Influenza epidemics in temperate regions show a characteristic seasonal pattern with peak incidence occurring in winter. Previous research has shown that low absolute humidity and school holidays can both affect influenza transmission. During an epidemic, transmission is strongly influenced by the depletion of susceptibles (i.e., increase in the number of those immune). To assess how much variability in influenza transmission intensity is due to each of these driving factors, we used a long time series of the number of weekly visits to general practitioners for influenzalike illness in the Netherlands from 1970–2011 and transformed this into a time series of weekly influenza reproduction numbers, which are a measure of transmission intensity. We used statistical regression techniques to quantify how the reproduction numbers were affected by each driving factor. We found a clear ranking of importance of driving factors in explaining the variation in transmission intensity. Most of the variation (30%) was explained by the depletion of susceptibles during the season, 27% was explained by between-season effects, and 3% was explained by absolute humidity. School holidays at the Christmas period did not have a statistically significant effect on influenza transmission. Although the influence of absolute humidity was small, its seasonal fluctuations may determine when sustained influenza transmission is possible and may thus drive influenza seasonality.

epidemic modeling; influenza; seasonality; transmission dynamics

Despite decades of research, the mechanisms that drive influenza seasonality in temperate regions remain poorly understood. Seasonal introduction of novel influenza strains and the depletion of susceptibles (i.e., increase in the number of those immune) to these novel strains are important factors in shaping annual influenza epidemics (1, 2). Small seasonal changes in the transmissibility of influenza over the year could result in sustained annual cycles (3). Several factors have been proposed that may cause such small seasonal changes in transmissibility, such as changes in human behavior (4), school holidays (5), host susceptibility (6, 7), and absolute humidity. There is experimental evidence that low temperature and low humidity increase the transmissibility and survivability of influenza A viruses (8–10), and it has been observed that periods of low absolute humidity often precede periods with increased influenza-related mortality (11, 12). However, analysis of data from the 2009 influenza pandemic revealed that only a small part of the variation in influenza transmission could be explained by absolute humidity (13).

In this study, we analyzed time series of weekly incidence of general practitioner consultations by patients with influenzalike illness (ILI) in the Netherlands and quantified the effect of extrinsic factors (absolute humidity and school holidays) and intrinsic factors, such as the depletion of susceptibles. Existing approaches to analyzing time series of infectious diseases rely on fitting a discrete-time transmission model to case reports and then looking for seasonal trends in transmission rates (2, 14–16). In our method, we first inferred the effective reproduction number, defined as the number of secondary infections produced by a typical infective individual, and decomposed it into biologically relevant variables. We then translated this decomposition into a straightforward regression...
In these calculations, we take the generation interval and then calculated the effective reproduction number (equation 1) according to a method described by Wallinga and Lipsitch (21):

\[
R_u = \sum_{t=u}^{\infty} \frac{b(t)}{\sum_{a=0}^{\infty} b(t-a)g(a)},
\]

in which the influenza incidence at time \( t \) is denoted with \( b(t) \). In these calculations, we take the generation interval \( a \) to follow a \( \gamma \) distribution \( g(a) \) with a mean of 2.6 days and a variance of 1.9 (22–24), and we account for interval censoring (22). We calculated the weekly reproduction numbers as the geometric mean of the daily reproduction numbers.

**MATERIALS AND METHODS**

**Time series of ILI**

We used data on the weekly incidence of general practitioner consultations by patients with ILI in the Netherlands collected and reported by the Continuous Morbidity Registration Sentinel General Practice Network (hereafter, the sentinel network) of the Netherlands Institute for Health Services Research (Figure 1, Web Appendix 1.1–1.2, and Web Figures 1 and 2 available at http://aje.oxfordjournals.org/) (17) to infer the weekly effective reproduction number. The sentinel network covers about 0.8% of the Dutch population. In the Netherlands, all patients seek medical care through their general practitioners; therefore, the consultations within this sentinel network are representative of the healthcare-seeking behavior of the Dutch population. We used data from all 41 available influenza seasons from 1970 to 2011. Since 1996, throat swabs from selected patients with ILI in the sentinel network have been taken and tested for influenza virus (17). During the period with increased ILI incidence, there is a strong positive linear correlation between the incidence of general practitioner visits for ILI and laboratory-confirmed influenza (18, 19), which suggests that the incidence of ILI is approximately proportional to the incidence of influenza infection.

