

Supporting Information for “**On the relative role of different age groups during epidemics associated with respiratory syncytial virus**”

Section S1: Simulated epidemics

We use simulations of transmission dynamics in the context of RSV epidemiology to examine the relation between the age groups with the highest RR value for the simulated epidemics and the age groups for which the distribution of a given quantity of a perfect vaccine (1,000,000 doses) - alternatively, a removal of 1,000,000 individuals from the transmission process – would result in the largest reduction in the epidemic’s initial reproductive number. Our inference framework largely follows [1] (section 2.3 of the Methods), though here we utilize a more flexible model for the distribution of susceptibility to infection within each age group.

We simulated RSV outbreaks in an age-stratified population with 8 age groups (in years): (0–2,3-4,5-6,7-12,13-19,20-39,40-59,60+). Transmission dynamics are modeled in the stratified mass action SIR framework. Contacts between the different age groups (strata) are described by a symmetric matrix $C = (c_{ij})$, where c_{ij} is the average number of contacts per unit of time (day) between a pair of individuals in strata i and j . Our estimate of the matrix (c_{ij}) was informed by the POLYMOD study data [2,3], with further extrapolation for contact rates within age groups smaller than the ones considered in the POLYMOD study. Additionally, for each age group i , we have

1. Population size n_i , extracted from [4].
2. A proportion p_i of individuals within the age group i susceptible to infection, each having susceptibility q_i ($0 \leq p_i, q_i \leq 1$).

We assume that individual infectivity is age-independent. This assumption is partly based on the results in [5,6] that suggest no difference in the duration of viral shedding or viral loads by age in children. For susceptibility, let $E_i = p_i q_i$ be the average susceptibility of a person in an age group i . We note that for a given value of E_i , the variance for the susceptibility distribution of individuals in the age group i can vary from 0 (when $p_i = 1, q_i = E_i$) to $(1 - E_i)E_i$ (when $p_i = E_i, q_i = 1$), with $(1 - E_i)E_i$ being the maximal possible variance for the distribution of susceptibility X_i ($0 \leq X_i \leq 1$) within the age group i with the given mean $E(X_i) = E_i$.

The initial next generation matrix (N_{ij}) is therefore

$$N_{ij} = \lambda \cdot n_i \cdot p_i \cdot q_i \cdot c_{ij} \quad (*)$$

Here the parameter λ is fixed as described below. The largest eigenvalue of the initial next generation matrix is the effective initial reproductive number $R_{eff}(initial)$.

For an infected individual, their infectiousness will be spread in time via the serial interval distribution. We adopt the estimate of the serial interval distribution based on the data in [7] on RSV-positive household contacts of the initially recruited cases (Figure 1 in [7]). The serial interval is assumed to have a maximum length of 17 days, with a mean of 7.8 days.

The simulated epidemics are generated in a deterministic setting deterministically by specifying the initial distribution of susceptibility in each age group (namely the parameters p_i, q_i), and the numbers of the initially infected individuals. The latter are distributed at random between the 10 age groups at a rate of 6,000 a day for a period of 17 days (length of the serial interval) in a total population of 323.13 million (US population estimate in 2016, [4]). For the initial distribution of susceptibility in the different age groups, we assume that the average susceptibility E_i is decreasing with age. Given the high periodicity of the annual RSV epidemics, we assume a reasonably narrow range for the distribution of the parameter E_i , and draw this parameter uniformly from that range. Additionally, let

$$x_i = \log_{E_i}(p_i), \quad y_i = \log_{E_i}(q_i) \quad (**)$$

Since $E_i = p_i q_i$, we have $x_i + y_i = 1$. We will draw x_i from a Beta distribution, with $y_i = 1 - x_i$. Table S1 specifies the distributions for the (independently drawn) parameters E_i, x_i (with p_i evaluated as $E_i^{x_i}$, q_i evaluated as $E_i^{1-x_i}$, and the ranges for the parameters p_i, q_i for the simulated epidemics presented in Table S3).

Age group/ Parameter	Average initial susceptibility E_i	x_i (equation (**))
0-2 years	Uniform [0.9,1]	Beta(1,5)
3-4 years	Uniform [0.8,0.9]	Beta(1,3)
5-6 years	Uniform [0.6,0.75]	Beta(1,2)
7-12 years	Uniform [0.45,0.55]	Beta(1,2)
13-19 years	Uniform [0.4,0.5]	Beta(1,2)
20-39 years	Uniform [0.4,0.5]	Beta(1,2)
40-59 years	Uniform [0.35,0.45]	Beta(1,2)
60+ years	Uniform [0.3,0.4]	Beta(1,2)

Table S1: Parameters for the distribution of susceptibility within the different age strata.

We fix the value of the parameter λ in eq. (*) so that when the average initial susceptibility to infection in the different age groups is the largest possible for the ranges in Table S1 (namely equaling 1 for children aged 0-2 years, 0.9 for children aged 3-4 years, etc.), the epidemic's initial reproductive number is 1.4.

We have simulated 250,000 epidemics according to the model described above. Table S2 shows the number/percent of the 250,000 simulated epidemics for which the highest estimate of RR belongs to each age group, as well as number/percent of epidemics with the highest RR estimate belonging to a given age group for which vaccination of 1,000,000 individuals in that age group had the largest impact on reducing the epidemic's initial reproductive number/growth rate. The highest RR estimate belonged to children aged 3-4 years and 5-6 years in 47.6% of simulations each, to children aged 7-12 years in 2.8% of simulations, to adolescents aged 13-19 in 1.9% of simulations, and to adults aged 20-39 years in 0.05% (133/250,000) of simulations. Overall, epidemics for which the group with the highest RR estimate was also the group for which vaccination of 1,000,000 individuals yielded the biggest reduction in the epidemic's initial effective reproductive number/growth rate (*concordant epidemics*) constituted 73.5% of the simulated epidemics. The percentage of concordant epidemics varied from 86% for the simulated epidemics for which children aged 5-6 years had the highest RR estimate, to 68.3% for the simulated epidemics for which children aged 3-4 years had the highest RR estimate, to 0.1% for the simulated epidemics for which children aged 13-19 years had the highest RR estimate, to 0% (none) for the simulated epidemics for which either children aged 7-12 years or adults aged 20-39 years had the highest RR estimate.

