1968-1972. For these pandemics, reliable data on the number of symptomatic and the number of deaths is difficult to obtain. The best data, kept by the military, suggests different ethnicities had varying responses to the viruses, mainly evidenced by differing death rates. Case mortality rate is a parameter that describes the probability that a person infected with the disease will die either directly from the disease or related complications. An equally important and related parameter is attack rate, a parameter that characterizes the ratio of the number infected to the initially susceptible population. An initial population is defined as the population prior to the outbreak of the disease and is composed of those who are immune to the disease, and everyone else, the susceptible population. A disease model based on these two parameters provides a way to understand pandemic influenza and a basis for preparation for possible future pandemics. A simple model, illustrated in Figure 1, assumes that the initial population is composed of two groups: 'susceptible' and 'immune'. Those in the 'susceptible' group can get the disease and are then moved into the 'infected' category. Otherwise they remain 'not infected'. Persons in the 'infected' group either both get well and move into the 'recovered' category or they die from the disease and are assigned to the 'dead' category. For PI, the 'recovered' become part of the 'immune' subset of the population. With this simplified model, focus is on the two main parameters: attack rate and case mortality rate.

The Problem:

Using data from the three historical pandemics, the goal is to obtain reasonable distributions for both case mortality rate and attack rate. Ideally for estimation of PI consequences, these distributions will capture the full range of possible case mortality rates and attack rates which when propagated thru a simulation model would provide possible low to high estimates of infection and death. Using these distributions in a simulation of PI can provide assessment of consequences while capturing uncertainties in the outputs. Our goal is to incorporate historical PI data and expert opinion to create reasonable distributions relevant to possible future pandemics. Using solely historical data has shortcomings associated with poor data collection, limited quantity, and uncertainty about the relevance of past PI occurrence to possible future pandemics.

Using a statistical model to obtain distributions on these parameters has the potential to mitigate the effect of little, and questionably reliable data, and focus on features of realistic distributions, possibly incorporating expert opinion. Some association of future influenza pandemics with influenza pandemics of the past is expected. Therefore, a relatively accurate characterization of a distribution for past case mortality data is relevant. We assume influenza pandemics can be modeled as a stochastic process where each pandemic is an instantiation of a joint distribution of factors that characterize the disease. One of the factors is the case mortality rate, which is of most direct interest. The attack rate is of secondary interest; it determines the 'infected' population which is required to estimate deaths. Figure 1 illustrates how both attack rate and case mortality rate act on the total population and the infected population, and how they interact with each other.

Both attack rate and case mortality rate, as uncertain parameters in the simulation, require statistical distributions. Beta distributions seemed a reasonable form to assume for the distributions of these parameters because they naturally represent distributions of proportions. Below, the beta distributions assumed as initial candidates for distributions, or prior distributions, on these parameters were created using maximum and minimum values from expert opinion.

A pandemic is often characterized by three quantities:

a) attack rate = θ = number people sick / total population,

b) gross death rate = K×(deaths / total population), where K is a constant (like 1,000 or 100,000) so that gross death rate is per K unit population, and

c) case mortality rate = δ = deaths / (attack rate × total population).

A further simplification to the model sketched in Figure 1 is represented in Figure 2. The assumption is that everyone in the initial population has the same probability of becoming 'infected' and all 'infected' have the same probability of dying. Although the flow chart of Figure 2 may look as though the entire population gets infected and then dies, this is not the case; it merely illustrates the most important aspects of the model. In particular, it emphasizes the assumption of binomial distributions for the number of 'infected' and 'dead', with respective parameters θ and δ.

Several assumptions are implicit in this simplification of the problem. These include that the population is known and equals the number of susceptible people; there are no immune people. Additionally, each person in the initial and infected populations has the same probability of being infected or dying, and infection or death of one person does not influence infection or death of another, so the events of an individual becoming infected or dying are independent. Bayesian analysis requires assumption of prior distributions on the case mortality rate and attack rate parameters. In the following, beta distributions will be assumed as priors on attack rate and case mortality rate, the respective parameters of the binomial distributions assumed for number infected and dead. Many of these assumptions are imperfect, but based on discussions with subject matter experts, they are considered adequate for the current analyses.

