

Supplementary Methods – Bayesian Estimate of Yearly Pandemic Probability

The probability of a pandemic occurring undoubtedly varies seasonally and geographically within each year. Although some data is available on the relationship between seasonal variation in humans and avian influenza (Li et al. 2004) (Park, unpubl. data), we focus attention on the yearly probability of a pandemic globally. This does not imply that seasonal and geographic variation is ignored, but rather that this variation is lumped into a composite, yearly average for the globe as a whole. Given the limited available data on the occurrence of pandemics, more detailed analyses seem unwarranted. Similarly, although the probability of a pandemic is likely to vary from year to year, and perhaps also to change systematically over time (Lipatov et al. 2004; Webby & Webster 2003), too little data is available to conduct an analysis of these effects. Consequently we assume that the probability of a pandemic has remained constant over time, and as such the results should again be interpreted as averages.

We follow Bayesian techniques (Bayes 1763) and begin with a prior probability density describing the likelihood of the parameter of interest (here the yearly pandemic probability) taking on various values. The prior distribution is meant to reflect the belief, based on anecdotal evidence, in different values of the pandemic probability. We then use Bayes' rule to combine this with data on inter-pandemic intervals to calculate the posterior distribution (Berger 1985).

The inter-pandemic interval is a geometric random variable with parameter λ representing the yearly probability of a pandemic. It is believed that the pandemic occurring prior to that of 1918 occurred around 1890, giving an observed interval of 28 years (Cox & Subbarao 2000; Reid & Taubenberger 2003). The observed interval between the 1918 and 1957 pandemic is 39 years, and the observed interval between the 1968 and 1957 pandemic is 11 years (Cox & Subbarao 2000). (We exclude the 1977 pandemic because it is thought to have possibly resulted from an accidental lab release of H1N1; this exclusion should make the results conservative). If we assume that the occurrence of a pandemic in any given year is independent of other years, then the

probability of obtaining these three intervals for the set of 4 pandemics is

$(1 - \lambda)^{28-1} \lambda(1 - \lambda)^{39-1} \lambda(1 - \lambda)^{11-1} \lambda$ or $P(data | \lambda) = (1 - \lambda)^{75} \lambda^3$. We seek the posterior distribution $P(\lambda | data)$, using Bayes' Theorem. This is given by (Berger 1985)

$$P(\lambda | data) = \frac{F_{prior}(\lambda)P(data | \lambda)}{\int_0^1 F_{prior}(\lambda)P(data | \lambda)d\lambda},$$

where $F_{prior}(\lambda)$ is the prior distribution based on the anecdotal evidence. For simplicity we use a truncated Exponential distribution for the prior, with parameter α . Figure 1 illustrates the prior and posterior distributions, with parameter value $\alpha = 25$, which gives an expected pandemic probability of 0.04 in the prior distribution as the anecdotal evidence suggests. The expected value of the posterior distribution is 0.039, and a 95% support interval for this posterior distribution is found to be 0.007-0.076 (technically this is the 95% highest posterior density support interval; p.140 (Berger 1985)).

Bayesian analyses must necessarily involve a subjective choice for the prior distribution, and therefore, to determine the robustness of our results we conducted the same analysis using two other extreme choices for the prior distribution: a uniform prior (which corresponds to a complete lack of prior information) and a truncated Normal prior with a mode at 0.04 (which, with a small variance, corresponds to having high confidence in the pandemic probability being 0.04). Both of these alternative choices yield posterior distributions that are skewed to higher values, suggesting that our choice of an Exponential distribution results in a conservative estimate compared with other choices (Figure S1.1).

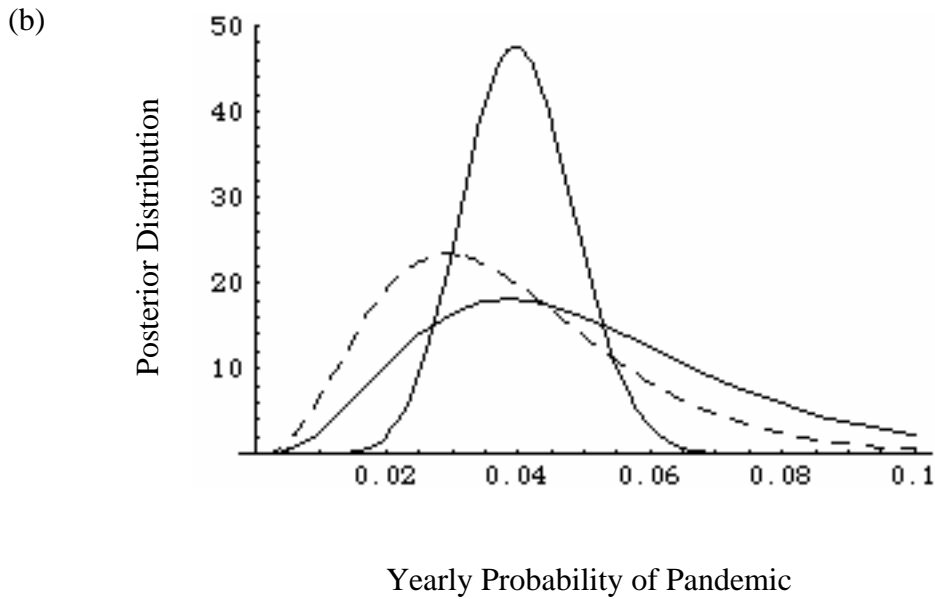
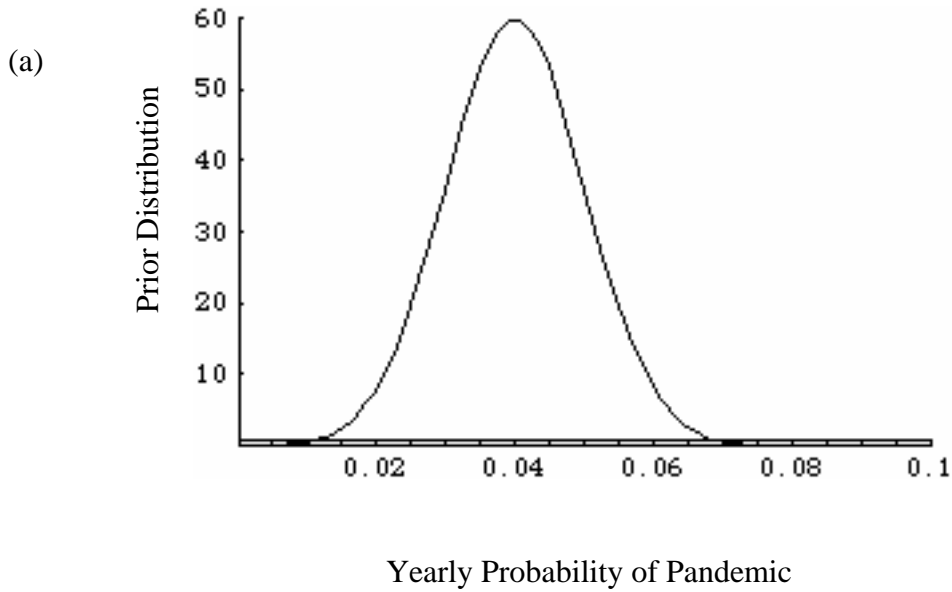


Figure S1.1 – (a) Uniform and truncated Normal prior distributions. (b) Solid curves are the posterior distributions corresponding to the uniform and truncated Normal priors (solid curve reaching the highest peak is that corresponding to the truncated Normal prior). Dashed curve is the posterior distribution of the main text for comparison.