Supporting Information for **On the role of different age groups during pertussis epidemics in California, 2014 and 2010**

Ayesha Mahmud1,\*, Marc Lipsitch1,2 , Edward Goldstein1

1. Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave., Boston, MA 02115, USA

2) Department of Immunology and Infectious Disease, Harvard T.H. Chan School of Public

 Health, 677 Huntington Ave., Boston, MA 02115, USA

**Section S1:** Selection of regions and starting and ending weeks for the 2010 and 2014 epidemics

Table S1 shows the starting and ending weeks for the major waves of the 2010 and the 2014 pertussis epidemics in 6 different regions in California. Each region comprises a number of counties whose names are listed in Table S1. The selection of counties that constitute different regions was largely guided by the list of economic regions of California [1], with some exceptions. In particular, we’ve grouped the San Diego, Imperial, San Bernardino, and Riverside counties into one region, excluded several counties east of the Sierra Nevada and in the northern part of the state, and included the Mariposa, Tuolumne, Calaveras, and Amador counties into region 3. The starting and ending weeks were selected by inspection of Figures 1 and 2 in the main text to be the weeks when the ascent of the first major wave of the epidemic began, and the descent of that wave ended, correspondingly. Regions that did not have a pronounced epidemic peak (as suggested by visual inspection) were not included in the relative risk (RR) analysis as such regions may comprise further sub-regions with asynchronous epidemics, leading to a misclassification of a large number of cases as being before or after the peak of the epidemic.

|  |  |  |
| --- | --- | --- |
| Region  |  2010 |  2014 |
| Starting week | Ending week | Starting week | Ending week |
| 1. San Diego, Imperial, San Bernardino, Riverside  |  X |  X |  8 |  35 |
| 2. Los Angeles, Orange, Ventura |  10 |  49 |  5 |  38 |
| 3. Kern, Tulare, Fresno, Madera, Merced, Mariposa, Tuolumne, Stanislaus, San Joaquin, Amador, Calaveras  |  6 |  40 |  10 |  38 |
| 4. San Benito, Santa Clara, Santa Cruz, San Mateo, Alameda, Contra Costa, San Francisco, Marin, Solano, Napa, Sonoma  |  8 |  38 |  6 |  33 |
| 5. Yolo, Sacramento, El Dorado, Yuba, Placer, Sutter, Butte, Nevada, Colusa, Glenn, Tehama, Shasta |  X |  X |  5 |  33 |
| 6. Santa Barbara, San Luis Obispo, Monterey |  3 |  39 |  X |  X |

**Table S1**: Starting and ending weeks for the major waves of the 2010 and the 2014 pertussis epidemics (calendar weeks for the corresponding year) in different regions in California. “X” indicates that the region wasn’t selected for the relative risk analysis.

The 2010 epidemic had 5,422 reported cases in the selected regions between the starting and ending weeks of the epidemic in those regions; the 2014 epidemic had 7,440 reported cases in the selected regions between the starting and ending weeks of the epidemic in those regions.

**Section S2:** Odds ratios for different pairs of age groups

The relative risk (RR) statistic (eq. 1 in the main text) allows for the simultaneous comparison of all age groups in terms of the relative depletion of susceptible individuals in those age groups before the epidemic peak. For a pairwise comparison of different age groups , we consider reported pertussis cases that were either in or , and evaluate the odds ratio for being in vs for cases before vs after the epidemic peak. The estimate means that the proportion of cases in among all cases in the two age groups had decreased after the peak, suggesting a higher depletion of susceptible individuals in compared to before the epidemic peak; the estimate suggests a higher depletion of susceptible individuals in compared to by the time of the epidemic peak. We compute the odds ratio using a logistic regression model, adjusting for whether the case was Hispanic, and whether the case was white. We note that the unadjusted odds ratio for a pair of age groups , is simply the ratio of the relative risks and in those age groups [2]:

Here and are given by eq. 1 in the main text, and eq. S1 can be established simply by plugging the expressions for and in eq. 1 in the main text into the ratio .

Tables S2 and S3 present the estimates of the odds ratios for different pairs of age groups during the 2010 and the 2014 epidemics in California.

Table S2 suggests a lower depletion of susceptible individuals in children aged 2-4y compared to all other age groups during the 2010 epidemic, and a higher depletion of susceptible individuals in children aged 9-13y compared to adults aged over 20y during the 2010 epidemic.