Also, there are issues related to limited available historical data. Reliable data on numbers of infected and dead is limited simply because it may not have been recorded. Many people do not receive treatment for flu-like symptoms, and deaths may occur prior to treatment or even record of infection. It is widely considered that for the three historical PI the best data were kept by the military. In this data, different ethnic groups were anecdotally observed to have varying responses to the virus. Specifically Native Americans in the U.S. and Indian militia in the British army in India were observed to suffer high losses. However, data on infections and deaths may have been mitigated by planned or ad hoc preventative measures, and uneven application of these measures in different ethnic groups may have resulted in the differential death rates among British troops versus India population. The actual data available is how many died, which is relevant to case mortality rate if number of infected is known, but how many were infected (relevant to attack rate) is not available. Glezen (1996) provided population data and the number of people who died from the virus. Although the data on the numbers of

illnesses and related deaths from the three historical pandemics is incomplete, subject matter experts familiar with the literature in this area provided information about accepted ranges of case mortality and attack rates for PI. These ranges included high and low estimates for attack rate and case mortality rate which would possibly lead to a pandemic. From Brundage (2006), the 20th century influenza pandemic attack rates ranged from 24.7% to 34.2% and overall case mortality rates ranged from 4.4 to 6.7 per 1,000.

Basic Tools and Procedure:

The model for this problem is a nested binomial model with priors for θ and δ. Bayes' Theorem is the method for obtaining posteriors for θ and δ. Bayes' Theorem demonstrates how 'probabilities change in light of the data', from Berry (1996). Bayesian analysis requires assumptions on data models and priors. Priors are distributions on the parameters which are not based on data but on prior opinion. Posteriors are distributions obtained by Bayes' Theorem from the data and the priors. Different priors may be tried to see how they affect the resulting posterior when data is incorporated.

The data model for this problem is: Dead ~ Binomial(Infected, δ). A binomial distribution seems a reasonable assumption since there are only two possible outcomes: 'recovered' or 'dead' for each individual. This binomial distribution depends on the number infected ('Infected') as well as case mortality rate, δ. 'Infected' and case mortality rate require distribution assumptions as well. 'Infected' is actually unobserved data and is assumed to be distributed as follows: Infected ~ Binomial(Population, θ). 'Infected' is also reasonably assumed modeled binomially because there are only two options for each individual: 'infected' or 'not infected'. This binomial distribution depends on the total population size ('Population') and attack rate, θ. Population sizes for each historic pandemic were given in Glezen (1996) and are listed in Table 1 along with deaths for the three historical PI. The attack rate requires assumption of another prior. Attack rate, $\theta \sim$ Beta(•, •), and case mortality rate, $\delta \sim$ Beta(•, •), are both assumed to have Beta distribution forms for their priors. These seem reasonable forms for these prior distributions because they are defined for values between 0 and 1.

Priors play a pivotal role in Bayesian statistical analysis because they capture how much expert opinion weighs in relative to how much the posterior will be determined by the data available. Here several different cases were tried including different combinations of priors. Both attack rate and case mortality rate were assigned two distributions. The subject matter experts initially proposed the prior for attack rate be θ~Uniform(0.247, 0.342) based on the interval formed by the lowest and highest estimated attack rates from historical data in Brundage (2006). Prior information on case mortality rate suggests it lies in the interval (0.005, 0.15), which is considered a conservative estimate of the range of PI case mortality rate. Seasonal flu has a case mortality rate of about 4×10^{-5} unless the patient is over 65, in which case it is about 0.01, for an average case mortality rate of 1.6×10^{-3}. However, seasonal flu is much milder than pandemic influenza, and is a lower bound on the case mortality rate. At the other extreme, for small isolated populations, the case mortality rate can be as high as 1.0.

However, when assessed as part of larger regions, the rate rarely exceeds 0.20. Recently, the discovery of H5N1 in humans with an associated case mortality rate of 0.55 (although some estimates are as low as 0.33) raise concern that a pandemic could have an upper limit near 0.5. Many experts fear that H5N1 will be the viral basis for the next pandemic. Alternatively, uninformed attack rate and case mortality rate priors might have assumed Uniform(0,1) priors, or equivalently $\theta \sim$Beta(1,1) and $\delta \sim$Beta(1,1). The WinBUGS software tool was used for analyses (Spiegelhalter et al., 2004). Since the initial uniform distributions proposed by subject matter experts did not work in the WinBUGS program, these uniform distributions were modified to beta distributions with parameter values chosen so that the proposed uniform intervals had high probability coverage with the beta distributions and the means of the beta distributions matched the means of the proposed uniform distributions.

