Hiroshi Nishiura
The University of Hong Kong and Japan Science and Technology Agency
e-mail: nishiura@hku.hk
Gerardo Chowell
Arizona State University
e-mail: gchowell@asu.edu
Carlos Castillo-Chavez
Arizona State University
e-mail: ccchavez@asu.edu

Validating early estimation of the transmission potential of pandemic influenza (H1N1-2009): Sample size estimation for post-epidemic seroepidemiological studies

Seroepidemiological studies before and after the epidemic wave of influenza (H1N1-2009) are useful for estimating final size with a potential to validate early estimates of the reproduction number, R, in modeling studies. Nevertheless, a glance at the literature shows that various seroepidemiological studies published so far have adopted a binomial sampling process to quantify the uncertainty of the *proportion* of infected individuals. In the present study, the use of an asymptotic distribution of the final epidemic size that allows for the computation of approximate 95% confidence intervals of the proportion of individuals in a population infected during an epidemic, is proposed since infection events are not independent. Let $\hat{\rho}$ be an observed final size, v be the coefficient of variation of the generation time distribution, and q be the proportion of initially immune individuals. Assuming that v and q are known, we propose the Wald approximation by which the $100(1-2\alpha)\%$ confidence interval for ρ is calculated as

$$\hat{\rho} \pm z_\alpha \sqrt{\frac{\hat{\rho}^3(1-\hat{\rho}) + v^2\hat{\rho}(1-\hat{\rho})^2 \ln^2(1-\hat{\rho}/(1-q))}{n\left[\hat{\rho} + (1-\hat{\rho})\ln(1-\hat{\rho}/(1-q))\right]}} \quad (1)$$

where n is the sample size and z_α denotes $1-\alpha$ quantile of the standard normal distribution. This approach allows the comparison of observed final sizes against model studies based predictions ($R = 1.15, 1.40$ and 1.90) while yielding simple formulae for determining acceptable sample sizes for future seroepidemiological studies. Eleven published seroepidemiological studies of H1N1-2009, which took place after observing the peak incidence in a number of countries, are used in the testing of the methodology. Observed seropositive proportions in six studies appear to be significantly smaller than those predicted from $R = 1.40$; four of the six studies sampled serum less than one month after the reported peak incidence. Comparisons of observed final sizes against $R = 1.15$ provide evidence that all eleven studies do not significantly deviate from the prediction with $R = 1.15$ while comparisons with $R = 1.90$ suggest that the final sizes in nine studies would be overestimated. Sample sizes of published seroepidemiological studies were too small to assess the validity of model predictions except when $R = 1.90$ was used. We recommend the use of the proposed approach in determining the sample size of post-epidemic seroepidemiological studies, calculating the 95% confidence interval of observed final size, and conducting relevant hypothesis testing instead of the use of methods that rely on a binomial proportion,