METHODOLOGY AND SENSITIVITY ANALYSIS OF ESTIMATES OF R

CHARACTERIZATION OF THE FIRST 2 PANDEMIC WAVES IN COPENHAGEN

Summer wave. The weekly numbers of medically treated ILIs started increasing dramatically in Copenhagen during the week of June 23, denoting the onset of the summer pandemic wave. Morbidity increased at an exponential rate during the first 3–4 weeks of the epidemic, indicating that depletion of susceptible individuals did not play a role during this period, and so a true R could be measured for those first 3–4 weeks. The growth rate was constant during the first 3 weeks, with an increase by a factor of $\gamma = 10.4$/week. After these first 3 weeks, the growth rate was significantly reduced, probably because of a combination of the depletion of susceptible individuals and behavioral changes [31].

Fall wave. The fall wave may best be characterized as starting on September 8, because that week had the lowest incidence of disease during that period. However, the growth rate was not constant during any 3-week period, and , consequently, our best estimate for the growth rate is the maximal weekly growth factor of $\gamma = 3.48$, observed for the week starting September 29 and ending the following week.

METHOD FOR ESTIMATING R BASED ON ILI DATA

We employed a 2-step process to estimate R. During the first step, we recognize that the data are given on a weekly basis (i.e., at discrete points in time), and we employ a standard statistical analysis (see, e.g., reference [37]) of these data, to determine whether the epidemic, in fact, increases exponentially. During the second step, we translate the discrete data by obtaining a measure of R, by using a discrete version of the method proposed by Wallinga and Lipsitch [20].

Utilizing these data, which, in fact, are discrete counts, we observe that, because it is rare that an individual will seek medical attention for this disease, we expect that the number of cases reported in a given week will be Poisson distributed. We are assuming that all reported cases of ILI are attributable to the Spanish flu, because this allows us to assume that the observations are Poisson-distributed, which, in turn, allows us to estimate their variance. Subtracting a fixed background level would increase our estimate of the growth factor (for details, see the sensitivity analysis). Provided that all ILI cases are attributable to the Spanish flu, our basic assumption about the structure of the data is that y_i, the number of cases in week i, is Poisson distributed with intensity λ_i.

During the initial phase of the epidemic, when exponential growth is expected, the weekly number of cases will grow geometrically, as $\alpha\gamma^t$, where $\alpha\gamma$ is the initial size of the infected population during the first week, γ is the weekly growth factor, and t is the number of weeks since the reference week. Eventually the exponential growth phase will discontinue as a result of the depletion of susceptible individuals. The hypothesis for exponential growth, H_0, then becomes

$$H_0: \lambda_t = \alpha\gamma^t, \quad \text{for } t = 1 \ldots T.$$

To look for the onset of the epidemic, we varied the starting date, and to look for when the depletion of susceptible individuals affected the growth rate, we varied the length of the period T. The exponential growth phase of the epidemic was surprisingly short—in most cities, only 3 weeks—suggesting that depletion of susceptible hosts sets in at this time. Taking the summer wave in Copenhagen as an example, only ~1000 cases had been observed when the growth rate started to decline. This early decline can hardly be explained solely as a reduction in susceptible individuals, because such an explanation would require an unrealistically high level of underreporting and subclinical seroconversion. If those 1000 cases were to correspond to an observable reduction (say, a 5% reduction) in susceptible individuals, there should have been 27 seroconverted hosts for every reported case. This number may be compared with the estimate by Nguyen-van-Tam and Hampson [38], who observed that ~50% of the infections in a pandemic are subclinical. The rapid deviation from exponential growth seems to be characteristic of epidemic data for pandemic influenza and may be due to a mixture of behavioral, biological, and climatic factors [31, 33].

The cities of Stockholm and Oslo give rise to additional remarks. The case reports for Stockholm are surprisingly few. During the period from June 30 through September 7, a total of 1316 cases are reported in Stockholm, with a case-fatality rate of ~111 deaths per 1000 reported cases of respiratory illness. The case-fatality rates in the other Scandinavian cities were an order of magnitude lower (table 3), suggesting that the Stockholm reporting system was incomplete. For that reason, in table 3, we have not included morbidity data or case-fatality rates for Stockholm. In contrast, a total of 18,613 cases were reported during the summer wave in Oslo (June 30–August 11), with 143 deaths. For the city of Oslo, our data start with 2165 cases reported in the week beginning June 30, when the summer wave in Oslo was well on its way. Low [8] reports of an unspecified source who claimed that the Oslo epidemic started in the middle of June (which concurs with [27]). Such a starting date would be similar to that found in Copenhagen, although the incidence of infection in Oslo during the summer wave was considerably higher than that in Copenhagen. Thus, we speculate that the first case occurred 3–4 weeks earlier in Oslo and loosely guess that the weekly growth factor is $\gamma = \sqrt[4]{2165} - \sqrt[3]{2165}$, corresponding to R = 2.4–3.2. The case rates in Copenhagen and Oslo were 45/1000 and 71/1000, respectively. Also, the case-fatality rate was higher

in Oslo, 56/100,000, compared with 10/100,000 in Copenhagen, suggesting that the difference in recorded incidence between the 2 cities reflects a true difference in disease incidence rather than a variation in reporting practices.

