Seasonal variation in SARS-CoV-2 transmission in The Netherlands, 2020-2022: statistical evidence for an inverse association with solar radiation and temperature

Klinkenberg, D.; Backer, J.; Reusken, C.; Wallinga, J.

Seasonal variation in SARS-CoV-2 transmission in the Netherlands, 2020-2022: statistical evidence for a negative association with temperature.

Don Klinkenberg, Jantien Backer, Chantal Reusken, Jacco Wallinga

Acknowledgments
Author affiliations: Center for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands (Don Klinkenberg, Jantien Backer, Chantal Reusken, Jacco Wallinga); Department of Viroscience, Erasmus MC, Rotterdam, the Netherlands (Chantal Reusken); Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, the Netherlands (Jacco Wallinga).
Data availability: all data and code are available online at github.com/rivm-syso/covid-seasonality
We thank all collaborators in the Data Analytics, Research, and Automated Reporting (DARA) team in the Center for Infectious Disease Control for processing and curating all COVID-19 surveillance data.
Conflicts of Interest: none.

Keywords
coronavirus; seasonality; time series regression; reproduction number; respiratory infections

ABSTRACT
In temperate regions, respiratory viruses such as SARS-CoV-2 are better transmitted in Winter than in Summer. Understanding how temperature and humidity affect SARS-CoV-2 transmissibility can enhance projections of COVID-19 incidence and improve estimation of the effectiveness of control measures. During the pandemic, transmissibility was tracked by the reproduction number R_t. This study aims to determine whether information about the daily temperature and absolute humidity improves predictions of R_t in the Netherlands from 2020 to 2022, and to quantify the relationship between R_t, daily temperature and absolute humidity. We conducted a regression analysis, accounting for immunity from vaccination and previous infection, higher transmissibility of new variants, and changes in contact behaviour due to control measures. Results show a linear association between $\log R_t$ and daily temperature, indicating a ratio of R_t in Winter versus Summer of 1.5 (95% CI, 1.2-1.8). Including absolute humidity in the model did not improve predictions. The possibility that this association arises from unrelated seasonal patterns was dismissed, as weather data from earlier years provided poorer fits, and incorporating mobility data did not affect results. This suggests a causal relationship between temperature and SARS-CoV-2 transmissibility, enhancing confidence in using this relationship for short-term predictions and other epidemiological analyses.

INTRODUCTION

Many infectious diseases display seasonal patterns in incidence caused by seasonality in exposure (e.g. Lyme disease, Q-fever) or seasonality in contact rates (e.g. measles[1]), as well as by other mechanisms. With many pathogens, associations of transmission potential with weather variables have been found, mainly temperature and humidity[1-3]. Associations with weather variables are especially pronounced for respiratory viruses, where the causal mechanism may be related to the pathogen (e.g. virus survival[4]), the individual host (duration of shedding[4]) or host-host interaction (e.g. more contacts indoors[5]).

Already early in the COVID-19 pandemic a relation between weather variables and SARS-CoV-2 transmission has been hypothesized and tested, to assess the risk for a large second wave by the end of 2020 in northern temperate regions[6-8]. Many studies focused on the transmission potential, which can be expressed as the time-dependent reproduction number R_t, defined as the mean number of secondary cases infected per primary case. Large-scale studies across many geographical locations showed associations of R_t with temperature and humidity, with often small and non-linear effect sizes. A study among 409 cities in 26 countries worldwide found a maximum R_t at a temperature around 10˚C and at an absolute humidity around 6.6 g/m^3, with effect sizes of 0.087 and 0.061 compared to minimal values at 20˚C and 11 g/m^3, respectively[9]. In a multi-county analysis in mainland USA, almost 20% of variation in R_t was explained by temperature, humidity and ultraviolet radiation, but the effects were concentrated at the extreme ends of the ranges of the variables. For instance, R_t was largest at a temperature of 20˚C, by about 5% compared to any other temperate under 30˚C; and specific humidity affected transmissibility only at very low or very high values, suggesting that the model better explained spatial than seasonal variation in R_t[10]. In a study in 47 prefectures in Japan, the reproduction number R was found to decrease with about 6% when comparing temperatures of 0˚C vs 30˚C, and with 5% when absolute humidity increased to 10g/m^3 or higher[11]. In general, across many studies, temperature was found to be inversely associated with SARS-CoV-2 transmission, whereas the association with humidity was unclear[12].

The impact of weather on SARS-CoV-2 transmission has significant implications for epidemiological analyses, such as forecasting incidence by dynamic transmission models or estimating the effects of sets of control measures. To account for a weather effect in such analyses, it should be properly and independently quantified for the target population at hand. This requires the target population to be sufficiently large, so that the epidemic is driven by transmission events within the population, affected by the local weather, instead of reflecting epidemics in other locations by many import cases. It also requires the epidemic in the target population to have a sufficiently uniform climate, because the mechanisms by which weather may impact transmissibility are not self-evident, and may be different in different climates. The importance of focussing on single target populations is further illustrated by the fact that estimated effect sizes for smaller populations tend to be large , e.g. 46% higher R_t in winter compared to summer explained by temperature and absolute humidity in France[13], or 22% in London and Paris[14], whereas those from larger populations are smaller or reflect spatial rather than seasonal variation[10,11].