For each season, we defined the epidemic period as the 11 weeks centered on the week with peak ILI incidence during the season (Figure 1; Web Figures 1 and 2). These 11 weeks correspond to the average duration of an epidemic in the Netherlands as described by Donker and Gravestein (20). To test whether the results depend on the precise definition of the epidemic period, we repeated the analyses with a shorter epidemic period of 9 weeks, a longer period of 13 weeks, a test whether the results depend on the precise definition of the epidemic period, we repeated the analyses with a shorter epidemic period of 9 weeks, a longer period of 13 weeks, a period of 11 weeks with the seventh week centered on the week of peak ILI incidence, and a period of 6 weeks leading up to and including the week with the peak ILI incidence.

**Reproduction numbers**

Intensity of influenza transmission is measured by the effective reproduction number, which is defined as the number of secondary infections caused by a typical single infective individual. The effective reproduction number can be obtained from the time series of incidence of general practitioner consultations by patients with ILI. We transformed the weekly incidence to daily incidence (Web Appendix 1.3, Web Figure 3) and then calculated the effective reproduction number \( R_u \) at day \( u \) (equation 1) according to a method described by Wallinga and Lipsitch (21):

\[
R_u = R_0 S_{ij} \prod_k d_{ijk} \beta_k.
\]

The fraction of susceptibles at the start of week \( i \) of season \( j \) is \( S_{ij} \approx S_{ij0} e^{-h_{ij}} \), where \( S_{ij0} \) represents the initial fraction of susceptibles at the start of season \( j \), \( h_{ij} \) is the observed cumulative incidence of general practitioner consultations by patients with ILI up to week \( i-1 \) of season \( j \), and \( z_j \) is a seasonal effect that adjusts \( h_{ij} \) to the (unknown) number of influenza infections and, thus, determines how quickly susceptibles deplete. By taking the logarithms of equation 1 and adding a normally distributed error term \( \varepsilon_{ij} \), we arrive at a linear regression model (Web Appendix 2):

\[
\log(R_{ij}) = \log(R_0 S_{ij0}) + z_j h_{ij} + \sum_k \beta_k \log(d_{ijk}) + \varepsilon_{ij}.
\]

The effect of driving variable \( k \), \( d_{ijk} \) is estimated through the regression coefficient \( \beta_k \). If the coefficient \( \beta_k \) is positive, then there is a positive association between \( R_{ij} \) and \( d_{ijk} \); if the coefficient \( \beta_k \) is negative, then there is a negative association between \( R_{ij} \) and \( d_{ijk} \). The compound parameter for the basic reproduction number and fractions of susceptibles at the start of each season, \( \log(R_0 S_{ij0}) \) and the parameter that scales ILI incidence to actual influenza infection, \( z_j \), are treated as “nuisance variables.” We will refer to the seasonal differences in these parameters as the “between-season effect.”

**Driving variables**

We retrieved data on daily maximum temperature and maximum relative humidity over the period 1970–2011 from the Royal Netherlands Meteorological Institute’s weather station in De Bilt, the Netherlands (Figure 1; Web Figures 1 and 2). This station is located centrally in the Netherlands, such that all general practitioner sentinels are within a 200-km radius of the station. Regional differences in humidity and temperature in the Netherlands are small. The daily maximum absolute humidity was calculated from the maximum relative humidity and maximum temperature. We calculated the weekly absolute humidity by taking the arithmetic mean of the daily values from Monday through Sunday. We normalized these values by dividing by the overall mean of maximum absolute humidity values (12.7 g/m³) and included this in our analysis as a continuous explanatory variable. The functional relationship between the reproduction number of influenza and absolute humidity and maximum temperature can be expressed as a basic reproduction number \( R_0 \) modified by the fraction of susceptibles at the start of that week \( S_{ij0} \) and modified by \( k \) external driving variables \( d_{ijk} \) scaled by \( \beta_k \):

\[
R_{ij} = R_0 S_{ij} \prod_k d_{ijk} \beta_k.
\]
humidity can be rewritten as the relationship used by Shaman et al. (11).