Table 1. Historical PI data from Glezen (1996)

Pandemic Years	Number Died	Population Size
1918-1920	675,000	103,262,929
1957-1960	115,700	173,723,700
1968-1972	111,927	203,211,926

More precisely, expert opinion derived prior beta distributions parameters were set so that they have high probability of lying in the intervals (0.247, 0.342), for attack rate, and (0.005, 0.15), for case mortality rate, and so that the mean of the Beta distributions was .2945 for attack rate and .0775 for case mortality rate. Here is the reasoning that went into choosing prior beta distributions. The illustration is for attack rate, θ, but the same method was applied to case mortality rate, δ. For a distribution Beta(a,b), a and b are wanted such that the mean is the same as the mean of the uniform distribution Uniform(0.247,0.342), or 0.2945, and the probability of being in interval (0.247, 0.342) is a high value accepted by the subject matter experts. The approach takes $a=n\times0.2945$ and $b=n\times0.7055$ so that the mean of the beta distribution is fixed at 0.2945. Then n is selected so that the subject matter experts' probability condition is met as nearly as possible. So

$$\theta \sim Beta(a,b) \quad = \quad Beta(n\times0.2945, n\times0.7055)$$

$$E(\theta)=\frac{a}{a+b} \quad = \quad \frac{n\times0.2945}{n\times0.2945+n\times0.7055}=$$

$$\frac{n*0.2945}{n}=0.2945$$

The prior distribution assumed for attack rate is mapped to Beta(70.091,167.909) which assigns nearly 90% probability to the interval (0.247,0.342). The expert opinion-derived prior distribution for case mortality rate is mapped to Beta(1.0075,11.9925) which has a mean of 0.0775 and assigns almost 80% probability to the interval (0.005,0.15).

Three different combinations of prior cases are investigated. Case one uses the expert-informed attack rate prior and a uniform case mortality rate prior. This allowed us to see the effect of data on the case mortality rate prior. Case two had uniform priors on both the attack and case mortality rates; thus assuming completely uninformed knowledge of the possible values of these parameters. Case three had expert-informed attack and case mortality rate priors.

With these model assumptions for the data and prior distributions, Bayes' Theorem allows us to update prior distributions with the available historical PI data to get posterior distributions on these inputs to be used with the simulation model. Bayes' Theorem derives a posterior distribution from the multiplication of the likelihood and priors. The likelihood is from our data model and is the likelihood of the number of 'dead' in the three historical PI. The 'infected' data is not directly observed but also occurs three times, once for every pandemic. The attack rate and case mortality rate are both beta distributions. Bayes' Theorem allows us to multiply these likelihood and probability density functions together to obtain a posterior distribution:

$$posterior \propto likehood(data) \times prior .$$

From the data $Dead_i \sim$ Binomial($Infected_i$, δ), we have the likelihood

$$\prod_i \delta^{Dead_i} (1-\delta)^{Infected_i - Dead_i} .$$

The prior based on $Infected_i \sim$ Binomial($Population_i$, θ), $\theta \sim Beta(a_\theta, b_\theta)$, and $\delta \sim Beta(a_\delta, b_\delta)$ is

$$\prod_i \theta^{Population_i} (1-\theta)^{Population_i - Infected_i} \times \theta^{a_\theta - 1} (1-\theta)^{b_\theta - 1} \times \theta^{a_\delta - 1} (1-\theta)^{b_\delta - 1} .$$

WinBUGS was used to perform the Bayesian statistical analysis. WinBUGS implements Bayes' Theorem and makes draws from the posteriors using a Markov chain Monte Carlo (MCMC) algorithm. These draws can then be used to characterize the posterior. R was used to create the graphs in the following results section (R Core Development Team, 2004).

Results:

Case 1 used uniform case mortality rate prior $\delta \sim$ Beta(1,1) and expert-informed attack rate prior $\theta \sim$ beta(70.091,167.909). We use the kernel density function 'density()' in R to smooth the histogram of the draws from the posterior distributions generated by WinBUGS, and plot a graphical representation of the distributions. In Figure 3, the case mortality rate posterior distribution is the solid line and the uniform case mortality rate prior is the dashed line. The data significantly alters the prior distribution with the

posterior distribution exhibiting a pronounced peak. In Figure 4 the attack rate posterior and prior distributions are again respectively the solid and dashed line. The data had a more modest effect on the attack rate posterior as changed from the expert-informed prior distribution. Table 2 gives several quantile values for the case mortality and attack rate posteriors. Quantiles are the x-axis values for θ, attack rate, or δ, case mortality rate at positions along the graph corresponding to cumulative probability values (such as 0.50 quantile, which is also called the median).