Table S3 suggests, among other things, a greater depletion of susceptible individuals in adolescents aged 14-15y compared to adolescents aged 11-13y and 16-19y during the 2014 epidemic; a higher depletion of susceptible individuals in adolescents aged 11-19y compared to other age groups during the 2014 epidemic; a greater depletion of susceptible individuals in children aged 7-10y compared to children aged 2-6y during the 2014 epidemic; and a greater depletion of susceptible individuals in children aged 9-10y compared to infants aged <1y and adults aged over 20y during the 2014 epidemic.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1-2y | 3-4y | 5-6y | 7-8y | 9-10y | 11-13y | 14-15y | 16-19y | 20+y |
| <1y | **1.45****(1.13,1.85)** | **1.56****(1.18,2.06)** | 1.01(0.73,1.39) | 0.92(0.71,1.2) | 0.85(0.68,1.08) | 0.83(0.66,1.04) | 0.82(0.58,1.17) | 0.92(0.67,1.26) | 1.06(0.88,1.29) |
| 1-2y |  | 1.06(0.77,1.47) | **0.67****(0.47,0.96)** | **0.62****(0.45,0.84)** | **0.57****(0.43,0.76)** | **0.55****(0.42,0.72)** | **0.53****(0.36,0.79)** | **0.61****(0.43,0.87)** | **0.72****(0.56,0.92)** |
| 3-4y |  |  | **0.66****(0.45,0.96)** | **0.6****(0.43,0.83)** | **0.56****(0.41,0.75)** | **0.54****(0.4,0.73)** | **0.53****(0.35,0.8)** | **0.6****(0.41,0.87)** | **0.69****(0.52,0.91)** |
| 5-6y |  |  |  | 0.9(0.63,1.29) | 0.85(0.61,1.18) | 0.81(0.59,1.13) | 0.85(0.56,1.29) | 0.91(0.62,1.35) | 1.06(0.78,1.44) |
| 7-8y |  |  |  |  | 0.94(0.71,1.24) | 0.9(0.69,1.19) | 0.95(0.65,1.38) | 1(0.7,1.41) | 1.17(0.91,1.5) |
| 9-10y |  |  |  |  |  | 0.95(0.75,1.2) | 1.02(0.72,1.45) | 1.07(0.78,1.47) | **1.27****(1.03,1.56)** |
| 11-13y |  |  |  |  |  |  | 1.07(0.75,1.52) | 1.1(0.8,1.51) | **1.33****(1.09,1.63)** |
| 14-15y |  |  |  |  |  |  |  | 1.06(0.7,1.6) | 1.24(0.89,1.72) |
| 16-19y |  |  |  |  |  |  |  |  | 1.17(0.87,1.57) |

**Table S2**: Odds ratios for being before vs after the peak of the 2010 pertussis epidemic in California for reported pertussis cases in members of one age group vs another for different pairs of age groups. Significant estimates are marked in bold.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1-2y | 3-4y | 5-6y | 7-8y | 9-10y | 11-13y | 14-15y | 16-19y | 20+y |
| <1y | **1.43****(1.11,1.84)** | **1.52****(1.15,2.02)** | **1.67****(1.21,2.29)** | 0.85(0.66,1.09) | **0.75****(0.6,0.94)** | **0.62****(0.5,0.76)** | **0.43****(0.34,0.53)** | **0.6****(0.47,0.76)** | 1.02(0.79,1.3) |
| 1-2y |  | 1.07(0.79,1.45) | 1.17(0.84,1.63) | **0.61****(0.46,0.8)** | **0.53****(0.42,0.69)** | **0.43****(0.34,0.55)** | **0.3****(0.24,0.38)** | **0.41****(0.32,0.53)** | **0.71****(0.55,0.93)** |
| 3-4y |  |  | 1.09(0.77,1.55) | **0.57****(0.42,0.76)** | **0.51****(0.38,0.66)** | **0.41****(0.31,0.53)** | **0.28****(0.21,0.37)** | **0.38****(0.29,0.5)** | **0.67****(0.5,0.88)** |
| 5-6y |  |  |  | **0.52****(0.38,0.73)** | **0.46****(0.34,0.63)** | **0.37****(0.28,0.5)** | **0.26****(0.19,0.35)** | **0.35****(0.26,0.48)** | **0.61****(0.45,0.84)** |
| 7-8y |  |  |  |  | 0.89(0.7,1.13) | **0.71****(0.57,0.89)** | **0.49****(0.39,0.62)** | **0.66****(0.52,0.84)** | 1.15(0.9,1.48) |
| 9-10y |  |  |  |  |  | **0.79****(0.65,0.97)** | **0.55****(0.45,0.67)** | **0.73****(0.59,0.91)** | **1.29****(1.03,1.62)** |
| 11-13y |  |  |  |  |  |  | **0.69****(0.57,0.83)** | 0.92(0.76,1.12) | **1.65****(1.33,2.03)** |
| 14-15y |  |  |  |  |  |  |  | **1.33****(1.09,1.62)** | **2.38****(1.93,2.94)** |
| 16-19y |  |  |  |  |  |  |  |  | **1.8****(1.45,2.24)** |