Age group	Epidemics with the leading RR belonging to age group	Concordant epidemics with the leading RR belonging to age group
0-2 years	0	0
3-4 years	119104 (47.6%)	81353 (68.3%)
5-6 years	119077 (47.6%)	102411 (86%)
7-12 years	7502 (2.8%)	0 (0%)
13-19 years	4634 (1.86%)	6 (0.13%)
20-39 years	133 (0.05%)	0 (0%)
40-59 years	0	0
60+ years	0	0

Table S2: Number/percent of simulated epidemics for which the leading RR estimate belonged to each age group, and the number/percent of epidemics with the highest RR estimate belonging to a given age group for which vaccination of 1,000,000 individuals in that age group had the largest impact on reducing the epidemic's initial reproductive number/growth rate.

Next, we compared, for each age group, those epidemics for which the highest RR estimate belonged to that age group, but vaccination of 1,000,000 individuals in that age group did not have the largest impact on reducing the epidemic's initial reproductive number/growth rate (*discordant epidemics*) with the overall collection of the 250,000 simulated epidemics in terms of the distribution of susceptibility in the given age group. Specifically, we examined the proportion of susceptible individuals p_i and the average

susceptibility per susceptible individual q_i . The results of this comparison are summarized in Table S3. Table S3 suggests that for each age group, discordant epidemics with the highest RR estimate belonging to that age group had, on average, a lower fraction of susceptible individuals in that age group but a higher susceptibility to infection per susceptible individual compared to the whole collection of 250,000 simulated epidemics.

Age group	All epidemics		Discordant epidemics with the leading RR estimate belonging to age group	
	Fraction of susceptible individuals (p_i)	Susceptibility per susceptible (q_i)	Fraction of susceptible individuals (p_i)	Susceptibility per susceptible (q_i)
0-2 years	0.991(0.96,1)	0.958(0.91,1)	-----	-----
3-4 years	0.961(0.88,1)	0.886(0.82,0.96)	0.949 (0.87,1)	0.897(0.83,0.97)
5-6 years	0.881(0.7,1)	0.773(0.63,0.94)	0.737(0.64,0.83)	0.874(0.77,0.97)
7-12 years	0.804(0.55,0.99)	0.639(0.48,0.9)	0.593(0.49,0.70)	0.885(0.77,0.98)
13-19 years	0.779(0.51,0.99)	0.598(0.43,0.88)	0.539(0.44,0.64)	0.883(0.76,0.98)
20-39 years	0.779(0.51,0.99)	0.598(0.43,0.88)	0.512(0.45,0.57)	0.935(0.87,0.99)
40-59 years	0.753(0.46,0.99)	0.556(0.38,0.87)	-----	-----
60+ years	0.725(0.41,0.99)	0.513(0.33,0.85)	-----	-----

Table S3: Fraction of individuals susceptible to infection in each age group, and susceptibility to infection among the susceptible individuals in that given age group for all the 250,000 simulated epidemics, as well for those simulated epidemics with the highest RR estimate belonging to that age group for which vaccination of 1,000,000 individuals in that age group did not have the largest impact on reducing the epidemic's initial reproductive number/growth rate.

Section S2: RR estimates during the 2001-02 through the 2010-11 RSV seasons

In this section we present the estimates/confidence bounds for the RR statistic in the 10 age groups that we have used in our analyses for the 2001-02 through the 2010-11 RSV seasons in the US (with the estimates plotted in Figure 1 in the main body of the text). In addition to the estimates produced using the inference method described in the main body of the text, we also exhibit the estimates of the RR statistic obtained using the 3-week moving average for the hospitalization counts in each state to find the peak week for the hospitalization counts in that state - the values of the RR statistic are then estimated from eq. 1 in the main body of the text as described in the Methods.

Tables S4 and S5 present the results of the estimation according to the inference method in the main body of the text, while Tables S6 and S7 present the RR estimates with the 3-week moving average used to find the peak weeks. While there are some differences in the estimates of RR statistic between Tables S4 and S5 vs. S6 and S7,

the age groups with the leading RR estimate coincide for the two methods (3-week moving average vs. actual hospitalization counts for the peak week) for all the seasons in the data save for the 2001-02 epidemic. Also, using the actual peak week (rather than a 3-week moving average peak) generally resulted in a somewhat more clear separation (in terms of the RR values) of the age groups with the leading RR estimate.

Season/Age	<1	1y	2y	3-4y	5-6y
2001-02	0.964 (0.95,0.97)	1.172 (1.12,1.23)	1.181 (1.08,1.28)	1.146 (1.01,1.29)	1.284 (0.99,1.64)
2002-03	0.94 (0.93,0.95)	1.211 (1.15,1.27)	1.33 (1.21,1.46)	1.312 (1.16,1.48)	1.41 (1.06,1.85)
2003-04	0.93 (0.92,0.94)	1.206 (1.15,1.26)	1.272 (1.17,1.39)	1.396 (1.24,1.57)	2.103 (1.6,2.73)
2004-05	0.945 (0.93,0.96)	1.121 (1.07,1.18)	1.369 (1.25,1.5)	1.408 (1.25,1.58)	1.363 (1.04,1.76)
2005-06	0.935 (0.92,0.95)	1.194 (1.14,1.25)	1.476 (1.35,1.62)	1.534 (1.36,1.73)	1.386 (1.08,1.76)
2006-07	0.949 (0.94,0.96)	1.164 (1.11,1.22)	1.29 (1.18,1.41)	1.473 (1.31,1.65)	1.398 (1.09,1.77)
2007-08	0.963 (0.95,0.97)	1.184 (1.13,1.24)	1.296 (1.19,1.41)	1.217 (1.08,1.36)	1.176 (0.93,1.47)
2008-09	0.989 (0.98,1)	1.113 (1.06,1.17)	1.104 (1.01,1.2)	1.19 (1.07,1.32)	0.978 (0.78,1.21)
2009-10	0.905 (0.89,0.92)	1.271 (1.21,1.33)	1.259 (1.17,1.35)	1.504 (1.37,1.65)	1.721 (1.42,2.08)
2010-11	0.973 (0.96,0.99)	1.164 (1.11,1.22)	1.215 (1.13,1.31)	1.35 (1.23,1.48)	1.391 (1.16,1.65)
2011-12	0.942 (0.93,0.96)	1.178 (1.12,1.24)	1.297 (1.19,1.41)	1.425 (1.29,1.57)	1.22 (1,1.48)

Table S4: Seasonal RR estimates (with the leading seasonal R estimate in bold) for select age groups (children under the age of 7 years) obtained using the inference method in the main body of the text for RSV epidemics in the US between 2001-2012. Data from the State Inpatient Databases of the Healthcare Cost and Utilization Project (HCUP).