Table 2. Quantiles for Case 1 Posteriors of δ and θ

Probability	0.025	0.05	Median, 0.50	0.95	0.975
δ	0.01697	0.01762	0.0212	0.02604	0.02716
θ	0.2324	0.2426	0.2916	0.3416	0.3519

Case 2 assumes the uninformed priors so both δ and θ are Uniform(0,1) or equivalently Beta(1,1). Figure 5 shows δ posterior with a solid line and δ uniform prior as a dashed line. Again, the data has a large influence on the uninformed prior. In Figure 6 the data also exhibit a pronounced effect on the uninformed θ prior. The posterior distribution is irregular and perhaps could be better smoothed by the kernel density function. Table 3 lists quantiles values for δ and θ posteriors.

Table 3. Quantiles for Case 2 Posteriors of δ and θ

Probability	0.025	0.05	Median, 0.50	0.95	0.975
δ	0.01096	0.01177	0.05078	0.6452	0.7867
θ	0.00779	0.009542	0.1224	0.5208	0.5542

Case 3 has expert-informed priors for both attack rate and case mortality rate, θ~Beta(70.091,167.909) and δ~Beta(1.0075,11.0025). Figure 7 shows δ posterior as a solid line with the informed δ prior dashed. The data has a large influence on the case mortality rate prior. Figure 8 shows θ posterior with informed θ prior, solid and dashed respectively. The data has little influence on the expert-informed attack rate prior with the posterior barely different from the prior. We observe this is similar to case 1 where the expert-informed attack rate prior exhibits little effect from the available historical PI data. Table 4 lists quantiles values for δ and θ posteriors.

Table 4. Quantiles for Case 3 Posteriors of δ and θ

Probability	0.025	0.05	Median, 0.50	0.95	0.975
δ	0.01696	0.01762	0.02115	0.02591	0.027
θ	0.2332	0.2432	0.2923	0.342	0.3523

Table 5 gives a summary of the results across cases. Observe that the attack rate posteriors for the cases with expert-informed priors (highlighted in blue) are little

changed from the priors. For both expert-informed and Uniform(0,1) case mortality rate priors with expert-informed attach rate priors, the posteriors on case mortality rate were almost the same and substantially differ from the priors with the inclusion of the PI historical data.

For case 2, with both attack rate and case mortality rate priors completely uninformed, the attack rate posterior is rather low, much lower than suggested by expert opinion, although at the same time the 0.95 probability interval is much wider than that suggested by expert opinion. In this case, the effect on case mortality rate seems to be that, so to have the number of deaths observed in the PI historical data, the case mortality rate has to be higher as evidenced by the median of the posterior of case mortality rate. It is interesting to observe that in this case, assuming little knowledge of case mortality, the median of the posterior of case mortality is close to the median of the expert-informed prior, although again the 0.95 probability interval for the posterior of case mortality is much higher and broader than what might be expected by subject matter experts.

Table 5. Medians and 0.95 probability intervals for the priors and posteriors calculated from 3 cases of prior assumptions for PI attack rate, θ and case mortality rate, δ.

		θ, PI attack rate		δ, PI case mortality rate	
		Median	0.95 probability interval	Median	.95 probability interval
Priors	Beta(1,1)	0.5	(0.025,0.975)	0.5	(0.025,0.975)
	Beta(70.091,167.909)	0.2939	(0.2384, 0.3539)		
	Beta(1.0075,11.9925)			0.0567	(0.0022,0.2657)
Posteriors	Case 1 priors θ~Beta(70.091,167.909) δ~Beta(1,1)	0.2916	(0.2324, 0.3519)	0.0212	(0.0170, 0.0272)
	Case 2 priors θ~Beta(1,1) δ~Beta(1,1)	0.1224	(0.0078, 0.5542)	0.05078	(0.0110, 0.7867)
	Case 3 priors θ~Beta(70.091,167.909) δ~Beta(1.0075,11.9925)	0.2923	(0.2332, 0.3523)	0.0212	(0.0170, 0.0270)

Future Work

There are several steps that could be taken to improve this work. First is the replacement of the very simple disease model with a more appropriate model framework, such as an SIR or SEIR (Susceptible, Exposed, Infected, Recovered) model. In this

approach we modeled the attack rate, a value that was not used by the customer, but needed for the simple disease model used in the analysis. An SEIR model requires the input of the reproductive number (often called R_0). This parameter describes the average number of people who will be infected by a person who is already infected with the disease. The customer actually desired distributions for both case mortality rate and R_0, Thus using the SEIR model would have been more appropriate from the beginning, but time constraints did not allow this approach initially.