In this study, we estimated the association between SARS-CoV-2 transmissibility and temperature and absolute humidity during 25 months of the pandemic in the Netherlands. We quantified this relationship and estimated the relative change in reproduction number in Winter vs Summer. We assessed sensitivity of the relationship to potential mediation of the effect through mobility as measured by mobile phone locations, and we established whether the estimated association can be attributed to the general seasonal trend of temperature and absolute humidity or to the specific course of the weather during the pandemic.

METHODS

We used the log of the daily *instantaneous* reproduction number R_t, which was estimated during the pandemic by the National Institute of Public Health and the Environment (RIVM) in the Netherlands[15,16] (Figure 1a), and the variance of the estimator. It was estimated by dividing the symptom onset incidence at time $t + 5$ (incubation period is estimated to be 5 days) by the number of infectious individuals, i.e. the symptom onset incidence in the days up to $t + 5$ weighted by the serial interval distribution[17]. Because of autocorrelation and day-of-week effects in the time series of daily estimated R_t, we only used the time series of $\log R_t$ estimates of Fridays (Figure 1a, see Appendix S1).

Weekly genomic surveillance data were obtained from RIVM's open data[18]. They were used to estimate $\log R_t^{\text{var}}$, i.e. the relative mean transmissibility at time t, caused by the circulating variants Alpha, Beta, Gamma, and Delta, relative to the wildtype variant[19] (Figure 1b, see Appendix S2). After the emergence of Delta, $\log R_t^{\text{var}}$ was kept constant, because the emergence of Omicron in December 2021 was for an important part due to immune escape.

Cumulative incidence CumIncid(t) per day of the pandemic was calculated by RIVM, by combining serological data and daily hospitalisation data, and previously used to calculate the prevalence of infected individuals[20]. It was available until 2 July 2021, after which CumIncid(t) was kept constant. Vaccine coverage data were obtained from RIVM's open data[21]. We used the weekly national coverage of the completed primary vaccination, interpolated to obtain daily coverages. To calculate vaccine-induced immunity VaccImm(t) in the population, the coverage was multiplied by 75% vaccine effectiveness as estimated against the Delta variant[22], because the primary vaccination series was quickly followed by the rise of the Delta variant. Cumulative incidence and vaccine-induced immunity were combined to calculate the times series of the susceptible fraction of the population as $s_t = \left(1 - \text{CumIncid}(t)\right) \cdot \left(1 - \text{VaccImm}(t)\right)$ (Figure 1b).

The time series (109 weeks) of weekly reproduction numbers from 22 February 2020 to 29 March 2022 was partitioned into 41 intervention periods (see Appendix S3), separated by changes in COVID-19 control measures or the start or end of school holidays. After a preliminary analysis for outlier detection based on Cook's distance[23] (Appendix S1), the period between 26 June 2021 and 9 July 2021 was split into two separate weeks for the main analyses. This was a period when bars and nightclubs were allowed to open, followed by a rapid surge in new cases and closure of bars and nightclubs in the period thereafter.

We used the weekly averaged (Tuesday to Monday) temperature T_t (degrees Celsius) and absolute humidity ah_t (g/m^3) as weather variables (Figure 1c,d). We used observations from the "De Bilt" weather station, located centrally in the Netherlands[24], and calculated absolute humidity from relative humidity and temperature as $ah = 2.1674 \cdot 6.112 \cdot \exp\left(17.67 \cdot T / (T + 243.5)\right) \cdot rh / (T + 273.15)$ [25].

We used two mobility datasets. The Google mobility data (Figure 1e) were based on mobile phone locations, indicating activity levels at several locations relative to pre-pandemic levels[26]. We used weekly averaged mobility indices for the locations *work*, *grocery*, *home*, *transit*, *retail*, and *parks*. The NVP data (Nederlands Verplaatsingspanel, 'Netherlands Translocation Panel', Figure 1f) has measured mobility and activities of 10,000 participants in the Netherlands since 2019, with a smartphone app and occasional surveys[27]. Weekly presence at five different locations were monitored during the COVID-19 pandemic, relative to the first week of March 2020: *home*, *short_walk*, *supermarket*, *office*, and *healthcare*.