Season/Age	7-10y	11-17y	18-49y	50-64y	65+y
2001-02	0.727 (0.51,1.01)	0.801 (0.53,1.15)	0.667 (0.47,0.91)	0.888 (0.59,1.28)	0.415 (0.3,0.55)
2002-03	0.983 (0.68,1.36)	1.212 (0.78,1.81)	1.005 (0.68,1.43)	0.952 (0.63,1.38)	0.747 (0.52,1.03)
2003-04	1.641 (1.12,2.38)	1.426 (0.93,2.13)	0.815 (0.57,1.13)	1.017 (0.7,1.44)	1.063 (0.81,1.38)
2004-05	1.317 (0.91,1.85)	0.94 (0.6,1.39)	0.845 (0.58,1.19)	0.813 (0.55,1.14)	0.636 (0.44,0.89)
2005-06	1.245 (0.87,1.73)	0.939 (0.61,1.38)	0.72 (0.53,0.95)	0.664 (0.5,0.86)	0.647 (0.51,0.81)
2006-07	1.386 (0.98,1.9)	0.758 (0.49,1.11)	0.711 (0.52,0.94)	0.497 (0.35,0.68)	0.499 (0.37,0.65)
2007-08	0.972 (0.69,1.32)	0.808 (0.56,1.12)	0.499 (0.38,0.64)	0.548 (0.41,0.71)	0.385 (0.31,0.47)
2008-09	1.047 (0.78,1.39)	0.592 (0.42,0.81)	0.52 (0.4,0.66)	0.553 (0.42,0.71)	0.499 (0.41,0.59)
2009-10	1.635 (1.26,2.1)	1.021 (0.74,1.36)	0.821 (0.66,1)	0.727 (0.6,0.86)	0.602 (0.51,0.7)
2010-11	0.982 (0.77,1.23)	0.718 (0.54,0.94)	0.541 (0.44,0.65)	0.592 (0.49,0.7)	0.417 (0.37,0.47)
2011-12	1.367 (1.02,1.8)	0.847 (0.6,1.17)	0.826 (0.67,1)	0.577 (0.47,0.69)	0.658 (0.57,0.75)

Table S5: Seasonal RR estimates for select age groups (individuals over the age of 7 years) obtained using the inference method in the main body of the text for RSV epidemics in the US between 2001-2012. Data from the State Inpatient Databases of the Healthcare Cost and Utilization Project (HCUP).

Season/Age	<1	1y	2y	3-4y	5-6y
2001-02	0.965 (0.95,0.98)	1.161 (1.11,1.22)	1.182 (1.09,1.29)	1.167 (1.04,1.31)	1.165 (0.9,1.48)
2002-03	0.939 (0.93,0.95)	1.211 (1.15,1.27)	1.31 (1.19,1.44)	1.347 (1.19,1.52)	1.476 (1.11,1.93)
2003-04	0.925 (0.91,0.94)	1.22 (1.16,1.28)	1.301 (1.19,1.42)	1.459 (1.3,1.64)	2.09 (1.57,2.73)
2004-05	0.946 (0.93,0.96)	1.108 (1.05,1.16)	1.367 (1.24,1.5)	1.463 (1.29,1.65)	1.196 (0.91,1.54)
2005-06	0.938 (0.93,0.95)	1.173 (1.12,1.23)	1.412 (1.29,1.54)	1.521 (1.35,1.71)	1.447 (1.12,1.84)
2006-07	0.949 (0.94,0.96)	1.164 (1.11,1.22)	1.272 (1.16,1.39)	1.495 (1.32,1.68)	1.441 (1.12,1.83)
2007-08	0.962 (0.95,0.97)	1.195 (1.14,1.25)	1.298 (1.19,1.41)	1.258 (1.12,1.41)	1.109 (0.87,1.39)
2008-09	0.985 (0.97,1)	1.139 (1.08,1.2)	1.127 (1.04,1.23)	1.187 (1.06,1.32)	0.955 (0.76,1.19)
2009-10	0.908 (0.9,0.92)	1.288 (1.23,1.35)	1.249 (1.16,1.34)	1.444 (1.32,1.58)	1.703 (1.39,2.06)
2010-11	0.969 (0.96,0.98)	1.177 (1.12,1.23)	1.235 (1.15,1.33)	1.374 (1.25,1.5)	1.392 (1.17,1.66)
2011-12	0.943 (0.93,0.96)	1.171 (1.12,1.23)	1.287 (1.18,1.4)	1.382 (1.25,1.53)	1.331 (1.08,1.62)

Table S6: Seasonal RR estimates (with a 3-week moving average for hospitalization counts used to find the peak week in each state) for select age groups (children under the age of 7 years, with the leading seasonal R estimate in bold) for RSV epidemics between 2001-2012. Data from the State Inpatient Databases of the Healthcare Cost and Utilization Project (HCUP).