The customer found the distributions on case mortality rate very valuable, particularly the distributions based on uninformed priors. The customers felt that the case 2 posterior distribution best captured the likelihood of a deadly H5N1 based pandemic and reflected an upper limit on the regime of plausible distributions. Ultimately the customers chose to use a Beta(1.2,58.8) to model the case mortality rate. They assessed this distribution to be more conservative than the case 2 result, but more plausible than cases 1 and 3.

Acknowledgements:

This work was performed as a summer undergraduate student project at Los Alamos National Laboratory by the lead author. Dr. Michael S. Hamada of the Statistical Sciences Group, CCS-6, was the student's mentor and advisor. Mr. Dennis Powell of the Risk Analysis and Decision Support Systems Group, D-6, and Dr. Leslie M. Moore, CCS-6, posed the problem and were collaborators. Dr. Jeanne Fair of the Biosecurity & Public Health Group, B-7, also provided subject matter expertise and review. Other advisors from CCS-6 on statistical tools and methods included Dr. Christine Anderson-Cook and Dr. Kary Myers. The authors gratefully acknowledge these individuals and C. C. Essix for their support and encouragement.

References:

Berry, Donald A. (1996), *Statistics: A Bayesian Perspective*, Wadsworth Publishing Company: Belmont.

Brundage, J. F. (2006), "Cases and Deaths During Influenza Pandemics in the United States", *American Journal of Preventative Medicine*, 31(3):252–256.

Glezen,W.P. (1996), 'Emerging Infections: Pandemic Influenza', *Epidemiologic Reviews*, 18:64-76.

Population Census, http://www.census.gov/popest/archives/1990s/popclockest.txt.

R Development Core Team (2004),. R: A Language and Environment for Statistical

Computing, R Foundation for Statistical Computing: Vienna. (http://www.R-project.org).

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2004), *WinBUGS Version 1.4 User Manual.*

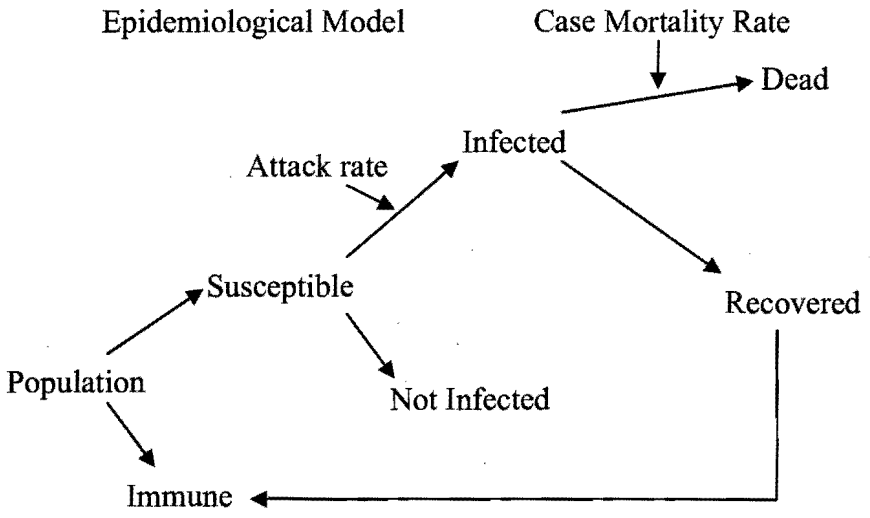

Figure 1. Simple flow chart of typical disease progression in pandemic influenza simulation with emphasis on roles of attack rate and case mortality rate.

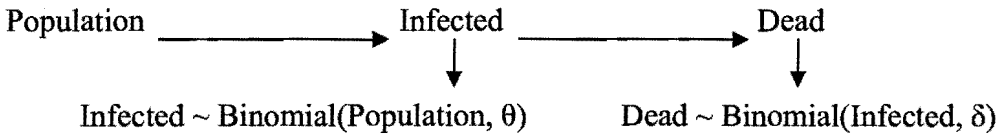

Figure 2. Simplified influenza progression model. Infected and Dead are assumed to have binomial distributions.