RESULTS OF ESTIMATES OF R BASED ON ILI DATA

Table 5 shows the results of testing the hypothesis of exponential growth H_0. For each wave, the longest period in which the epidemic shows exponential growth is listed in the fourth column. In the eighth column, $-2 \log Q$ gives the χ^2-distributed log-likelihood test quantity, and the ninth column gives the 95% acceptance level for the test. Considering the fairly large observation size, we accept the hypothesis at a slightly lower acceptance level for the Copenhagen summer wave. The parameter γ gives the estimated weekly growth factor, and R gives the corresponding value of reproduction number, as discussed in the next section.

The error on the estimate of γ is obtained by observing that, if D^2l denotes the Hessian matrix of the logarithmic likelihood function evaluated at the estimators, then the variance of the estimators are given by the inverse of D^2l. We take the standard error on γ to be the square root of the variance.

SENSITIVITY ANALYSIS

Our analysis of the ILI and hospitalization data assumes that all observed ILI cases are attributable to the Spanish influenza. However, because the ILI diagnosis includes a background of noninfluenza-related cases, it would be natural to subtract the baseline obtained from a Serfling-type analysis. To see the effect of neglecting the baseline, we repeated our analysis of our best data set, namely, the ILI cases in Copenhagen from June 23 and onward, after removing 12 cases per week (the Serfling June baseline level) from the total incidence. Removing 12 cases per week results in acceptance of the hypothesis and an estimate of $\gamma = 16 \pm 2$, corresponding to an estimate of R = 2.7–3.2. Thus, in this sense, our estimate is conservative, giving a lower bound on the growth rate γ. Because we have assumed that the observations are Poisson-distributed, the maximum likelihood estimator of γ weighs the contributions of the observations according to their size. Once again, taking the Copenhagen summer wave of ILI cases as an example, we find that changing the initial observation of 14 cases per week to 19 cases per week reduces the estimate of γ from 10 ± 1 to 9.5 ± 0.8. We conclude that our approach of using "raw" morbidity data to estimate R produces conservative estimates and that the maximum likelihood method that we used reduced the sensitivity of our results to an error attributable to our choice of not removing noninfluenza (baseline) events from the morbidity series.

TRANSLATION OF WEEKLY GROWTH FACTOR INTO R

To convert the estimated γ into R, information about the course of the typical infection is required. Recent analysis of viral-shedding data on influenza infections in drift years suggests that the mean generation time of an infection (the serial interval) may be as short as 2.6 days [3, 4, 20], whereas previous studies have used an approach in which the duration of infection was assumed to be 6 days, with a serial interval around 4 days [1, 6, 21]. To our knowledge, during the 1918 influenza pandemic in Copenhagen, the course of infection has never been quantified, and so it remains unclear how close the serial interval, estimated during a normal drift period, matches the actual serial interval (see [33] for further discussion). Because earlier estimates of R for the 1918 transmission, based on epidemiological data, use the long serial interval [1, 10, 39], we give estimates based on both the long and short serial intervals, for comparison.

In general, we can describe the course of infection by the relative infectivities on day i given by ν_i, with $\sum_i \nu_i = 1$. At the onset of the epidemic, where we can neglect the depletion of susceptible individuals, the growth process is now described by a simple Leslie-type projection matrix. Thus, if $J(t)$ denotes the (eight-dimensional) vector of hosts who were infected 1–8 days ago, then the number of hosts that are infected the next day is determined by $R\sum_{i=1}^{8} \nu_i \mathrm{J}_i(\mathrm{t})$, and the vector $J(t+1)$ is given by

$$J(t+1) = \begin{pmatrix} R\nu_1 & R\nu_2 & \cdots & R\nu_7 & R\nu_8 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} J(t).$$

Neglecting possible transients, we determined the daily growth factor by the dominant eigenvalue of the matrix, and γ by the dominant eigenvalue raised to the power of 7.

Figure 3*A* illustrates the relationship between R and γ, based on data from Ferguson et al. [4] and from Mill et al. [1] representing a short and a long serial interval, respectively. Including possible transients would lower the growth and would require a larger value of R to explain the same growth factor; hence, our estimate of R gives a lower bound.

ESTIMATES OF R BASED ON MORTALITY DATA

Table A1 illustrates the deaths attributed to influenza and pneumonia during the fall wave of 1918, as reported from the available sources. The weekly growth factor was determined by 2 of the methods proposed by [1] for mortality data. We determined the highest growth factor that was observed between 2 consecu-

Table A1. Number of deaths attributed to Spanish influenza during the 1918 fall waves and the corresponding estimates of the growth rate (γ) and the reproduction number (R).