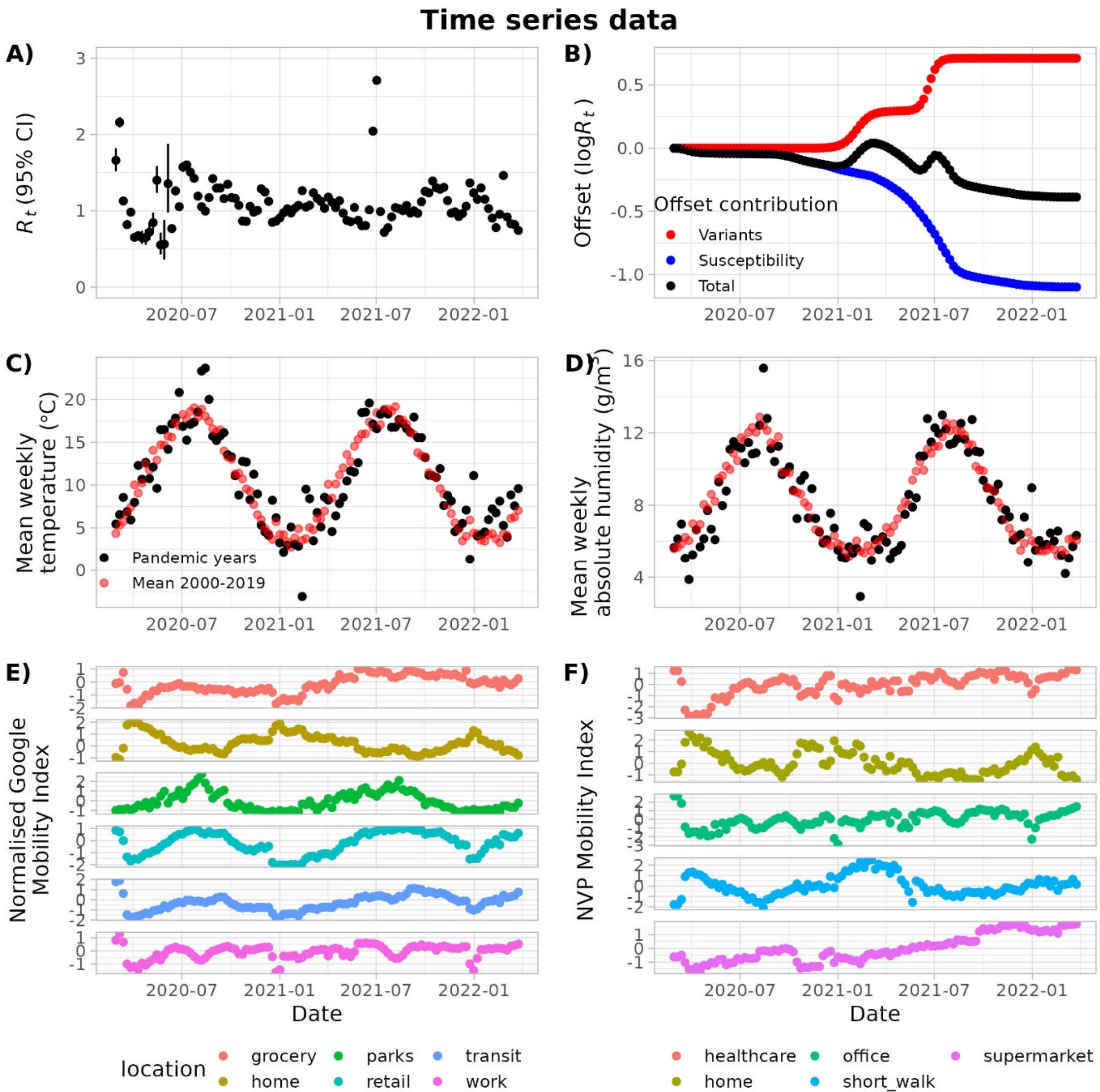

Figure 1. The time course of all datasets. (A) the response variable $\log R_t$, with the 95% confidence interval indicating the uncertainty of the estimator, the inverse of which is used to weigh the observations; (B) the offsets due to increased transmissibility from new variants, and decreased transmissibility from immunity due to infection and vaccination; (C) the mean weekly temperature during the pandemic and in the years 2000-2019; (D) the mean weekly absolute humidity during the pandemic and in the years 2000-2019; (E) the six variables of the Google Mobility Index; (F) the five variables of the NVP Mobility data.

We used the Oxford Stringency Index for the Netherlands as a compound measure for the stringency of control measures (Oxford COVID-19 Government Response Tracker[28]). It was used to indicate changes in control measures and compare these to changes in the estimated regression coefficients.

We conducted weighted time series regression analyses to estimate the association of the SARS-CoV-2 reproduction number R_t and weather variables between 22 February 2020 and 29 March 2022 in the Netherlands. The full regression equation for the main analysis was

$$\log R_t \sim offset\left(\log R_t^{\text{var}} + \log s_t\right) + period_t + s\left(T_t\right) + s\left(ah_t\right) + s\left(T_{t-7}\right) + s\left(ah_{t-7}\right), \quad (1)$$

in which the lags are 7 days because of the use of weekly data, and weekly averaged temperature and absolute humidity. The observations were weighted by $1/\mathrm{var}(\log R_t)$ to reflect uncertainty in the estimated $\log R_t$. Offsets were included to account for a known increase in transmissibility because of the emergence of new variants (up to and including Delta), and for a known decrease in the susceptible fraction resulting from immunity due to natural infection (until 2 July 2021) and primary vaccination. The factor *period* was added to correct for the effect of different contact rates due to changing intervention measures or school holidays and to absorb other mechanisms for gradual changes in transmissibility, such as compliance to measures, behaviour unrelated to measures, and population immunity not included in the offsets, by booster vaccination, natural waning of immunity and immune escape due to the emergence of the various Omicron variants.

We fitted 13 models with the last four terms of equation (1), i.e. temperature and/or absolute humidity, in the same week or lagged by one week, with a linear relation or with thin-plate regression splines. For this, we used generalised additive models (R package mgcv[29], version 1.9.0), with the REML optimisation method. We compared the models by AIC_c, Akaike Information Criterion with an additional correction[30] because of the high number of parameters (including 41 periods) relative to observations (109 weeks): $AIC_c = AIC + 2 \times df \times (df+1)/(n-df-1)$, in which AIC and *df* are calculated output from the mgcv package (*df* is estimated with spline models). A lower AIC_c by at least 2 points difference indicates a better model. All AIC_c values are reported as ΔAIC_c, relative to the model in equation (1) without the terms containing temperature and absolute humidity.