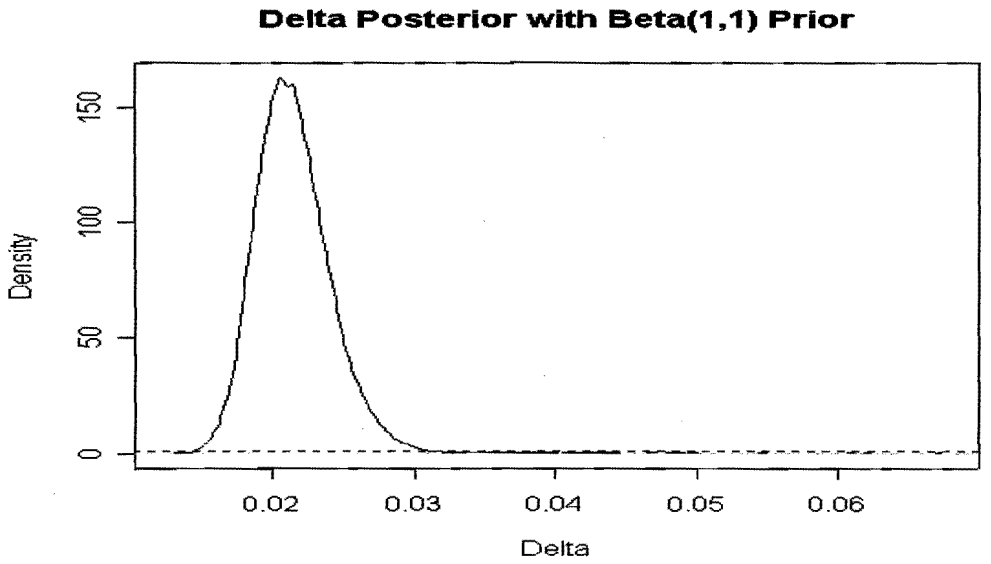

Figure 3. Case 1 posterior distribution of case mortality rate, δ, (solid line) plotted with case mortality rate prior (dashed line, uniform on (0,1)). ('Delta'= δ)

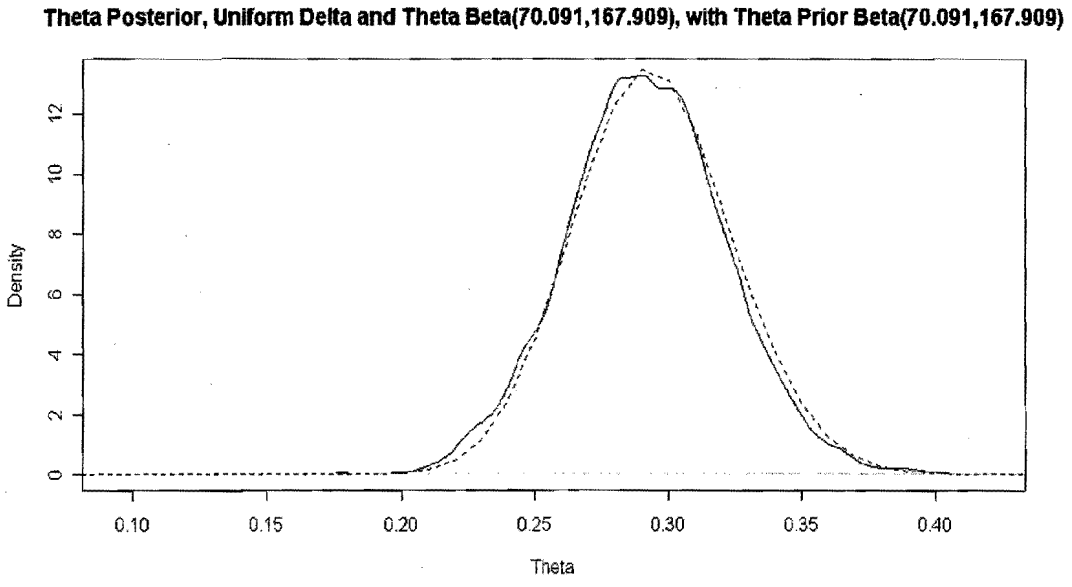

Figure 4. Case 1 posterior distribution of attack rate, θ, (solid line) plotted with expert-informed attack rate prior (dashed line). ('Theta'= θ, 'Delta'= δ)