The school holidays around Christmas and New Year’s Day (hereafter, Christmas holidays) in the Netherlands last for 2 weeks. We included these Christmas holidays in our analysis as a binary explanatory variable. Over the entire period, 42 of 451 weeks were classified as Christmas holidays. The timing of the Christmas holidays during the epidemic periods from 1970 to 2011 is shown in Figure 1. The timing of other school holidays is hard to pinpoint, because the recommended timing of those holidays is staggered, varying by year and by region, and each school is free to deviate from the recommended timing.

**Test on simulated data**

We tested whether the statistical analysis can assess how much variability in influenza transmission is due to driving factors such as absolute humidity and school holidays. To this end, we simulated a time series of weekly incidence over the same number of seasons as the actual time series (Web Appendix 3.1). We use a susceptible-exposed-infected-recovered transmission model, in which we know exactly how the transmission rate depends on absolute humidity and on school holidays. We simulated different data sets with different parameter values for the effects of these driving factors. We applied the proposed statistical analysis to the simulated data sets and tested whether we could identify differences in parameter values between data sets, and we tested whether the proportion of variance explained by a driving factor was an appropriate measure of the actual strength of the explanatory variable in driving the transmission intensity. The tests of the statistical analysis on simulated data showed that the analysis can detect substantial modulation of the transmission intensity by a driving variable and can attribute such modulation to the correct driving variable (Web Appendix 3.2, Web

![Figure 1. Time series of incidence of influenzalike illness in the Netherlands over the period 1970–2011 (bottom line), absolute humidity of de Bilt, the Netherlands (top line), and school holidays at Christmas (black bars) during an 11-week epidemic period centered on the peak incidence of influenza. (Normalized absolute humidity represents absolute humidity values divided by the overall mean absolute humidity of 12.7 g/m^3.)](image-url)
Tables 1–4). A sudden decrease in transmission is smoothed out and partly shifted to earlier dates (Web Appendix 3.3, Web Figure 4).

Permutation test

In a temperate region such as the Netherlands, absolute humidity tends to be low in the winter. Influenza epidemics also typically occur in the winter, and the reproduction number in this period is therefore high. The fact that the reproduction number and the absolute humidity correlate does not necessarily mean there is a causal relationship. We used a permutation procedure to test whether the association between absolute humidity and reproduction number might be due to confounding by season. In this procedure, we selected the absolute humidity data for each of the 11-week epidemic periods and randomized these absolute humidity data across seasons, repeating the analysis with the randomized explanatory variables. The means that, for example, we correlated the reproduction number in the 1986–1987 season to the absolute humidity in the 1972–1973 season. We conducted 10,000 permutations, and sampling was done with replacement. For each permutation, we then calculated how well the absolute humidity correlated with the time series of influenza reproduction numbers. If the correlation between absolute humidity and influenza reproduction number in the actual time series was significantly stronger than the correlations we observed in the randomized samples, we concluded that confounding by season was an unlikely explanation for this correlation.

RESULTS

The incidence of general practitioner consultations by patients with ILI showed marked peaks for each of the 41 seasons, and the peak incidence tended to decrease over time (Figure 1). The estimated reproduction numbers for each of the epidemic periods were in the range of 0.8–1.5 with a median of 1; typically, the reproduction number is approximately 1.2 at the start of an epidemic period and declines to approximately 0.8 at the end of an epidemic period (Figure 2).

Most (60%) of the observed variance in the influenza reproduction numbers was explained by the depletion of susceptibles during an influenza epidemic, by the between-season effect, and by absolute humidity (Table 1). As expected, a considerable part (30%) of this explained variance in reproduction numbers was due to the depletion of susceptibles during an influenza epidemic. The between-season effects explained 27% of the variance. The time series of absolute humidity explained 3% of the variance.