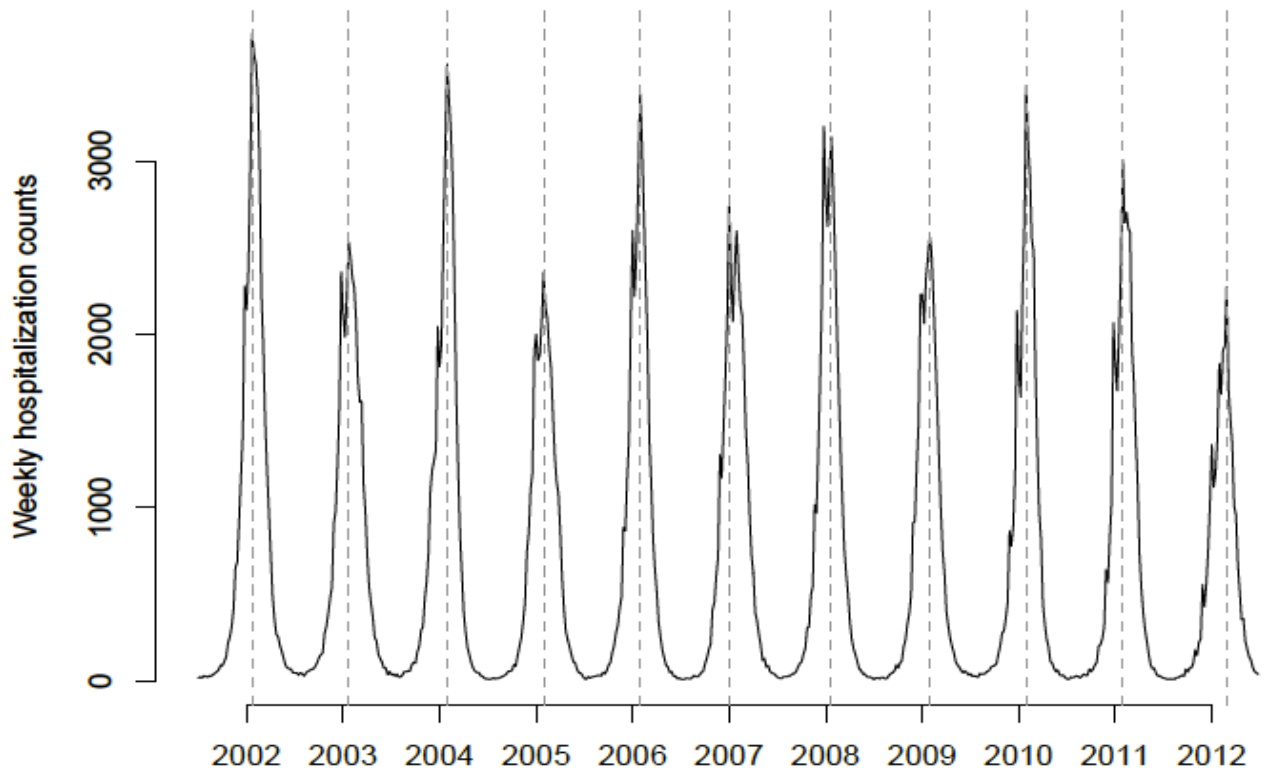
Season/Age	7-10y	11-17y	18-49y	50-64y	65+y
2001-02	0.727 (0.51,1.01)	0.801 (0.53,1.15)	0.667 (0.47,0.91)	0.888 (0.59,1.28)	0.415 (0.3,0.55)
2002-03	0.983 (0.68,1.36)	1.212 (0.78,1.81)	1.005 (0.68,1.43)	0.952 (0.63,1.38)	0.747 (0.52,1.03)
2003-04	1.641 (1.12,2.38)	1.426 (0.93,2.13)	0.815 (0.57,1.13)	1.017 (0.7,1.44)	1.063 (0.81,1.38)
2004-05	1.317 (0.91,1.85)	0.94 (0.6,1.39)	0.845 (0.58,1.19)	0.813 (0.55,1.14)	0.636 (0.44,0.89)
2005-06	1.245 (0.87,1.73)	0.939 (0.61,1.38)	0.72 (0.53,0.95)	0.664 (0.5,0.86)	0.647 (0.51,0.81)
2006-07	1.386 (0.98,1.9)	0.758 (0.49,1.11)	0.711 (0.52,0.94)	0.497 (0.35,0.68)	0.499 (0.37,0.65)
2007-08	0.972 (0.69,1.32)	0.808 (0.56,1.12)	0.499 (0.38,0.64)	0.548 (0.41,0.71)	0.385 (0.31,0.47)
2008-09	1.047 (0.78,1.39)	0.592 (0.42,0.81)	0.52 (0.4,0.66)	0.553 (0.42,0.71)	0.499 (0.41,0.59)
2009-10	1.635 (1.26,2.1)	1.021 (0.74,1.36)	0.821 (0.66,1)	0.727 (0.6,0.86)	0.602 (0.51,0.7)
2010-11	0.982 (0.77,1.23)	0.718 (0.54,0.94)	0.541 (0.44,0.65)	0.592 (0.49,0.7)	0.417 (0.37,0.47)
2011-12	1.367 (1.02,1.8)	0.847 (0.6,1.17)	0.826 (0.67,1)	0.577 (0.47,0.69)	0.658 (0.57,0.75)

Table S7: Seasonal RR estimates (with a 3-week moving average for hospitalization counts used to find the peak week in each state) for select age groups (individuals over the age of 7 years) for RSV epidemics between 2001-2012. Data from the State Inpatient Databases of the Healthcare Cost and Utilization Project (HCUP).

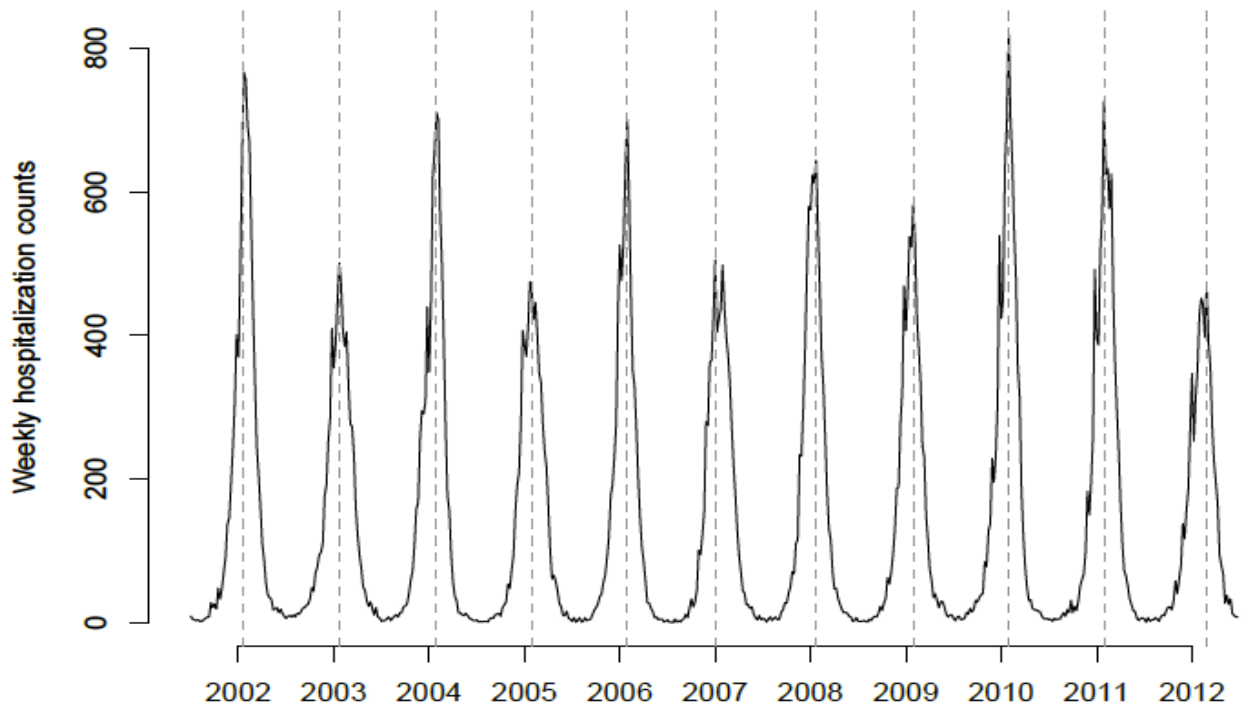
Section S3: Plots of the weekly RSV hospitalization counts

Figure S1 exhibits the weekly hospitalization counts with RSV in the diagnosis (both primary and contributing, [ICD-9] codes 079.6, 466.11, 480.1) in different age groups for the states included in our analyses (see the Methods section in the main body of the text). It suggests a temporal upward trend in RSV hospitalizations, particularly in the older age groups. Additionally, it allows for the examination of the seasonal counts of hospitalization in each age group relative to the seasonal peak weeks for hospitalization counts in all age groups (plotted in dashed grey lines). Furthermore, Figure S1 suggests an apparent biannual cycle in the height of the peak for RSV hospitalization counts in several age groups, including infants.

