Commentary

Bias in Pertussis Incidence Data and Its Implications for Public Health Epidemiology

Eli E. Goldwyn, PhD; Pejman Rohani, PhD

After a decline of more than 99% from 1934 to 1976, pertussis incidence in the United States has increased dramatically over the last 3 decades.[1,2] Although the initial decrease was attributed to the introduction of vaccination programs,[3] the much-publicized recent resurgence has occurred despite continued high-vaccination uptake estimates.[4] Unfortunately, an incomplete understanding both of the underlying factors causing this resurgence in incidence[2,5] and of the key drivers of pertussis transmission in general confounds creation of effective vaccination strategies and public health responses. A powerful approach to dissecting and elucidating the relative contributions of these drivers is the development of disease transmission models and their confrontation—via statistical inference methodology—with highly resolved incidence data.[6,7] Despite the existence of highly resolved data collected through the Supplementary Pertussis Surveillance System (SPSS), and generous stated policy guidelines regarding data sharing/release,[8] these data are not publicly accessible for researchers. Here, we discuss the lack of resolution and important biases in the incidence reports that are publicly available and argue in favor of more transparency in data sharing policies and broader dissemination of incidence data such as those collated through the SPSS.

In addition to mortality and morbidity, the public health consequences of pertussis include an economic burden from lost productivity,[9] as well as financial expenditures in areas such as associated medical costs, public awareness campaigns, and increased vaccinations. The most significant increase in pertussis in the United States (and other developed countries with vaccine programs) has occurred among adolescents and adults, with smaller increases observed for infants and older children.[5] Nonetheless, infants make up the plurality of reported cases and nearly all of the fatalities, which primarily occur in those who are undervaccinated.

To reduce the prevalence of pertussis, the Advisory Committee on Immunization Practices recommends several targeted strategies for booster vaccinations. These recommendations include vaccination for middle-school students (now a requirement for school attendance in several states), efforts to "cocoon" infants (which include prenatal[10] or postnatal[11] maternal vaccination, as well as vaccination of other household contacts), and the vaccination of health care workers. As of February 2012, the committee recommends that every adult who has not had a booster receive one.[12] The Global Pertussis Initiative has considered 7 vaccination strategies,[13] which, in addition to the aforementioned recommendations, include selective immunization of child care workers and improvements in infant and toddler immunization strategies.

Contributing to the difficulty in creating optimal vaccination strategies is an uncertainty surrounding the cause of the resurgence. This is largely due to a number of incompletely understood and hotly debated factors in the epidemiology and transmission of pertussis: (*a*) the uncertain frequency of subclinical pertussis infections and their contribution to transmission;[14,15] (*b*) the determinants and duration of protective immunity, both resulting from infection and immunization;[16-18] (*c*) continued speculation regarding

---

Author Affiliations: Department of Mathematics, University of California, Davis (Dr Goldwyn); Department of Ecology and Evolutionary Biology, and Center for the Study of Complex Systems, University of Michigan, Ann Arbor (Dr Rohani); and Fogarty International Center, National Institutes of Health, Bethesda, Maryland (Dr Rohani).

This work is supported by a grant from the National Institutes of Health (R01AI101155).

Disclosur: The authors declare no conflicts of interest.

Correspondence: Eli E. Goldwyn, PhD, Department of Mathematics, University of California, 1 Shields Ave, Davis, CA 95616 (eegoldwyn@ucdavis.edu).

DOI: 10.1097/PHH.0b013e31826d7f95

Journal of Public Health Management and Practice, 2012, 1–4

the impact of vaccination on transmission;[19] (*d*) evolutionary dynamics that may have led to the emergence of novel pertussis strains;[20] (*e*) the consequences of contact networks on shifting age distribution of pertussis incidence;[21] and (*f*) emerging evidence that transmission is not only seasonal but also seasonality varies according to age groups.[22]

Confronting mathematical disease transmission models with high-resolution data through the use of statistical inference is a valuable tool often used to elucidate the underlying drivers and relative roles in disease dynamics. Such models have successfully identified central mechanisms in the epidemiology of a variety of infectious diseases,[7,23-27] including pertussis in other nations.[16,18,26,28] Beyond shedding light on the underlying determinants of epidemiology, these models have played a critical role in predicting the effects of various vaccination policies and assessing the economic impact of control strategies across a variety of infectious diseases.[6,27] They have, moreover, aided the determination of optimal vaccination strategies, including the schedule and level of uptake necessary to prevent outbreaks.[28-30] In the context of pertussis in countries other than the United States, models have been invaluable for estimating the effectiveness of various permutations of adolescent, adult, and cocooning vaccination strategies.[31-33] Furthermore, such approaches have permitted analyses of the cost-effectiveness of a variety of strategies, including each of the 7 strategies recommended by the Global Pertussis Initiative in the United States,[30] and specifically focus on boosters for 4- and 15-year-olds,[31] adults,[32] and pediatric health care workers.[33]