	Copenhagen			Oslo	Stockholm	Gothenborg
	All causes of death [14]	Respiratory deaths [14]	Reported deaths from influenza or pneumonia [8]	Reported deaths from influenza or pneumonia [8]	Reported deaths from influenza or pneumonia [8]	Reported deaths from influenza or pneumonia [8]
Population, no.	543,000	534,000	534,000	260,000	413,000	197,000
Deaths/week, no.						
1 September	0	0	2	0	5	5
8 September	0	2	1	9[a]	7	2
15 September	0	4	5	10	16	8
22 September	0	8[a]	6[a]	23	41	30
29 September	10[a]	20	11	16	65	48
6 October	28	40	32	24	157	104
13 October	93	102	22	33	196	186
20 October	252	260	36	92	243	117
27 October	344	350	45	133	170	76
Estimated γ						
Max 1 week	3.3	2.6	2.9	2.8	2.6	3.8
First 2 weeks	9.3	5.0	5.3	2.6	NA	NA
Corresponding estimate of R (short-serial interval)						
Max 1 week	1.6	1.4	1.5	1.5	1.4	1.6
First 2 weeks	1.5	1.4	1.4	1.2	NA	NA
Corresponding estimate of R (long-serial interval)						
Max 1 week	1.8	1.6	1.7	1.7	1.6	1.9
First 2 weeks	1.7	1.5	1.5	1.3	NA	NA

NOTE. Low [8] did not include deaths attributed to bronchitis in Copenhagen. Also, the respiratory deaths and the reported deaths from influenza or pneumonia, in Copehagen, are determined by the use of a Serfling seasonal regression model, using weekly data. γ is determined in 2 ways, as proposed by [1]: the first row gives the maximal growth factor from 1 week to the next (max 1 week), and the last row gives the observed γ over the first 3 weeks after death counts exceeded the threshold level γ^2. NA, growth was not observed for 3 consecutive weeks or death counts never fell below threshold level between the summer and fall waves.

[a] This value indicates the first week in which deaths exceeded the threshold level of 1/100,000.

tive weeks (table A1; see the row labeled "max 1 week"). Following Mills et al. [1], we also identified the first week in which (excess) deaths exceeded 1/100,000 and then determined by how much the death count grew during the next 2 weeks. If the weekly death count did not fall sufficiently to allow for the identification of a fall threshold, or if death count did not grow monotonically for the first 3 weeks after the threshold, then we did not apply this method to the data. Finally, we transformed the weekly growth factor to R by using the same methods as in the analysis of the ILI data. For the fall wave, our estimates based on mortality data are consistent with those based on morbidity data. We applied the same method for deaths occurring during the summer wave. Our estimates of R based on mortality data are consistently well below the estimates based on morbidity data (results not shown). However, one must keep in mind that the actual numbers of deaths during the summer months were small. In addition, the period from infection to death was longer than that from infection to onset of symptoms, so that the transient period before exponential growth is obtained may be long enough to clutter the analysis.

Table 5. Statistical analysis of the weekly growth factor (γ), based on ILI data from 4 Scandinavian cities, during the summer and fall waves of 1918.

City	Data	Start	Length, weekly	γ	R_{short} (2.6 days)	R_{long} (4 days)	$-2 \log Q$	χ^2, 95% confidence	N[a]
1918 summer wave									
Copenhagen	Hospitalization	23 June	3	35 ± 17	2.8–4.0	3.6–5.4	3.68[b]	3.84	136
	ILI	23 June	3	10 ± 1	2.2–2.4	2.8–3.0	6.21[c]	3.84	928
Gothenburg	ILI	30 June	3	31 ± 7	3.1–3.6	4.0–4.8	0.23[b]	3.84	592
Oslo	ILI	14 June?	?	7–12	2.0–2.5	2.4–3.2	. . .[d]	. . .[d]	. . .[d]
1918 fall wave									
Copenhagen	Hospitalization	1 September	4	1.9 ± 0.1	1.2–1.3	1.34–1.41	3.68[b]	5.99	253
	ILI	8 September	3	1.7 ± 0.1	1.22–1.24	1.29–1.33	6.21[c]	3.84	1784
Gothenburg	ILI	8 September	3	2.6 ± 0.1	1.40–1.44	1.55–1.60	0.23[b]	3.84	1826
Oslo	ILI	25 August	4	2.2 ± 0.1	1.3–1.4	1.4–1.5	0.1[b]	5.99	569

NOTE. The reproduction number (R) is estimated for the summer and fall waves in 4 Scandinavian cites and based on each data type (influenza-like illness [ILI] and hospitalizations) and 2 serial-interval parameter values—short duration (2.6 days) [4] and long duration (~4 days) [1]. Ranges represent 95% confidence intervals, on point estimates.

[a] The observation size indicates the total number (N) of cases reported during the period.
[b] This value indicates that the hypothesis is accepted at the 95% level.
[c] This value indicates that the hypotheis is accepted at the 99% level.
[d] See text for method.