We did two sets of analyses to challenge our results against the possibility that our results are due to day-of-year as a confounder, i.e. that there was an association between $\log R_t$ and the weather only because both are seasonal..

First, we took the best fitting model from the main analysis and added variables on mobility (Google and NVP). All mobility variables per dataset were fitted in all possible combinations, with and without a lag of one week, creating $2^7 = 128$ Google models and $2^6 = 64$ NVP models. This tested robustness of the estimated magnitude of the weather effect, and addressed the question if the estimated associations are mediated by changes in behaviour as measured in the mobility data. If so, we would expect the strength of the association between $\log R_t$ and the weather to decrease.

Second, we fitted models with mean daily temperature and absolute humidity of the years 2000-2019, and with the five time series with the exact temperature and absolute humidity data of earlier years, all starting with a leap year: 2000-2002, 2004-2006, 2008-2010, 2012-2014, and 2016-2018. This addressed the question if $\log R_t$ can be explained by the general seasonal trend of the weather rather than the specific fluctuations in temperature and absolute humidity in the two years of the pandemic.

We did sensitivity analyses to address some specific choices we made. First, we repeated the main analysis with estimates of reproduction numbers for different days than Friday. Second, we repeated the main analysis with a single two-week period from 26 June 2021 to 9 July 2021, instead of two separate one-week periods. Third, we repeated the main analyses without offsets, to assess if the *period* variable will compensate for this; if so, we can be more confident that *period* will also absorb other gradually changing effects, in particular loss of immunity and the rise of the Omicron variant.

RESULTS

We see a clear preference (lower AIC_c) for the model with only a negative linear association between $\log R_t$ and temperature in the same week (Table 1). Including absolute humidity in addition to temperature did not result in a lower AIC_c value.

According to the selected model, R_t is expected to decrease by 2.2% (95% CI, 1.2% ; 3.2%) with every degree Celsius increase, which results in a ratio of R_t in Winter (around 1 February) vs Summer (around 1 August) of 1.5 (Table 1, Figure 2).

Table 1. Supports and estimated seasonal effects of models in the main analysis, including offsets and *period*. The selected model with lowest ΔAIC_c is bold-faced.

Model components				Model performance		R_t Winter vs Summer (95% CI)
Temperature	Absolute humidity	Spline or Linear	Lagged variables	ΔAIC_c [a]	*R*-squared	
-	-	-	-	0	0.68	1 (1 ; 1)
X	**-**	**Linear**	**-**	**-19.2**	**0.74**	**1.5 (1.2 ; 1.8)**
-	X	Linear	-	-10.5	0.72	1.5 (1.2 ; 1.9)
X	X	Linear	-	-14.6	0.74	1.4 (1.2 ; 1.8)
X	-	Linear	X	-10.5	0.73	1.5 (1.2 ; 1.9)
-	X	Linear	X	-6.2	0.71	1.4 (1.1 ; 1.9)
X	X	Linear	X	-6.9	0.74	1.4 (1.2 ; 1.8)
X	-	Spline	-	-15.1	0.75	1.8 (1.4 ; 2.5)
-	X	Spline	-	-10.5	0.72	1.5 (1.2 ; 1.9)
X	X	Spline	-	-9.4	0.76	1.8 (1.4 ; 2.5)
X	-	Spline	X	-9.2	0.75	1.8 (1.3 ; 2.6)
-	X	Spline	X	-6.2	0.71	1.4 (1.1 ; 1.9)
X	X	Spline	X	2.5	0.77	1.9 (1.4 ; 2.7)

[a] Corrected Akaike's Information Criterion, relative to that of the least complex model (first line)

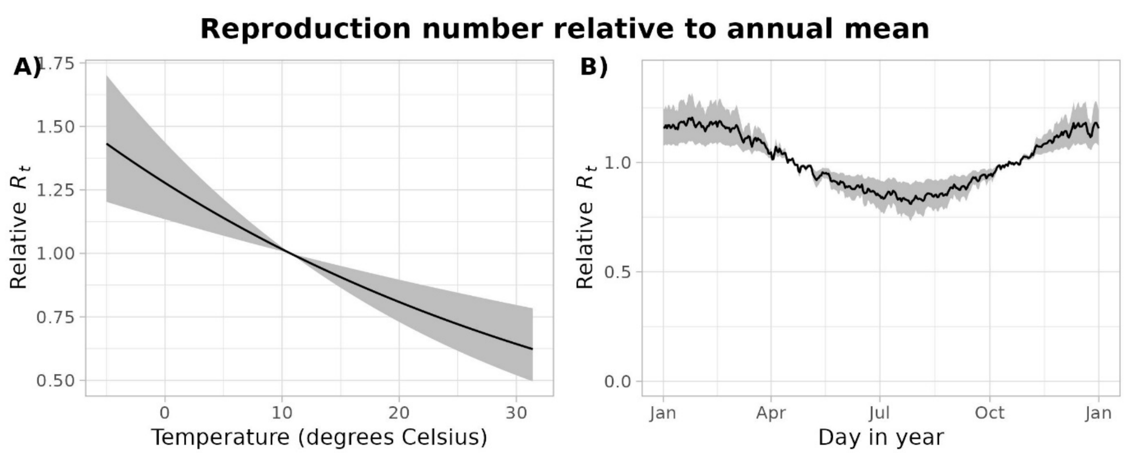

Figure 2. The change in reproduction number due to the association with temperature, expressed relative to the annual mean. (A) The direct relation with temperature; (B) the annual cycle using the relation between reproduction number and temperature, and the mean daily temperatures of the years 2000-2019.