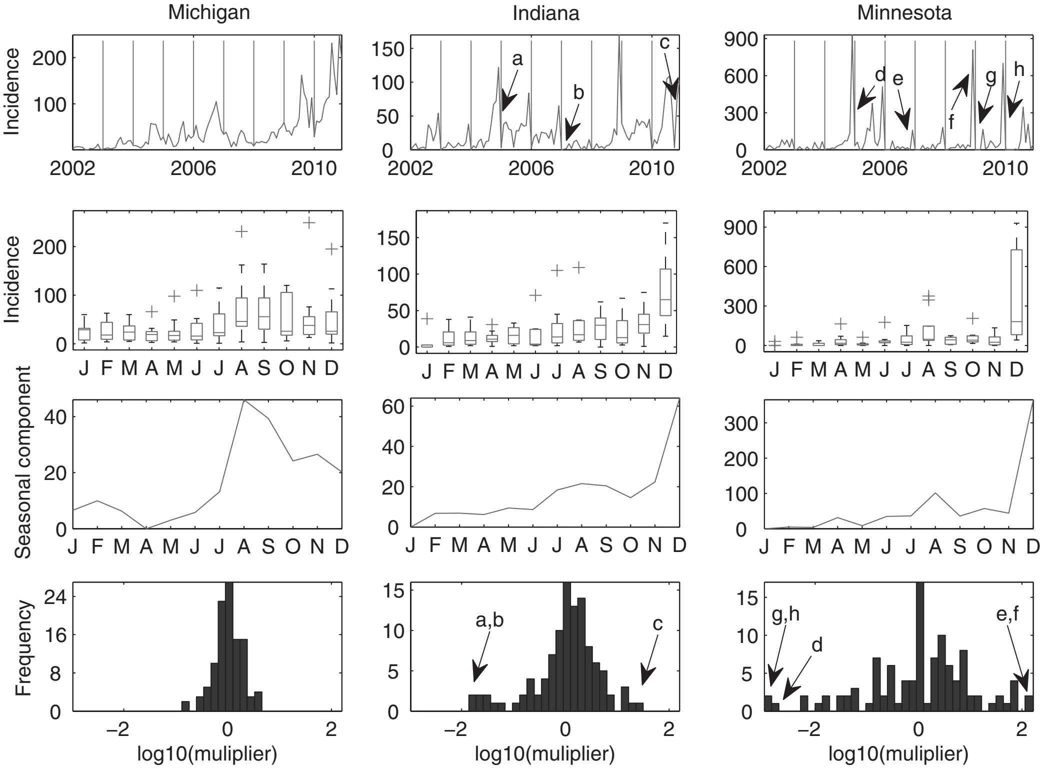
Ultimately, the veracity of model predictions relies first and foremost on careful "ground truthing" against epidemiological data. Since 1951, pertussis cases in the United States have been passively reported by physicians or laboratories to the local and state health departments. These health departments complete further investigations of each case and then report limited data to the Center for Disease Control and Prevention through its National Notifiable Disease Surveillance System (NNDSS). Starting in 1979, additional information has been collected through the SPSS, including individual patient date of onset, demographic data, vaccination history, clinical symptoms, and hospitalization records.[34] Despite the collation of this wealth of data and a recent movement to improve sharing of data (particularly when data gathering is taxpayer funded[35]), access remains an urgent problem. At present, the only publicly available nationwide information comes from the NNDSS and is released via the Morbidity and Mortality Weekly Reports in the form of (*a*) total number of weekly cases in the United States, (*b*) monthly incidence in each state, or (*c*) annual incidence in each county.

A significant hurdle in the use of these data is that provisional NNDSS data are not corrected for state-specific reporting procedures and are subject to delays or batching due to outbreaks, staffing issues, and other health conditions taking priority.[36] Specifically, "provisional data may be batched reported during outbreaks and at other times, including at the end-of-the year when surveillance staff are trying to finalize the data for a given year."[36] Although these provisional data are said to be finalized approximately 6 months after the end of that calendar year, as we show in the Figure, this is not necessarily the case.

To illustrate the impediments faced in interpreting such data, the Figure displays pertussis incidence data from the NNDSS for 3 representative states for the years 2002-2011. Michigan appears to have finalized data, whereas Indiana and Minnesota have imprecisely batched data, demonstrating reporting biases that cannot be explained strictly by state-specific reporting protocol. The top row of the Figure shows the incidence of pertussis per 100 000 individuals, with vertical lines corresponding to January of each year. These vertical lines delineate the large spikes in December incidence relative to November and January for the states of Indiana and Minnesota. Although we expect incidence to vary within each year (as it does in Michigan) due to seasonality in transmission, we do not expect the large differences in incidence that are seen between December and its neighboring months. Michigan experiences nothing similar to this pattern despite being both nearby geographically and similar in terms of vaccination rates and the average incidence during this time period. The second row of the Figure displays the same time series data in a box plot of monthly incidence. We formally quantify these monthly differences in the third row of the Figure through signal processing techniques by plotting the seasonal component of the data. Although we have highlighted Indiana and Minnesota, many other states demonstrate a similar phenomenon of surprisingly high incidence in December relative to other nearby months.

