American Journal of Epidemiology

Vol. 164, No. 10
DOI: 10.1093/aje/kwj318
Advance Access publication September 12, 2006

Invited Commentary

Invited Commentary: Challenges of using Contact Data to Understand Acute Respiratory Disease Transmission

M. Elizabeth Halloran

From the Fred Hutchinson Cancer Research Center and the University of Washington, Seattle, WA.

Received for publication April 7, 2006; accepted for publication April 21, 2006.

Wallinga et al. (1) have done an excellent job of demonstrating how simple data can be used to improve our estimation of transmission parameters for infectious disease models. Their analysis involves several steps illustrating a unifying framework, from collecting data on social contacts to model-fitting with relevant infectious disease data. First, there is estimation of the contact matrix from age-specific data on conversations. Second, assumptions are made to estimate the age-specific transmission parameters of a particular transmission model. Third, comparison is made with infectious disease outcome data for goodness of fit and model choice.

The problem is very important. Contact patterns are crucial determinants in both the spread of an infectious disease and the decision on which interventions would be most effective. Sometimes a group of investigators pulls together several ideas, in preliminary form, showing the way for future research. Such is the case with this paper by Wallinga et al. I would like to comment on four areas that deserve additional research: 1) data structure, 2) model-dependent transmission parameters, 3) statistical inference, and 4) infectious disease data for model-fitting.

In the paper by Wallinga et al. (1), the data on which the contact matrix analysis was based were remarkably simple. A random sample of people in Utrecht, the Netherlands, were asked about the number of conversations they had with people of different age groups during a typical week. These simple data were adequate to estimate an age-structured contact matrix. The simple age-structured matrix was in turn adequate to estimate transmission parameters for an age-structured model. In this type of model, the mixing groups are mutually exclusive; that is, a person can belong to only one age group. The transmission parameters that are estimated in this paper are specific to the type of model being used.

However, many current models being used to study the effects of interventions in populations, such as those for pandemic influenza (2–5) and smallpox (6), have more complex population structures. In these models, people can mix in several different places, including households, schools, and workplaces, as well as have age-specific components to the mixing within the different mixing groups. These more complex patterns are required to analyze the effect of household interventions, such as targeted antiviral prophylaxis, school closures, or quarantines. I would like to see more empirical studies like the one presented in this paper with these more complex models in mind (7). A natural extension of the data structure would be to ask people not only about the age groups of people with whom they have had conversations but also where they had the conversations. Such data would allow estimation of transmission parameters for models with more complex mixing structures. Since the interpretation of transmission parameters is model-specific, such data for estimation of transmission parameters would be very important.

The third area, statistical inference, also needs further development to take honest account of the uncertainty in the estimates. The current analysis probably underestimates the uncertainty in the mixing matrix, which then carries over to an underestimate of uncertainty in the estimates for the transmission model. First, the bootstrap confidence intervals presented in Wallinga et al.'s table 1 are only for the mean values of the negative binomial distributions that were fitted to the data. There is no mention of k, the shape parameter, so we do not know the full uncertainty of the entire distribution. Second, to estimate the age-specific transmission parameters, Wallinga et al. keep the estimated means fixed at the maximum likelihood value. Heterogeneity in the number of contacts within age groups and the variability of the data

Correspondence to Dr. M. Elizabeth Halloran, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, MS LE-400, Seattle, WA 98109 (e-mail: betz@scharp.org).

are not taken into account. Eventually, joint estimation of the contact matrix and the transmission-model parameters, probably within a Bayesian framework, would give us a better idea of the overall uncertainty. Third, although Wallinga et al. address the issue of nonresponse in their paper, they do not examine the potential for bias in their estimates if indeed there had been nonignorable nonresponse. With only 59 percent of the original sample included in the analysis, the potential for selection bias was high. Thus, the uncertainty in the estimates of the basic reproduction number R_0 is probably much greater than is reflected in their table 2. Fourth, the influenza model was fitted using a sample of only 128 participants from the 1957 Asian influenza pandemic. The small sample also probably contributed more uncertainty to the fit of the model.

The fourth area of concern is the lack of availability of good household and community-level infection and disease data with which to compare the results of the model estimation. I would like to see many more empirical field studies carried out to collect the needed appropriate data.

I commend Wallinga et al. (1) for their creative use of an important data set. They have illustrated the potential for simple data to contribute to the important problem of estimation of contact structures in dynamic infectious disease models. I found the paper helpful. For example, my colleagues and I compared the next generation matrix of our influenza simulator (3) with that of Wallinga et al. It was qualitatively similar. In particular, the diagonal entries of both were higher than the off-diagonal entries, indicative of higher mixing with people of one's own age group. Also, in both, the rates were highest among school-age children. The similarity to the empirically estimated next generation matrix lent support to the validity of the one used for our artificially constructed population. Rather than view the points raised here as deficiencies in the paper by Wallinga et al., I view them as challenges for future research.

ACKNOWLEDGMENTS

This research was partially supported by National Institute of Allergy and Infectious Diseases grant R01-AI32042 and National Institute of General Medical Sciences MIDAS grant U01-GM070749.

Conflict of interest: none declared.

REFERENCES

1. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am J Epidemiol 2006; 164:936–44.
2. Longini IM Jr, Halloran ME, Nizam A, et al. Containing pandemic influenza with antiviral agents. Am J Epidemiol 2004;159:623–33.
3. Longini IM Jr, Nizam A, Xu S, et al. Containing pandemic influenza at the source. Science 2005;309:1083–7.
4. Ferguson NM, Cummings DA, Cauchemez S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 2005;437:209–14.
5. Germann TC, Kadau K, Longini IM Jr, et al. Mitigation strategies for pandemic influenza in the United States. Proc Natl Acad Sci U S A 2006;103:5935–40.
6. Halloran ME, Longini IM Jr, Nizam A, et al. Containing bioterrorist smallpox. Science 2002;298:1428–32.
7. Koehly LM, Goodreau SM, Morris M. Exponential family models for sampled and census network data. Sociol Methodol 2004;34:241–70.