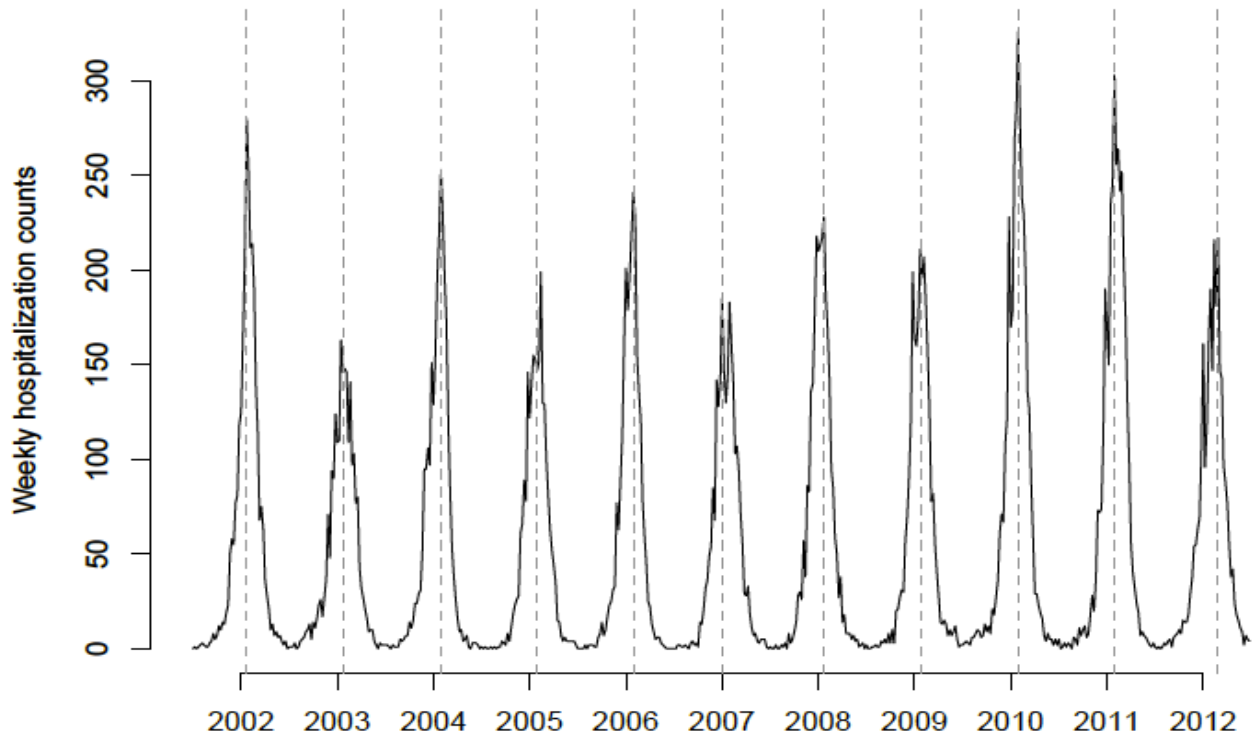
Age <1 year



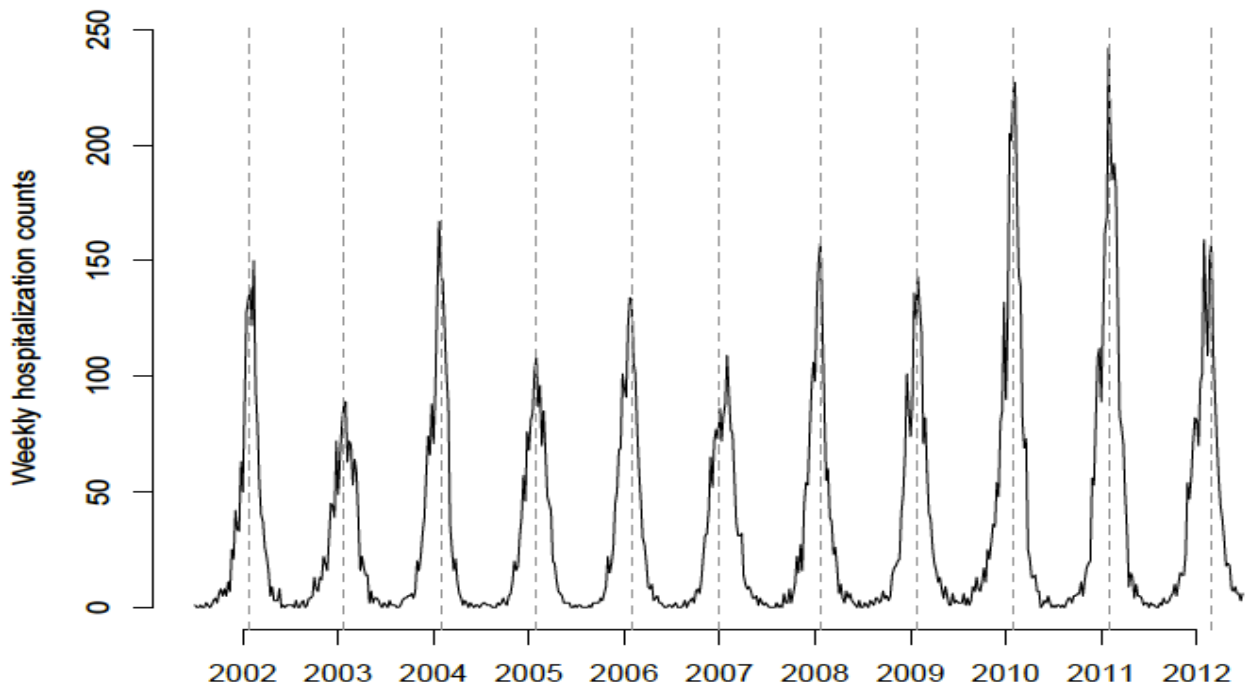
Age 1 year



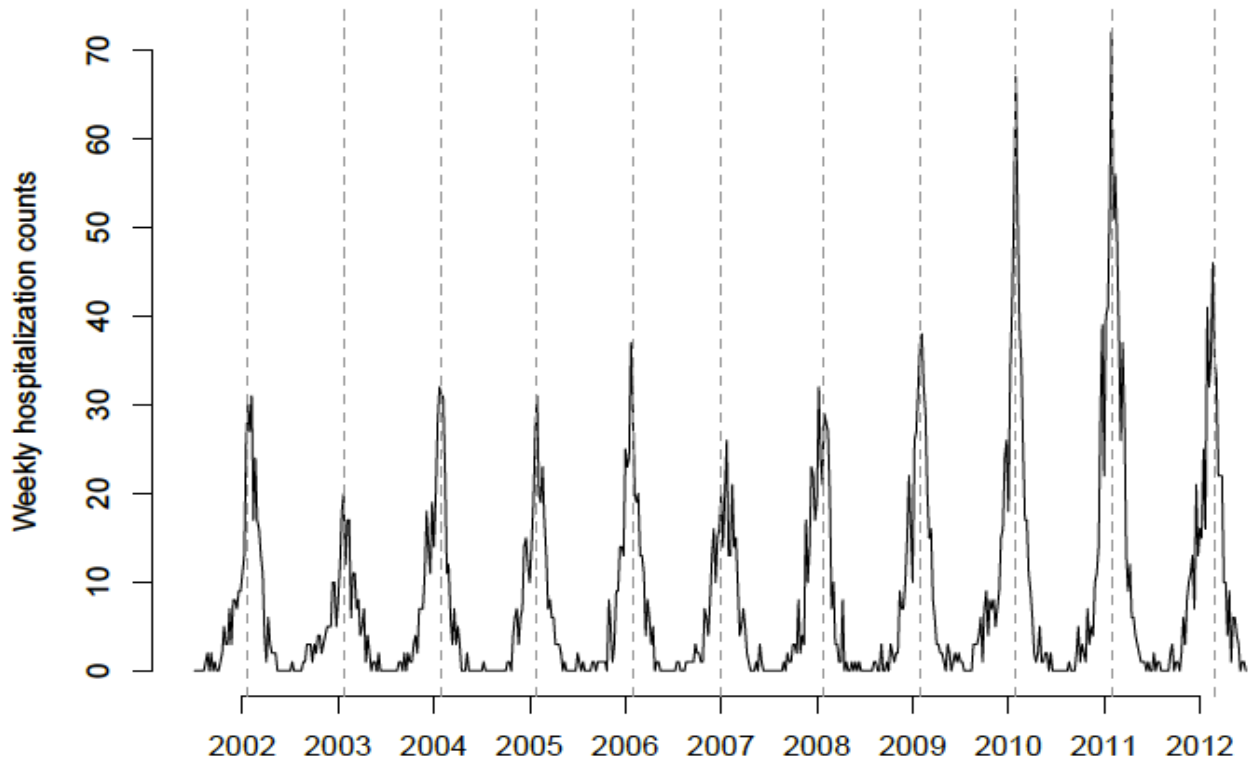
Age 2 years



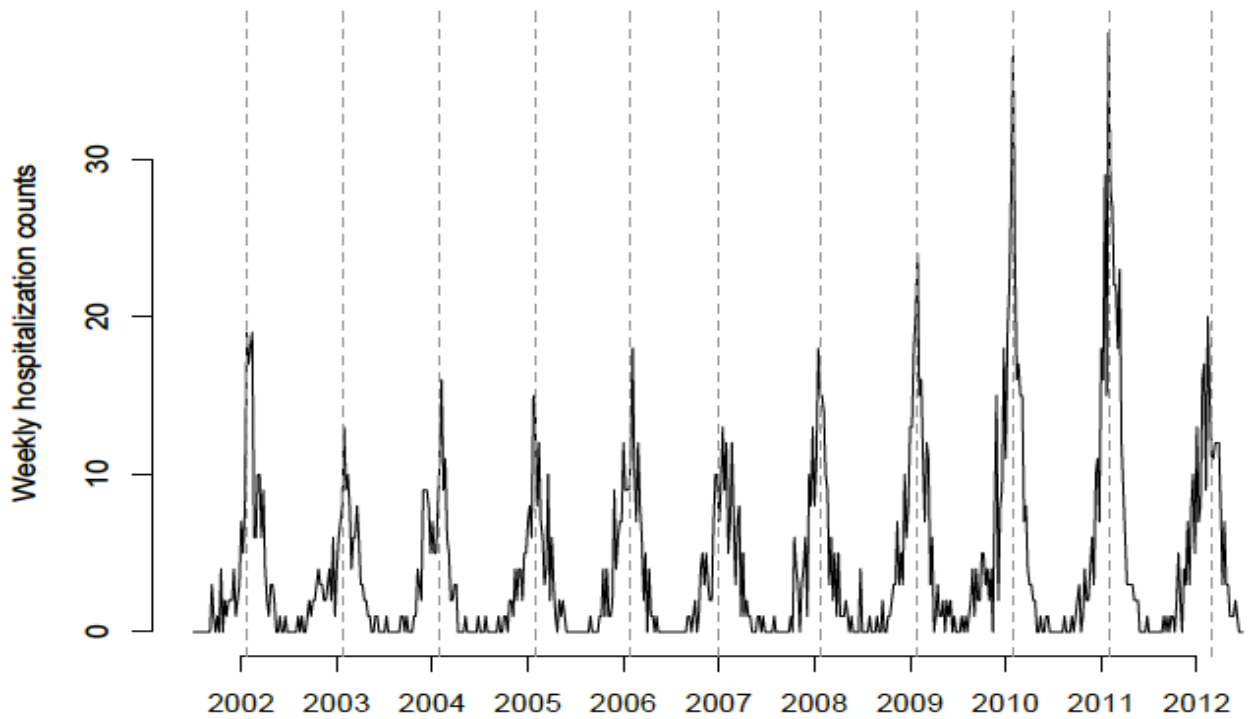
Age 3-4 years



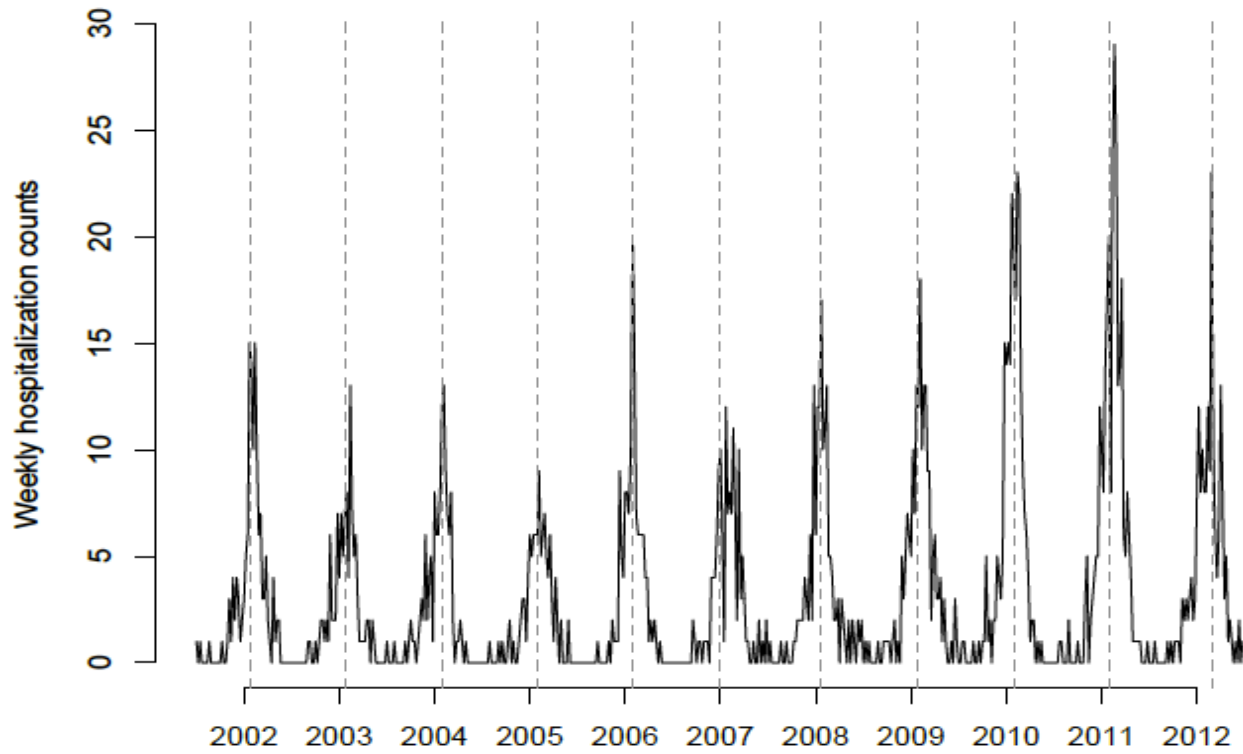
Age 5-6 years



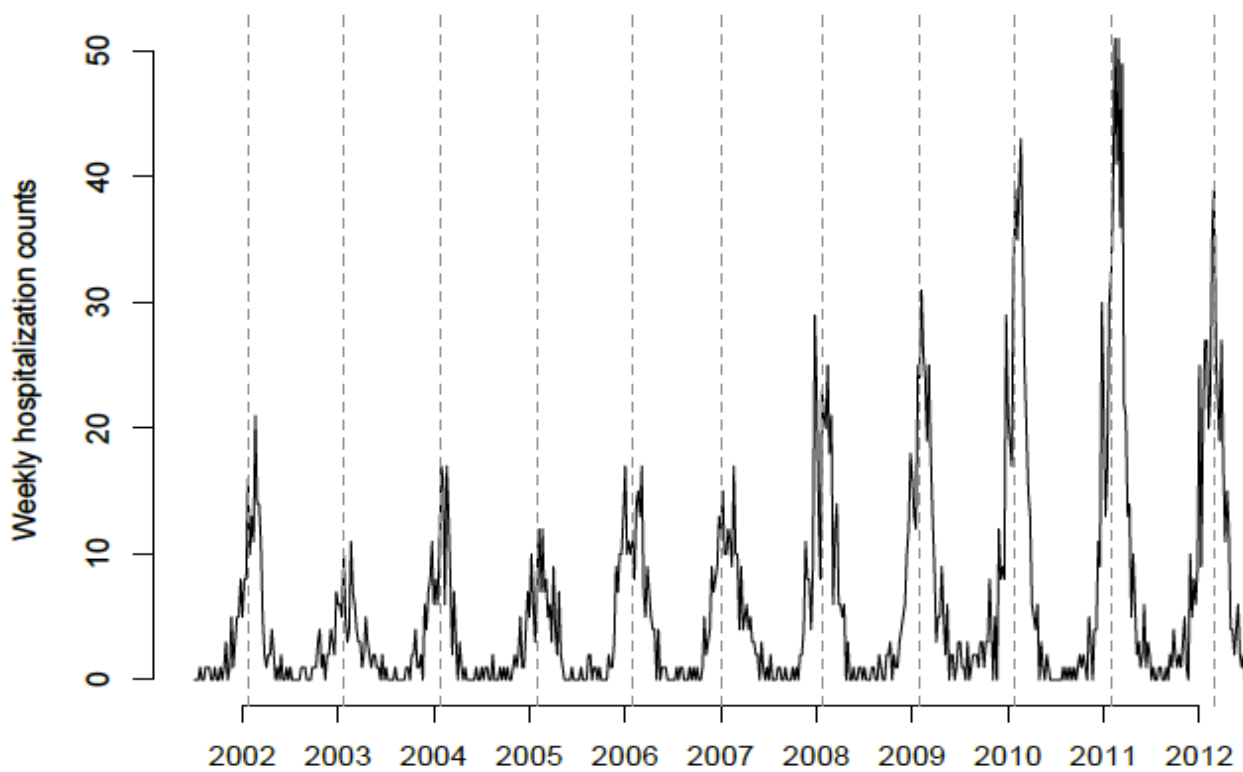
Age 7-10 years



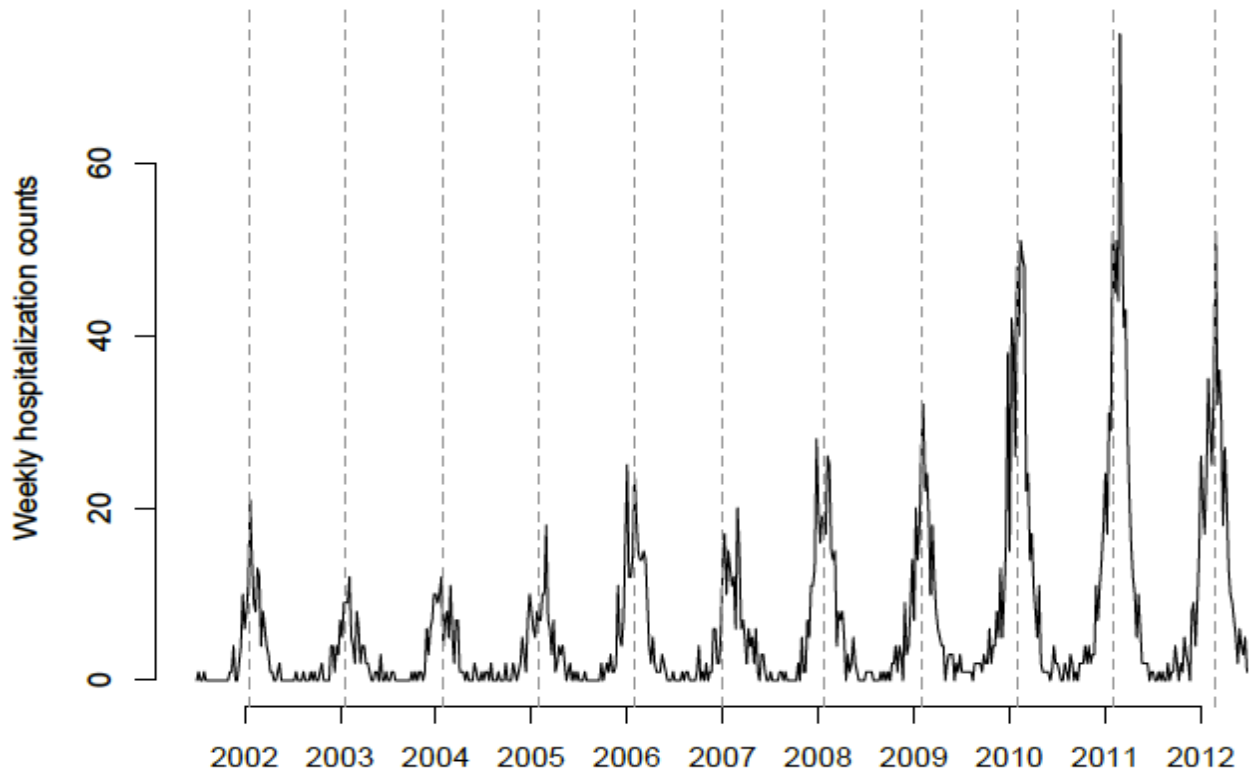
Age 11-17 years



Age 18-49 years



Age 50-64 years



Age 65+ years

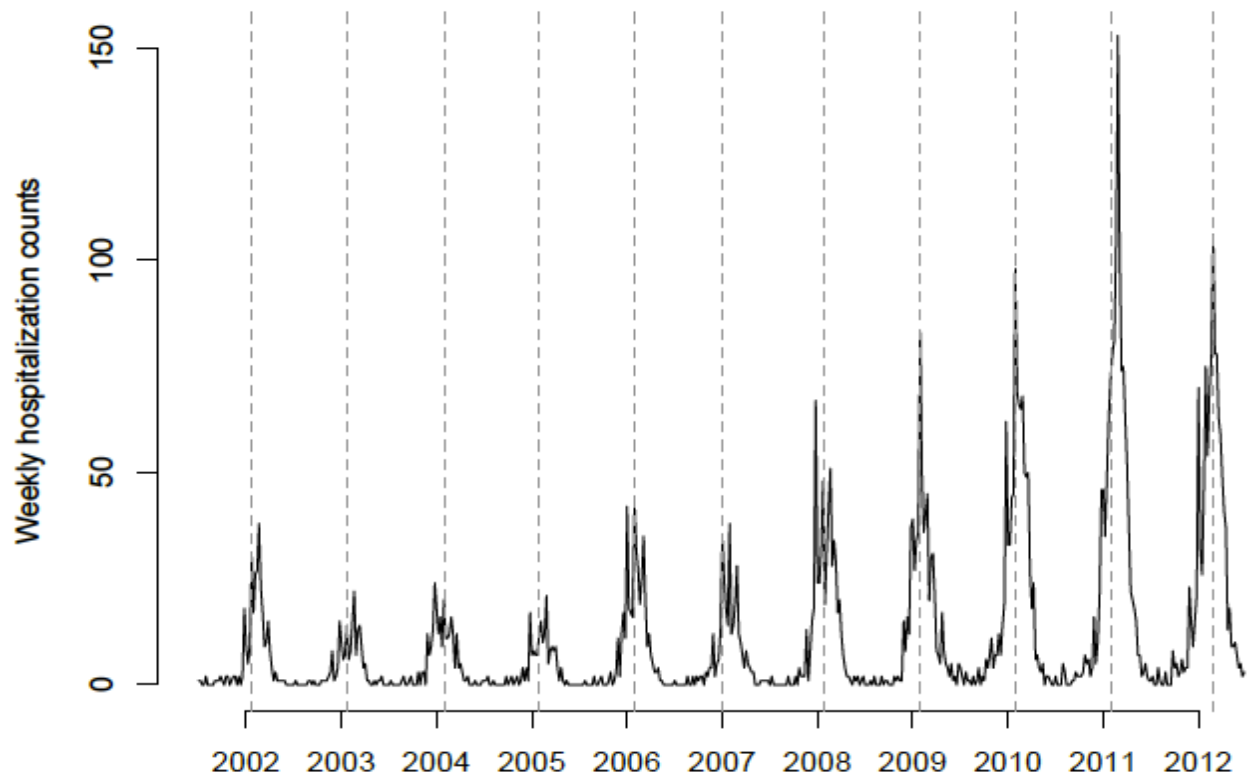