Period and offsets explain 68% of the variance in logR_t (Table 1). The estimates of the period effects turn out to reflect the stringency of control measures very well, as illustrated by the qualitative agreement with the Oxford Stringency Index (Figure 3): if measures become more stringent, the reproduction number decreases. This qualitative agreement ends in Autumn 2021, after which the period-associated reproduction number tends to be higher, especially when Omicron arises in 2022.

The effect of temperature in determining overall transmissibility explained another 6% of variance in logR_t (Table 1). This is clearly visible when comparing the observed R_t and the predictions from the models with and without temperature (Figure 4): especially during longer periods without changes in control measures (June to October 2020, and February to April 2021), fluctuations in R_t are much better explained in the model with temperature.

To assess robustness of these results, we first addressed the magnitude of the estimated association with temperature, i.e. the result that the ratio of R_t in Winter vs Summer is 1.5. We did this by adding Google or NVP mobility variables to the selected model from the main analysis. Some combinations of Google mobility variables did improve the model: the best model (lowest AIC_c) included only *work* (without lag), explaining another 3% of variance in logR_t; in fact, all models with lower AIC_c than the reference model (with temperature, without Google variables) included *work* (Table 2), always with a negative association with logR_t. Nevertheless, the estimated ratio of R_t in Winter vs Summer remained at 1.5 (95% CI: 1.3; 1.8). The NVP data did not improve the model fit as measured by AIC_c; the best model with any NVP variable contained only *office*, with a negative association (results not shown).

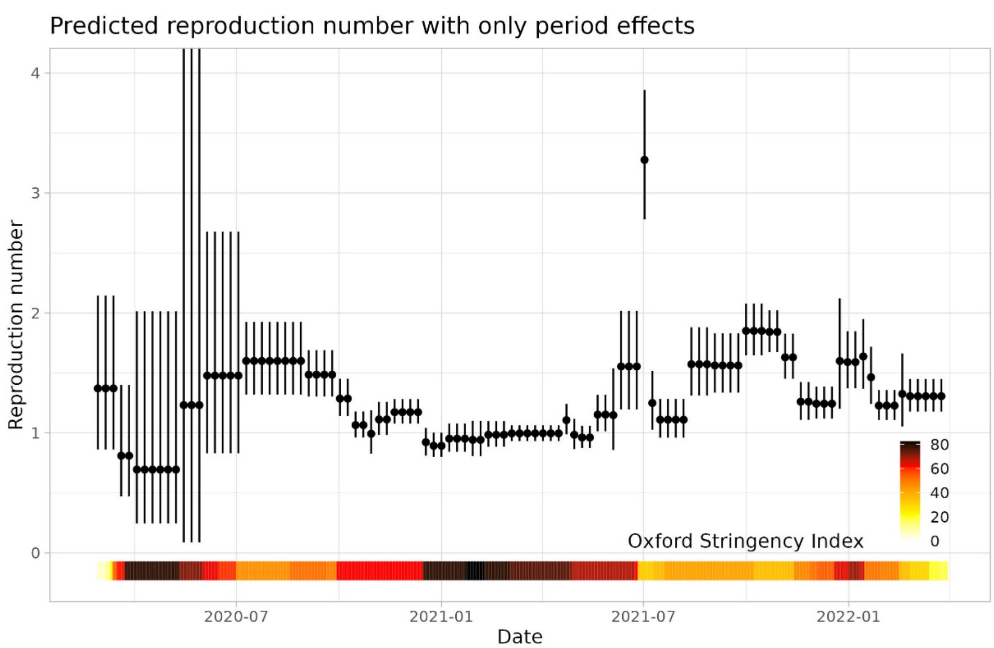

Figure 3. The reproduction number during the pandemic with mean annual temperature and without offsets, as predicted by the selected model with a linear temperature effect. It shows the contribution of the *period* estimates, mainly reflecting the effectiveness of interventions and adherence to interventions, but after the Summer of 2021 also waning of immunity and the emergence of the Omicron variants. The coloured bar indicates the Oxford Stringency index, a compound measure (scale 0-100) of overall stringency of government interventions[28].

