Response to Invited Commentary

Van Ballegooijen et al. Respond to “Evaluating Vaccination Programs Using Genetic Sequence Data”

W. Marijn van Ballegooijen*, Robin van Houdt, Sylvia M. Bruisten, Hein J. Boot, Roel A. Coutinho, and Jacco Wallinga

* Correspondence to Dr. W. Marijn van Ballegooijen, RIVM, National Institute for Public Health and the Environment, PO Box 1, 3720 BA Bilthoven, the Netherlands (e-mail: marijn.van.ballegooijen@rivm.nl).

Initially submitted September 28, 2009; accepted for publication October 8, 2009.

We thank professors Halloran and Holmes (1) for their thoughtful comments on our article (2). The idealized data set they describe will prove very useful for future studies that use genomic data to assess the effectiveness of vaccination programs. The advent of genomic data collection in epidemiologic studies stresses the need for adequate methods to analyze these kinds of data in order to estimate the impact of preventive measures such as vaccination. The coalescent analysis is a promising method for this purpose because it relates genetic information to epidemic dynamics. However, the coalescent approach is also a complex method. It is not always easy to grasp its mathematical underpinning, it is difficult to oversee all simplifying assumptions, and, when applied together with a Bayesian Markov chain Monte Carlo algorithm, it presents a number of computational and statistical challenges. We believe that at this stage, the field will benefit from a critical discussion of this method and would very much like to see studies that help to determine its strengths and pitfalls.

Finding the correct interpretation for changes in genetic diversity is the key challenge for the coalescent approach. In simplified, selectively neutral populations, genetic diversity is proportional to the incidence of infection. In the real world, however, we encounter selective sweeps in the pathogen population, changing risk behavior of the host population, superspreading events, and biased sampling of pathogens from the population of cases. In such realistic epidemiologic scenarios, the relation between genetic diversity and incidence of infection is more complex. In our study on hepatitis B virus in Amsterdam, the Netherlands, over the period 1992–2006 (2), we observe that the magnitude of change in genetic diversity does not correspond to the observed magnitude of change in incidence of infection. Therefore, we caution that genetic diversity may not be easily interpreted as if it were proportional to incidence of infection.

In their commentary, professors Halloran and Holmes describe a number of processes that could explain such a discrepancy in the change in genetic diversity and the change in incidence (1). We agree that it is important to assess the potential role of such processes for the inferences made. For the particular case of hepatitis B virus, it is possible to assess the relative likelihood of these possible explanations. Vaccine escape, for instance, is an important factor to consider in molecular epidemiologic analysis. Nevertheless, recent spread of a vaccine escape mutation in hepatitis B virus is very unlikely because the hepatitis B virus vaccine gives broad protection (3); even though the majority of the symptomatic acute hepatitis B virus infections in the Netherlands have been sequenced as of 2004, no patients infected with a vaccine escape mutant were found (4). Increased risk behavior is another important factor to consider. We have observations that support such an increased risk behavior for this specific host population. When we compare both factors, we think that it is more likely that the observed reduction in genetic diversity stems from an observed increase in risk behavior than from recent spread of a vaccine escape mutation.

Much of the coalescent analysis relies on the Bayesian skyline plot—Figure 4 in our article (2). This plot presents, at each time point, the median and 2.5 and 97.5 percentiles for a large number of plausible trajectories for genetic diversity through time (by “plausible” we mean that the trajectories are sampled from a Bayesian posterior distribution). We agree with professors Halloran and Holmes (1) that such a plot should not be used to assess the strength of evidence for a decline in genetic diversity. As an alternative, we suggest assessing the statistical evidence for a decline by examining the fraction of plausible trajectories that show a decline in genetic diversity. For our data set, we observe a decline in more than 95% of all plausible
trajectories and therefore conclude that there is statistical evidence for a decline in genetic diversity.

The plot reveals a large uncertainty about the genetic diversity at any point in time, which is of course due to the small number of sequences that were available. We note that genetic sequence data provide information about not only the sampled patients but also the transmission chains that eventually led to infection of the sampled patients. Any analysis of such small data sets should invoke appropriate statistical checks to prevent overinterpreting any information from these small data sets. The most obvious way to avoid overinterpretation of small data sets is to collect larger data sets; the data set we analyzed will grow larger as more samples are added over time. Nonetheless, forging a stronger foundation for coalescent analysis in future vaccination studies will benefit from both a careful mathematical reanalysis of assumptions in coalescent theory in realistic epidemiologic scenarios and the availability of a number of suitable data sets that enable testing of the theory against data. The description of idealized data sets offered by professors Halloran and Holmes (1) provides a welcome guide to collecting the required data sets.

ACKNOWLEDGMENTS

Author affiliations: Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, the Netherlands (W. Marijn van Ballegooijen, Hein J. Boot, Roel A. Coutinho, Jacco Wallinga); Department of Infectious Diseases, Public Health Service, GGD, Amsterdam, the Netherlands (Robin van Houdt, Sylvia M. Bruisten); Department of Human Retrovirology, Amsterdam Medical Center, University of Amsterdam, Amsterdam, the Netherlands (Roel A. Coutinho); and Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands (Jacco Wallinga).

Conflict of interest: none declared.

REFERENCES