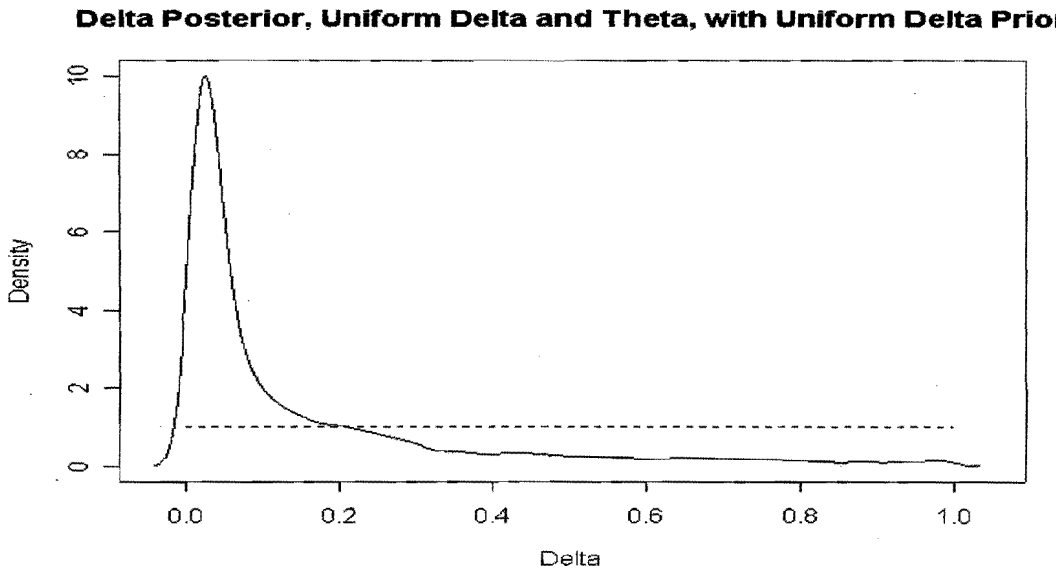

Figure 5. Case 2 posterior distribution of case mortality rate, δ, (solid line) plotted with case mortality rate prior (dashed line, uniform on (0,1)). ('Delta'= δ)

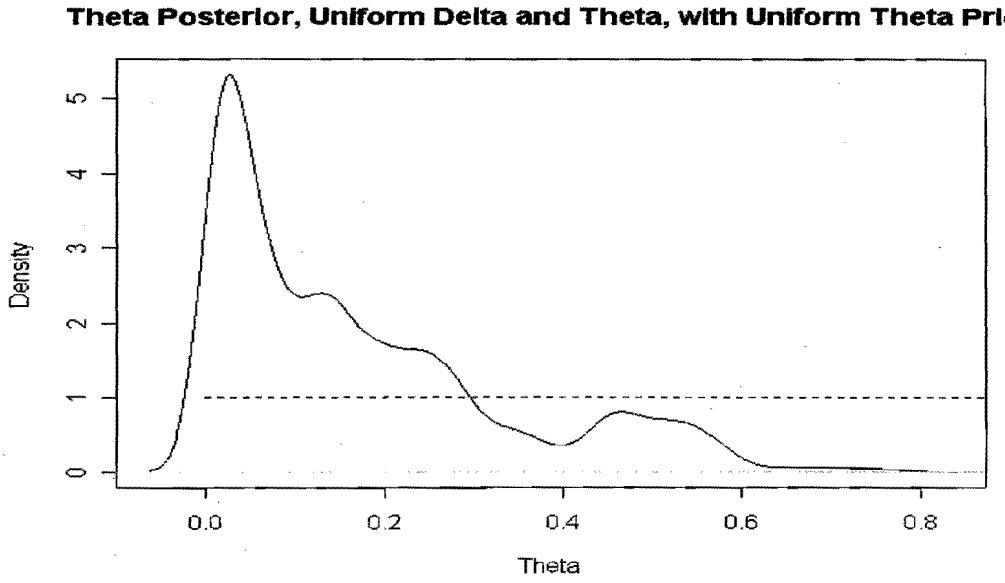

Figure 6. Case 2 posterior distribution of attack rate, θ, (solid line) plotted with expert-informed attack rate prior (dashed line). ('Theta'= θ, 'Delta'= δ)