Figure S1: Weekly hospitalization counts with RSV in the diagnosis (both primary and contributing, [ICD-9] codes 079.6, 466.11, 480.1) in different age groups. Seasonal peak weeks for hospitalization counts in all age groups represented by grey lines.

References:

1. Worby CJ, Chaves SS, Wallinga J, Lipsitch M, Finelli L, Wallinga J. On the relative role of different age groups in influenza epidemics. *Epidemics* 2015;13:10-16.
2. J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases *PLoS Med.*, 5 (2008), p. e74
3. 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(10):936-44.
4. US Centers for Disease Control and Prevention. Bridged-Race Population Estimates, 1990-2016 data request [Cited 2017 September 1]. Available from: <https://wonder.cdc.gov/Bridged-Race-v2016.HTML>
5. Okiro EA, White LJ, Ngama M, Cane PA, Medley GF, Nokes DJ. Duration of shedding of respiratory syncytial virus in a community study of Kenyan children. *BMC Infect Dis.* 2010 Jan 22;10:15
6. El Saleeby CM, Bush AJ, Harrison LM, Aitken JA, Devincenzo JP. Respiratory Syncytial Virus Load, Viral Dynamics, and Disease Severity in Previously Healthy Naturally Infected Children. *J Infect Dis.* 2011;204(7):996-1002
7. Crowcroft NS, Zambon M, Harrison TG, Mok Q, Heath P, Miller E. Respiratory syncytial virus infection in infants admitted to paediatric intensive care units in London, and in their families. *Eur J Pediatr.* 2008;167(4):395-9.