Changes in absolute humidity showed a strong negative correlation with changes in the reproduction number of influenza (Table 2; *P* < 0.001). In the permutation test, only 4 of 10,000 random samples (*P* = 4 × 10^-04) resulted in stronger correlations than the actual data. This shows that the strong correlation between the influenza reproduction number and absolute humidity is unlikely to be due to confounding by season. The estimated value of the regression coefficient that mediates this effect is −0.134 (Table 2). This means that if the absolute humidity is halved (e.g., from 0.8 to 0.4 in normalized values), the reproduction number increases by a factor 0.5^{−0.134} = 1.1. In most seasons, the model fit was improved considerably by the inclusion of absolute humidity, but there were also seasons in which there was almost no improvement and others in which the fit worsened (Figure 2). There appeared to be no relationship between the effect of humidity and the timing of the peak incidence (*P* = 0.75), the average humidity during the epidemic period (*P* = 0.21), or the magnitude of the peak incidence (*P* = 0.65).

We did not find a significant correlation between the Christmas holidays and the reproduction number of influenza. The weak evidence for an effect of Christmas holidays might be due in part to a low number of weeks (42 of a potential 82 weeks) in which the Christmas holidays occurred during the 11-week influenza epidemic period and suggests that, if an effect of Christmas holidays on transmission in the community exists, it is small.

In our analysis, we included data from an 11-week epidemic period centered on the peak incidence of general practitioner visits by patients with ILI. Shortening or lengthening the duration of the epidemic period selected for analysis to 9 or 13 weeks or moving the epidemic period 2 weeks forward had no effect on our outcomes; selectively including only the 6 weeks up to and including the week with peak incidence resulted in a stronger effect of absolute humidity and a weaker effect of depletion of susceptibles (Web Appendix 4.1, Web Table 5). Alternative driving variables and different time lags are presented in Web Appendix 4.2 and Web Figure 5.

The residuals of the model that included the depletion of susceptibles, absolute humidity, and between-season effects reveal some remaining negative autocorrelation at time lags of 2, 3, and 4 weeks (Web Appendix 5.1, Web Figure 6). The model thus captures most, but not all, of the seasonal variation in influenza reproduction numbers in the data. Closer inspection revealed that the reproduction numbers were underestimated in the weeks prior to the epidemic peak and overestimated in week 46 and in week 51, which is the week prior to Christmas (Web Appendix 5.1, Web Figure 7). Inspection of the “nuisance variables” that make up the between-season effect reveals that the seasonal variation in susceptibility is in line with expectations, capturing, for example, an increased susceptibility to influenza infection during the 2009-2010 season when the pandemic influenza A (H1N1) virus circulated (Web Appendix 5.2, Web Figure 8). Furthermore, the variables suggest that the general practitioner consulting rate of infected individuals decreased over seasons, and that the incidence of influenza infection remained constant over seasons.

The typical seasonal trajectories of absolute humidity through the season show a high occurring in the summer and a low occurring in the winter. By using the best fitting model, we used these trajectories to project the seasonal change in reproduction numbers, extrapolating beyond the range of values used in the fitting procedure. This revealed a small seasonal change in the reproduction number from values below the threshold value of 1 in the summer and crossing the threshold value where sustained influenza transmission becomes possible in October to a peak in January (Figure 3). Even though the explained variation in influenza reproduction number by absolute humidity is as low as 3%, it is a systematic seasonal
effect that allows us to predict how influenza transmission is driven throughout a season in temperate climates.

DISCUSSION

The model that included the depletion of susceptibles, between-season effects, and absolute humidity could explain 60% of the variance in influenza reproduction numbers. A large part of the remaining 40% of unexplained variance results from many factors, including noise that is inherent to the stochastic nature of disease transmission; measurement noise that is due to the use of data on general practitioner visits for ILI, which have an imperfect correlation to the actual incidence of influenza infections; and sampling error, because the sentinel stations cover less than 1% of the Dutch population. The absolute percentages of explained variance are

Figure 2. Influenza reproduction numbers inferred from the influenzalike illness (ILI) time series (black dots) and predicted reproduction numbers based on driving factors, including depletion of susceptibles and a between-season effect, excluding absolute humidity (black line) and including absolute humidity (grey line). The influenza reproduction number is defined as the number of secondary infections caused by a typical single infective individual. The difference between the black line and the grey line illustrates the improvement in fit due to inclusion of maximum absolute humidity.
affected by many external factors and are likely to be specific to this study. The relative ordering of the driving factors, however, is likely to be robust. Influenza transmission is strongly governed by the depletion of susceptibles and modulated by absolute humidity. If there is an effect of Christmas holidays on transmission, it is too weak to be detected.