To further demonstrate the inaccuracies in the monthly incidence, the bottom row of the Figure displays a histogram of the logarithm of the ratio between the number of reported cases in the current month and the previous month (if no case was reported in a particular month, we allowed there to be 1 case so that the ratio calculation would be defined; this is a reasonable assumption, given the effects of immigration and low reporting rates, estimated to be in the neighborhood of 10%). Although for Michigan, this ratio is never less than 1:7 or greater than 5:1, Indiana has monthly ratios as high as 25:1 and as low as 1:65. Even more astoundingly, Minnesota has ratios greater than 165:1 and less than 1:808.

FIGURE ● **A Comparison of the Reported Pertussis Cases Binned Monthly in Michigan, Indiana, and Minnesota From the National Notifiable Disease Surveillance System for 2002-2011[a].**



[a] The top row is the time-series data collected monthly, with vertical lines every January. We label outlier months (a)-(h) as those having especially large ratios in incidence change. These same letters correspond to those in the bottom row. The second row is a box plot of this same time-series data binned by month. The edges of the box represent the 25th and 75th percentiles, and the middle line is the median. The third row is the seasonal component of the data found by using the STL Decomposition Package in R. The seasonal component is found by removing the trend and the residual from the time-series data. For clarity, this function is shifted vertically so that the minimum is exactly zero. The bottom row is the $\log_{10}$ of the incidence in the current month divided by the incidence in the previous month (if the month had zero incidence, 1 incident was assumed). The minimum value in Minnesota corresponds to a drop from 808 reported cases to zero.

Although a better understanding of pertussis epidemiology is critical to formulate effective vaccination policies, without unbiased and highly resolved data the insights arising from quantitative exploration will be either severely restricted or, worse still, wrong. Ideally, pertussis epidemiology would be informed by careful analyses of incidence data arising from active surveillance, because passive systems rely on reporting and diagnostic practices that vary over time and geographic location and are known often to dramatically underestimate the true incidence.[4] Given the rarity of active surveillance, we argue that the dissemination of data from the SPSS (specifically the data of onset and age information) would substantially augment NNDSS information. In addition, with the potentially important role of spatially varying outbreaks in driving the epidemiology of pertussis,[26] the data in the SPSS come from such a large physical area could prove to be uniquely valuable.

Given the increasing health toll exacted on US populations by pertussis and the accompanying public hysteria, the need for better epidemiology studies that would lead to more effective control is timely and important. Yet despite the seemingly generous policy guidelines regarding data sharing/release,[8] in practice, data sharing in the United States is insufficient, with access to critical extant sources of data remaining both infrequent and idiosyncratic. With California experiencing 10 infant fatalities in 2010 alone, the urgent need for rigorous understanding and improved immunization strategies is clear.

REFERENCES

1. Rohani P, Drake JM. The decline and resurgence of pertussis in the United States. *Epidemics*. 2011;3:183-188.
2. Bass J, Stephenson S. The return of pertussis. *Pediatr Infect Dis J*. 1987;6:141-144.
3. Centers for Disease Control and Prevention. National immunization survey. http://www.cdc.gov/vaccines/stats-surv/nis/default.htm\#nis. Published 2010. Accessed August 30, 2012.