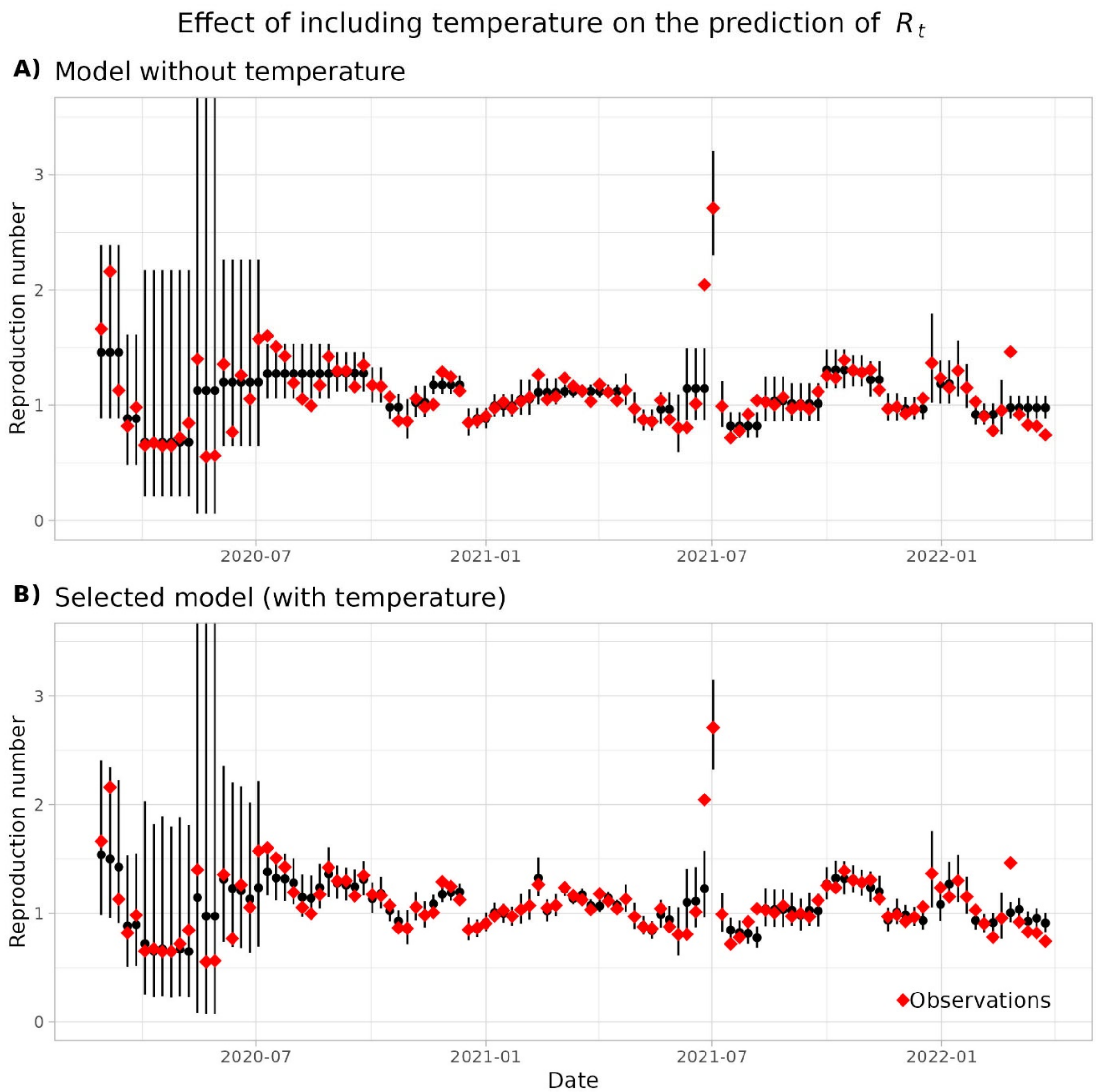

Figure 4. The reproduction number during the pandemic, predicted by (A) the model with only offsets and *period*, and (B) the selected model including a linear temperature effect.

Second, we addressed the question if the association of R_t with temperature was confounded by day of the year, i.e. if the association is only there because both are seasonal. We did this by fitting the $\log R_t$ data to alternative time series of temperature and/or absolute humidity: the means of the years 2000-2019, and five periods of exact measurements in earlier years. There were only three sets of weather variables showing better fits than the model without weather variables, suggesting an association to a general seasonal pattern, but none of the 18 models explained $\log R_t$ better than the exact temperatures from 2020-2022 (Table 3).

We repeated our main analyses with $\log R_t$ time series of other days than Fridays (Appendix S4), and found that our final model with only temperature as a linear term had highest support (lowest AIC_c) with data from all days except Wednesday and Thursday, in which case they still received considerable support (AIC_c was within two points of the minimum). The estimated ratio of reproduction number R_t in Winter vs Summer varied between 1.2 and 1.7, revealing that the outcome was not very sensitive to the choice of the day in the week.

We also repeated the main analysis with a single two-week period between 26 June 2021 and 9 July 2021 (Appendix S4), which resulted in a much more complex model with non-linear effects of temperature and absolute humidity receiving most support, and an estimated R_t in Winter vs Summer of 2.5 (1.7; 4.1). This finding confirms that results could become sensitive to the two data points indicated by evaluating Cook's distance in the preliminary analysis, thus justifying our choice of periods to use two separate one-week periods and reduce the influence of these data points.

Finally, we repeated the main analysis without the offset term, to check if the *period* variable indeed absorbs gradual changes in R_t due to effects not included in the model. Leaving out the offset did not change the estimated effects of temperature and absolute humidity, but did indeed result in *period* effects incorporating the offsets (Appendix S4).

Table 2. Additional analysis adressing mediation of the temperature effect on R_t by Google mobility variables. The final model of the main analysis (selected model) is shown with all 15/128 models having a ΔAIC_c lower than the selected model (lowest in bold).