By focusing on the reproduction number, which is a weighted ratio of incidence values, we must assume a stationary distribution of the generation interval for all transmission events, but we avoid having to specify how many individuals with an influenza infection develop ILI that is severe enough to cause them to visit their general practitioners. By restricting ourselves to an 11-week epidemic period with increased in-fluenza-related mortality (11). During the influenza A(H1N1) pandemic in 2009, only a small proportion of the variability of transmission intensity across 12 countries in Europe could be explained by absolute humidity (13). The statistical evidence for the relationship reported here is stronger than previously reported, and the permutation test demonstrated that confounding by season is negligible. Outside temperature is strongly correlated with absolute humidity, but it is less likely to affect transmission intensity because, during the winter, people are likely to spend much time indoors where the temperature is higher than that outside but the absolute humidity is similar to that outside (9, 11).

Between-season effects explained 27% of the variance in the reproduction number. The parameters that make up these effects have an epidemiologic interpretation, which allows

Table 1. Variance Explained by the Potential Driving Factors of Influenza Transmission in the Netherlands, 1970–2011

<table>
<thead>
<tr>
<th>Driving Factor</th>
<th>Regression Terms $M^a$</th>
<th>$R^2_b$</th>
<th>$\Delta R^2_b$</th>
<th>df</th>
<th>$R^2_{adj}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depletion of susceptibles</td>
<td>$zh_j$</td>
<td>0.30</td>
<td>0.30</td>
<td>449</td>
<td>0.30</td>
</tr>
<tr>
<td>Between-season effect</td>
<td>log($R_0S_0$) + $zh_j$</td>
<td>0.57</td>
<td>0.27</td>
<td>369</td>
<td>0.47</td>
</tr>
<tr>
<td>School holidays</td>
<td>log($R_0S_0$) + $zh_j$</td>
<td>0.57</td>
<td>0</td>
<td>368</td>
<td>0.47</td>
</tr>
<tr>
<td>Absolute humidity</td>
<td>log($R_0S_0$) + $zh_j$</td>
<td>0.60</td>
<td>0.03</td>
<td>368</td>
<td>0.50</td>
</tr>
</tbody>
</table>

$^a$ All fitted regression models take the form log($R_i$) = $M + \varepsilon$, and the regression terms $M$ differ for each model.

$^b$ $R^2$ is the variance of the influenza reproduction numbers that is explained by each model.

$^c$ $\Delta R^2$ is the proportion of the variance explained by a specific driving factor.

$^d$ $R^2_{adj}$ provides a measure of parsimony for each model.

Table 2. Estimates of the Strength of Driving Factors on Influenza Transmission in the Netherlands, 1970–2011

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Regression Terms $M^a$</th>
<th>Variable</th>
<th>Estimate</th>
<th>95% CI</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depletion of susceptibles</td>
<td>$zh_j$</td>
<td>$z$</td>
<td>-7.1</td>
<td>-6.1 to -8.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Between-season intercept</td>
<td>log($S_0$) + $zh_j$</td>
<td>Median $[S_0]_b$</td>
<td>1.12</td>
<td>1.02, 1.24</td>
<td></td>
</tr>
<tr>
<td>Between-season depletion of susceptibles</td>
<td>log($S_0$) + $zh_j$</td>
<td>Median $[z]_b$</td>
<td>-15.0</td>
<td>-23.9, -6.0</td>
<td></td>
</tr>
<tr>
<td>School holidays</td>
<td>log($S_0$) + $zh_j$ + $\beta_1$log($d_1$)</td>
<td>$\beta_1$</td>
<td>0.01</td>
<td>-0.02, 0.07</td>
<td>0.78</td>
</tr>
<tr>
<td>Absolute humidity</td>
<td>log($S_0$) + $zh_j$ + $\beta_2$log($d_2$)</td>
<td>$\beta_2$</td>
<td>-0.13</td>
<td>-0.19, -0.08</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.