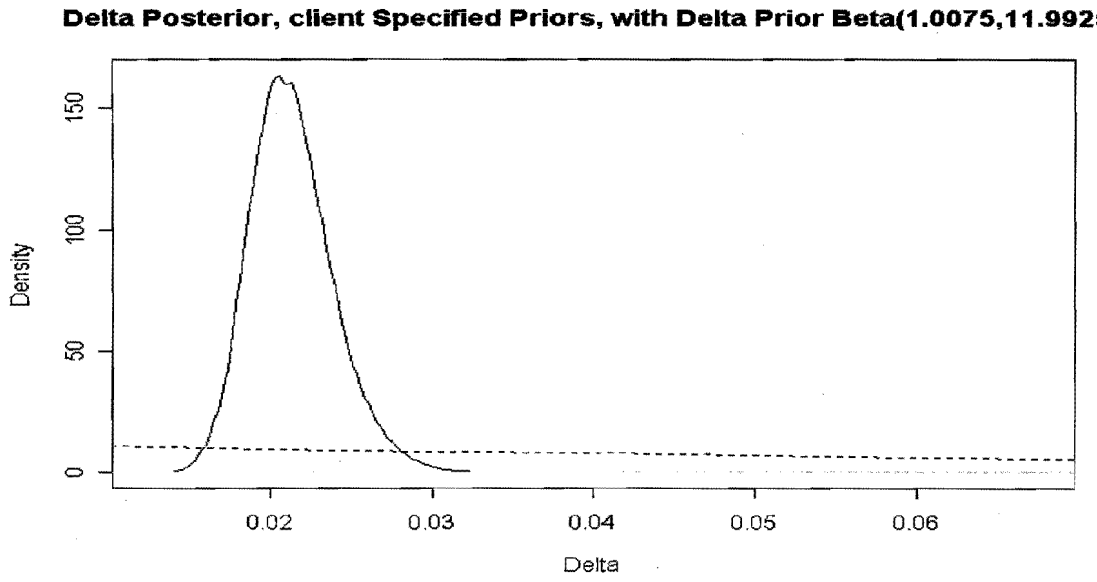

Figure 7. Case 3 posterior distribution of case mortality rate, δ, (solid line) plotted with case mortality rate prior (dashed line, uniform on (0,1)). ('Delta'= δ)

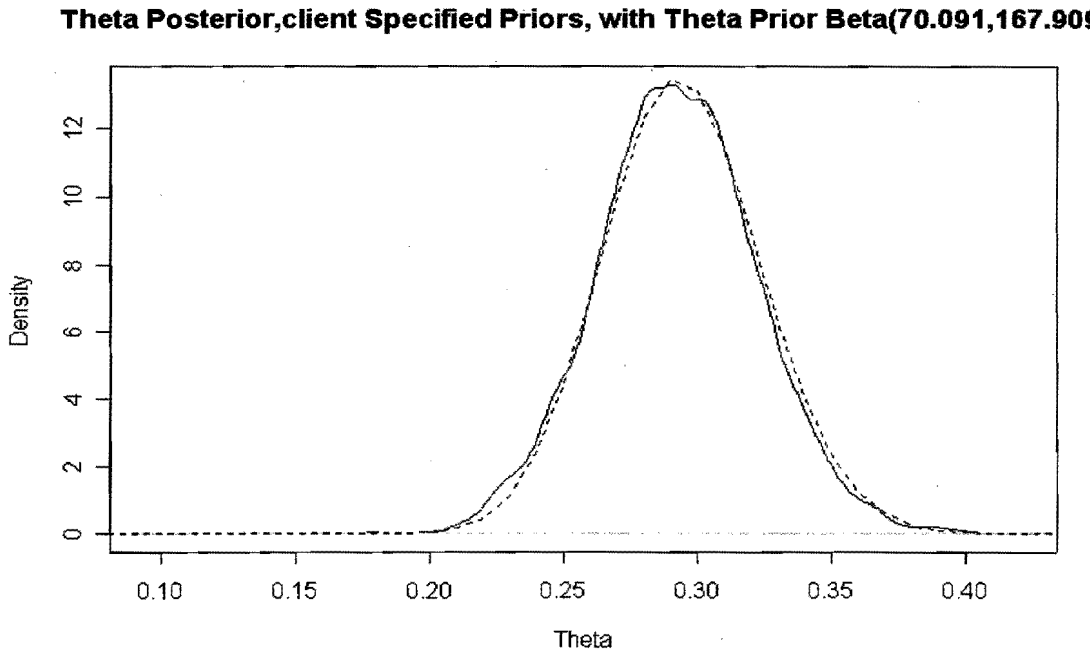

Figure 8. Case 3 posterior distribution of attack rate, θ, (solid line) plotted with expert-informed attack rate prior (dashed line). ('Theta'= θ)