4. Tan T, Trindade E, Skowronski D. Epidemiology of pertussis. *Pediatr Infect Dis J.* 2005;24:S10-S18.
5. Cherry JD, Grimprel E, Guiso N, Heininger U, Mertsola J. Defining pertussis epidemiology: clinical, microbiologic and serologic perspectives. *Pediatr Infect Dis J*. 2005;24:S25-S34.
6. Anderson RM, May RM. *Infectious Diseases of Humans Dynamics and Control*. Oxford: Oxford University Press; 1992.
7. Keeling M, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton University Press; 2008.
8. Centers for Disease Control and Prevention. Releasing and sharing data. http://www.cdc.gov/od/foia/policies/sharing.htm. Accessed August 30, 2012.
9. Cortese MM, Baughman AL, Brown K, Srivastava P. A "new age" in pertussis prevention: new opportunities through adult vaccination. *Am J Prev Med*. 2007;32:e177-e185.
10. Mooi F, de Greeff S. The case for maternal vaccination against pertussis. *Lancet Infect Dis*. 2007;7:614-624.
11. Healy CM, Rench MA, Baker CJ. Implementation of cocooning against pertussis in a high-risk population. *Clin Infect Dis*. 2011;52:157-162.
12. Center for Disease Control and Prevention. Updated recommendations for use of tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis (Tdap) vaccine in adults ages 65 years and older—Advisory Committee on Immunization Practices (ACIP), 2012. *MMWR Morb Mortal Wkly Rep*. 2012;61(25);468-470.
13. Forsyth K, Tan T, Von Knig CHW, Caro JJ, Plotkin S. Potential strategies to reduce the burden of pertussis. *Pediatr Infect Dis J* 2005;24:S69-S74.
14. Fine P, Clarkson J. Distribution of immunity to pertussis in the population of England and Wales. *J Hyg (Lond)*. 1984;92:21-26.
15. Broutin H, Mantilla-Beniers N, Simondon F, et al. Epidemiological impact of vaccination on the dynamics of two childhood diseases in rural Senegal. *Microbes Infect*. 2005;7:593-599.
16. Wearing H, Rohani P. Estimating the duration of pertussis immunity using epidemiological signatures. *PLoS Pathog*. 2009;5:11.
17. Wendelboe A, Rie AV, Salmaso S, Englund J. Duration of immunity against pertussis after natural infection or vaccination. *Pediatr Infect Dis J*. 2005;24:S58-S61.
18. Broutin H, Viboud C, Grenfell B, Miller M, Rohani P. Impact of vaccination and birth rate on the epidemiology of pertussis: a comparative study in 64 countries. Proc Biol Sci. 2010;277:3239-3245.
19. Cherry JD. The present and future control of pertussis. *Clin Infect Dis*. 2010;51:663-667.
20. Mooi FR. Bordetella pertussis and vaccination: the persistence of a genetically monomorphic pathogen. *Infect Genet Evol*. 2010;10:36-49.
21. Rohani P, Zhong X, King A. Contact network structure explains the changing epidemiology of pertussis. *Science*. 2010;330:982-985.
22. Tanaka M, Vitek C, Pascual F, Bisgard K, Tate J, Murphy TV. Trends in pertussis among infants in the United States, 1980-1999. *JAMA*. 2003;290:2968-2975.
23. Ferrari MJ, Grais RF, Bharti N, et al. The dynamics of measles in sub-Saharan Africa. *Nature*. 2008;451:679-684.
24. King AA, Ionides EL, Pascual M, Bouma MJ. Inapparent infections and cholera dynamics. *Nature*. 2008;454:877-880.
25. Metcalf CJE, Bjørnstad ON, Grenfell BT, Andreasen V. Seasonality and comparative dynamics of six childhood infections in pre-vaccination Copenhagen. *Proc Biol Sci*. 2009;276:4111-4118.
26. Rohani P, Earn D, Grenfell B. Opposite patterns of synchrony in sympatric disease metapopulations. *Science*. 1999;286:968-971.
27. Ferguson NM, Cummings DA, Cauchemez S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*. 2005;437:209-214.
28. Hethcote H. Simulations of pertussis epidemiology in the United States: effects of adult booster vaccinations. *Math Biosci*. 1999;158:47-73.
29. Coudeville L, Van Rie A, Getsios D, Caro JJ, Crépey P, Nguyen VH. Adult vaccination strategies for the control of pertussis in the United States: an economic evaluation including the dynamic population effects. *PLoS One*. 2009;4:e6284.
30. Van Rie A, Hethcote H. Adolescent and adult pertussis vaccination: computer simulations of five new strategies. *Vaccine*. 2004;22:3154-3165.
31. Edmunds W, Brisson M, Melegaro A, Gay NJ. The potential cost-effectiveness of acellular pertussis booster vaccination in England and Wales. *Vaccine*. 2002;20:1316-1330.
32. Lee GM, Murphy TV, Lett S, et al. Cost effectiveness of pertussis vaccination in adults. *Am J Prev Med*. 2007;32:e186-e193.
33. Greer AL, Fisman DN. Use of models to identify cost-effective interventions: pertussis vaccination for pediatric health care workers. Pediatrics. 2011;128:e591-e599.
34. Wharton M, Strebel P, Roush S. Chapter 23: national surveillance of vaccine-preventable diseases. In: *Manual for the Surveillance of Vaccine-Preventable Diseases*. 4th ed. http://www.cdc.gov/vaccines/pubs/surv-manual/chpt23-natl-surv-vpd.html. Published 2008. Accessed August 30, 2012.
35. Walport M, Brest P. Sharing research data to improve public health. *Lancet*. 2011;377:537-539.
36. Centers for Disease Control and Prevention. National Notifiable Diseases Surveillance System. http://www.cdc.gov/osels/ph_surveillance/nndss/phs/files/ProvisionalNationa%20NotifiableDiseasesSurveillanceData20100927.pdf. Accessed August 30, 2012.