$^a$ All fitted regression models take the form log($R_i$) = $M + \varepsilon$, and the regression terms $M$ differ for each model.

$^b$ For regression coefficients that were specific for each season, we estimated the range of $R_0S_0$ to be 1.00–1.25 and the range of $z_j$ to be -52.3 to -1.4.
us to check whether the outcomes of the statistical analyses make epidemiologic sense. Here we find that a large part of the between-season effect is captured by the decreasing general practitioner consulting rate by patients with ILI over the seasons. Only a small part is explained by variation in susceptibility between seasons.

The percentage of variance of the influenza reproduction number that is explained by absolute humidity is relatively small (3%), but the regular seasonal fluctuations of absolute humidity combined with the low value of the influenza reproduction number (hovering around unity) help to explain why influenza epidemics in the summer are unlikely (when the influenza reproduction number is lower than 1), and why the likelihood of influenza epidemics increases during the autumn and winter (when the influenza reproduction number is above 1). In this sense, influenza epidemics are locked in phase with the annual cycle of absolute humidity. Capturing this subtle but important effect of absolute humidity is an essential step toward predicting when sustained influenza transmission becomes possible and when an epidemic can occur.

We did not find convincing statistical evidence for reduced influenza transmission during the Christmas holidays in our analysis. This leaves the possibility that a small reduction occurred but was not detected, and we cannot rule out that the slight dip in transmission in the week prior to Christmas (Web Figure 7) is related to the Christmas holidays. Other studies have reported an effect of school holidays on transmission among children. A study of ILI data collected in a French sentinel network suggested that school holidays reduce influenza transmission to children by 20%–29% (5); and a study of confirmed influenza infection in the Canadian province of Alberta suggested that school closure reduced transmission among children by more than 50% (26). A study by Cowling et al. (27) did not detect a substantial effect of school closure on community transmission of influenza in Hong Kong. The apparent discrepancy between these findings is explained by our focus on the effect of Christmas holidays, rather than school holidays in general, and by our focus on transmission in the community, rather than transmission among children. During Christmas holidays, it is possible that a reduction in transmission due to school closure is offset by an increased number of family gatherings during Christmas.

The ranking of driving factors for influenza infection leads to a variety of questions about how environmental factors influence infectious disease dynamics. We address 4 of them here. First, are there other environmental drivers of influenza transmission? The statistical approach can be extended to include more potential driving factors, such as demographic data, and to include a time lag, as is common in the analysis of multivariate time series. Second, are other
infectious diseases with recurrent epidemics subject to the same seasonal forcing as influenza? It is now possible to use this statistical approach to compare the effect of drivers on reproduction numbers of other infectious such as the human respiratory syncytial virus, rhinovirus, norovirus, and rotavirus. Third, do the same driving factors for influenza arise in other regions that may have different, subtropical or tropical climates? This statistical approach can be extended to include time series of general practitioner visits by patients with ILI in other regions with different influenza monitoring systems and with different lengths. Fourth, what is the effect of interventions on the transmission of influenza? In some exceptional cases, schools are closed as a precautionary measure against influenza transmission (27), or vaccines become available during an influenza epidemic, as in the 2009 influenza pandemic. For such seasons, we can detect a statistically significant effect on transmission by categorizing the weeks before, during, and after interventions. When measuring the effect of interventions, it is important to measure those effects against background knowledge of seasonal drivers of epidemics. We believe that we can learn more about these seasonal drivers as more time series data on influenza become available.

ACKNOWLEDGMENTS

Author affiliations: Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, the Netherlands (Dennis E. te Beest, Michiel van Boven, Jacco Wallinga); Netherlands Institute of Health Services Research, Utrecht, the Netherlands (Mariëtte Hooi-veld); and Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, the Netherlands (Carline van den Dool).

This work was supported by the National Institute for Public Health and the Environment project S/210096.

We thank Jeffery Shaman, Marc Lipsitch, and Adam Meijer for their helpful comments on the manuscript. We thank the Royal Dutch Meteorological Institute for providing the weather data and the Netherlands Institute of Health Services Research for providing data on the general practitioner consultations by patients with influenzalike illness.

Conflict of interest: none declared.

REFERENCES