Google mobility variables	Lagged Google mobility variables	ΔAIC_c [a]	R-squared	R_t Winter vs Summer
None (selected model)	-	-19.2	0.74	1.5 (1.2;1.8)
work	**No**	**-27.5**	**0.77**	**1.5 (1.3;1.8)**
parks + transit + work	No	-26.3	0.78	1.6 (1.3;1.9)
parks + transit + work	Yes	-24.8	0.80	1.6 (1.3;2.0)
retail + parks + work	No	-24.0	0.78	1.5 (1.3;1.9)
transit + work	No	-23.2	0.77	1.6 (1.3;1.9)
retail + work	No	-23.2	0.77	1.6 (1.3;1.9)
work + home	No	-22.9	0.76	1.6 (1.3;1.9)
parks + work	No	-22.4	0.76	1.5 (1.2;1.8)
grocery + work	No	-22.3	0.76	1.6 (1.3;1.9)
parks + work + home	No	-22.1	0.77	1.6 (1.3;2.0)
work	Yes	-21.9	0.76	1.5 (1.3;1.9)
retail + parks + transit + work	No	-21.2	0.78	1.6 (1.3;1.9)
retail + work	Yes	-21.1	0.78	1.5 (1.3;1.8)
park + transit + work + home	No	-20.8	0.78	1.6 (1.3;2.0)
grocery + parks + transit + work	No	-20.5	0.78	1.6 (1.3;2.0)
transit + work	Yes	-19.9	0.78	1.5 (1.2;1.8)

[a] Corrected Akaike's Information Criterion, relative to that of the least complex model, i.e. without mobility, temperature or absolute humidity

Table 3. Additional analysis adressing confounding of weather variables and the reproduction number R_t by day-of-year (co-seasonality). The table shows ΔAIC_c values of model fits with the original weather time series during the pandemic and alternative time series of other years (lowest in bold). ΔAIC_c is calculated relative to that of the least complex model, i.e. without temperature or absolute humidity.

Years of weather data	Weather variables included		
	only temperature	only absolute humidity	temperature and absolute humidity
2020-2022 (pandemic years)	**-19.2**	**-10.5**	**-14.6**
means of 2000-2019	4.1	4.9	9.2
2016-2018	-4.9	1.1	-11.0
2012-2014	4.5	5.2	8.8
2008-2010	4.7	5.5	0.6
2004-2006	1.9	-0.2	5.1
2000-2002	5.2	4.7	9.7

DISCUSSION

We estimated the association between transmissibility of SARS-CoV-2 during the pandemic in the Netherlands, and the mean daily temperature and absolute humidity. We found a negative linear association with temperature to explain 6% of variance in $\log R_t$ during the pandemic, with reproduction number R_t decreasing by 2.2% for every degree Celsius increase. Extrapolating to mean weekly temperatures during the year, this means that R_t early February (Winter) is 50% higher than R_t early August (Summer) in the Netherlands. We also found a negative association with mobility related to work.

To estimate the effect of temperature and absolute humidity, we had to separate this effect from the much larger effect on SARS-CoV transmission during the pandemic by several other factors. The effects of variants, immunity from infection, and vaccination could be included as offsets up to Autumn 2021, after which waning immunity and immune escape due to the Omicron variant obscured the level of immunity in the population. However, the varying sets of sometimes rapidly changing social distancing measures had even greater impacts which could not be estimated independently. To account for these, the stepwise constant *period* was added to the model with changepoints at the times of changing control measures and school holidays. These could explain large changes in $\log R_t$ (68% of variance) due to changing control policies, but also allowed for more gradual changes due to other effects such as changing behaviour due to epidemic fatigue, and changes in immunity or variant transmissibility not accounted for in the offsets. The fact that the estimated effect of the temperature does not change when leaving out the offsets (Appendix S4) shows that such gradual changes are indeed absorbed by the *period* variable. Including the offsets is useful however, because it allows us to interpret the coefficients for *period* as the effects of control measures, which could therefore be compared with the Oxford Stringency Index[28] (Figure 2a).

We addressed the possibility that the association was only a result of two seasonally co-oscillating variables without any direct or indirectly causal link, by fitting $\log R_t$ to alternative datasets of temperature and absolute humidity in the Netherlands on the same dates but in different years, and a 20-year average. With most datasets (15/18) there was no statistical evidence for an association between $\log R_t$ and weather variables at all, and no dataset could explain $\log R_t$ better than temperature in 2020-2022, the years of the pandemic. This result is a strong indication that the actual temperature was causally related with SARS-CoV-2 transmission during the pandemic. Regardless of the causal nature of the relationship, a change in the actual temperature is thus a good predictor of a change in transmissibility. Prediction models can take advantage of this by using actual temperature data to anticipate changes in incidence that will result from such changes in transmissibility.

The association between temperature and reproduction ratio could have been mediated by behaviour. We tested for mediation by behaviour as measured in two large mobility datasets, both measured through smartphones. One of the Google mobility variables (*work*) did improve the model, but seemed not to mediate the temperature effect because the association between $\log R_t$ and temperature did not change. Even though was no convincing statistical evidence for an association between the NVP mobility variables and $\log R_t$, the best performing NVP variable was *office*, which had a time series that resembled the Google mobility variable work (Figure 1e,f).

Unique to our study is that we aimed to quantify changes in SARS-CoV-2 transmissibility attributable to changes in weather variables, to assess the magnitude of seasonal variation in transmissibility in the Netherlands. For a reliable quantification it is important to restrict estimation to a limited geographic area, because the relationship may differ in different geographic locations, making a single estimate from multiple locations difficult to interpret. Keeping the model simple also adds to reliability of the estimation, because if weather variables are included with many other explanatory variables, there is a risk that the different variables are dependent or display spurious associations, resulting in potentially biased estimates. We prevented this by using *period* as a factor, so that no parametric assumptions were needed about the relative effectiveness of interventions in different phases of the pandemic. Although various mechanisms may be responsible for the quantitive relationship between weather variables and transmissibility,, the fact that alternative weather time series are not associated, and adding mobility data does not decrease the estimated effect of weather variables on transmissibility, suggests that using temperature as a covariate will increase the performance of prediction models of SARS-CoV-2 transmissibility and resulting COVID-19 incidence.