**Table S3**: Odds ratios for being before vs after the peak of the 2014 pertussis epidemic in California for reported pertussis cases in members of one age group vs another for different pairs of age groups. Significant estimates are marked in bold.

**Section S3:** RR estimation in a Bayesian framework

Derivation of the confidence bounds for the RR estimates using eq. 2 in the main text relies on a normal approximation that is valid for large sample sizes. Here, we derive estimates/credible intervals for the RR statistic in different age groups in a Bayesian framework.

For each epidemic and each age group , cases occurring before the outbreak peak in each region included in the analysis were combined, with their total number denoted by , and the same applies to cases occurring after the peak, with their number denoted by . The expression for the relative risk for the age group is the ratio of the proportions of cases in the group among all reported cases in the population before the peak and after the peak as in eq. 1 in the main text (here in the sum runs over all age groups).

The estimates and confidence bounds for relative risks in each group can be obtained in a Bayesian framework based on the observations following the methodology in [2]. Briefly, if the number of cases of pertussis infection in each age group is large and case-reporting rates (proportion of cases of pertussis infection in a given age group that are reported to the California Department of Public Health) are low [3], the observed numbers of reported cases in each age group before and after the peak are Poisson distributed (with unknown true Poisson parameters). Posterior samples for the Poisson parameters (with a flat prior) corresponding to the observed counts are generated; for each i=1,..,100000, the corresponding parameters are plugged into equation (S2) to generate an estimate for the relative risk in the age group . The mean and the credible interval for the sample (i=1,..,100000) are then extracted. Table S4 exhibits the RR estimates in different age groups for the 2010 and the 2014 pertussis epidemics in California. Those estimates are very similar to the estimates in Table 1 in the main text.

|  |  |  |
| --- | --- | --- |
| Age group | 2010-11 epidemic | 2014 epidemic |
| <1y | 1.02 (0.9,1.14) | 0.74 (0.63,0.85) |
| 1-2y | 0.69 (0.56,0.83) | 0.55 (0.45,0.65) |
| 3-4y | 0.65 (0.51,0.81) | 0.5 (0.4,0.62) |
| 5-6y | 1 (0.76,1.3) | 0.46 (0.35,0.59) |
| 7-8y | 1.11 (0.89,1.36) | 0.88 (0.74,1.05) |
| 9-10y | 1.19 (1.01,1.4) | 0.99 (0.85,1.13) |
| 11-13y | **1.26 (1.08,1.46)** | 1.26 (1.12,1.41) |
| 14-15y | 1.18 (0.86,1.58) | **1.83 (1.61,2.07)** |
| 16-19y | 1.13 (0.86,1.46) | 1.41 (1.24,1.61) |
| 20+y | 0.95 (0.85,1.05) | 0.79 (0.67,0.92) |

**Table S4:** RR estimates for different age groups during the 2010 and the 2014 pertussis epidemics in California in a Bayesian framework

**References**

[1] List of economic regions of California. <https://en.wikipedia.org/wiki/List_of_economic_regions_of_California>

[2] Worby CJ, Kenyon C, Lynfield R, Lipsitch M, Goldstein E (2015) Examining the role of different age groups, and of vaccination during the 2012 Minnesota pertussis outbreak. Sci Rep 5: 13182.

[3] Rendi-Wagner P, Tobias J, Moerman L, Goren S, Bassal R, et al. The seroepidemiology of Bordetella pertussis in Israel--Estimate of incidence of infection. Vaccine (2010); 28:3285–3290