References

1. Martinez ME. The calendar of epidemics: Seasonal cycles of infectious diseases. *PLoS Pathog*. Nov 2018;14(11):e1007327. doi:10.1371/journal.ppat.1007327

2. Metcalf CJE, Walter KS, Wesolowski A, et al. Identifying climate drivers of infectious disease dynamics: recent advances and challenges ahead. *Proc Biol Sci*. Aug 16 2017;284(1860)doi:10.1098/rspb.2017.0901

3. te Beest DE, van Boven M, Hooiveld M, van den Dool C, Wallinga J. Driving factors of influenza transmission in the Netherlands. *Am J Epidemiol*. Nov 1 2013;178(9):1469-77. doi:10.1093/aje/kwt132

4. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog*. Oct 19 2007;3(10):1470-6. doi:10.1371/journal.ppat.0030151

5. Atchison CJ, Tam CC, Hajat S, van Pelt W, Cowden JM, Lopman BA. Temperature-dependent transmission of rotavirus in Great Britain and The Netherlands. *Proc Biol Sci*. Mar 22 2010;277(1683):933-42. doi:10.1098/rspb.2009.1755

6. Babin S. Use of Weather Variables in SARS-CoV-2 Transmission Studies. *Int J Infect Dis*. Nov 2020;100:333-336. doi:10.1016/j.ijid.2020.09.032

7. Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. Temperature, Humidity, and Latitude Analysis to Estimate Potential Spread and Seasonality of Coronavirus Disease 2019 (COVID-19). *JAMA Netw Open*. Jun 1 2020;3(6):e2011834. doi:10.1001/jamanetworkopen.2020.11834

8. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*. May 22 2020;368(6493):860-868. doi:10.1126/science.abb5793

9. Sera F, Armstrong B, Abbott S, et al. A cross-sectional analysis of meteorological factors and SARS-CoV-2 transmission in 409 cities across 26 countries. *Nat Commun*. Oct 13 2021;12(1):5968. doi:10.1038/s41467-021-25914-8

10. Ma Y, Pei S, Shaman J, Dubrow R, Chen K. Role of meteorological factors in the transmission of SARS-CoV-2 in the United States. *Nat Commun*. Jun 14 2021;12(1):3602. doi:10.1038/s41467-021-23866-7

11. Wagatsuma K. Association of Ambient Temperature and Absolute Humidity with the Effective Reproduction Number of COVID-19 in Japan. *Pathogens*. Nov 1 2023;12(11)doi:10.3390/pathogens12111307

12. Wei Y, Dong Z, Fan W, et al. A narrative review on the role of temperature and humidity in COVID-19: Transmission, persistence, and epidemiological evidence. *Eco Environ Health*. Jun 2022;1(2):73-85. doi:10.1016/j.eehl.2022.04.006

13. Paireau J, Charpignon ML, Larrieu S, et al. Impact of non-pharmaceutical interventions, weather, vaccination, and variants on COVID-19 transmission across departments in France. *BMC Infect Dis*. Mar 30 2023;23(1):190. doi:10.1186/s12879-023-08106-1

14. Xu R, Rahmandad H, Gupta M, et al. Weather, air pollution, and SARS-CoV-2 transmission: a global analysis. *Lancet Planet Health*. Oct 2021;5(10):e671-e680. doi:10.1016/S2542-5196(21)00202-3

15. Covid-19 reproductiegetal. 2023. https://data.rivm.nl/meta/srv/dut/catalog.search#/metadata/ed0699d1-c9d5-4436-8517-27eb993eab6e

16. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol*. Sep 15 2004;160(6):509-16. doi:10.1093/aje/kwh255

17. Gostic KM, McGough L, Baskerville EB, et al. Practical considerations for measuring the effective reproductive number, Rt. *PLoS Comput Biol*. Dec 2020;16(12):e1008409. doi:10.1371/journal.pcbi.1008409

18. Covid-19 rapportage van SARS-CoV-2 varianten in Nederland via de aselecte steekproef van RT-PCR positieve monsters in de nationale kiemsurveillance. National Institute for Public Health and the Environment (RIVM). 2024. https://data.rivm.nl/covid-19/COVID-19_varianten.csv

19. Campbell F, Archer B, Laurenson-Schafer H, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Euro Surveill*. Jun 2021;26(24)doi:10.2807/1560-7917.ES.2021.26.24.2100509

20. COVID-19 besmettelijke personen per dag. National Institute for Public Health and the Environment (RIVM). 2021. https://data.rivm.nl/meta/srv/dut/catalog.search#/metadata/097155aa-75eb-4caa-8ed3-4c6edb80467e

21. Covid-19 Cumulatieve opkomst voor tenminste één COVID-19 vaccinatie, vaccinatiegraad basisserie, vaccinatiegraad herhaalprik tegen corona binnen de najaarsronde 2022 naar gemeente en veiligheidsregio per geboortecohort per week. National Institute for Public Health and the Environment (RIVM). 2022. https://data.rivm.nl/covid-19/COVID-19_vaccinatiegraad_per_gemeente_per_week_leeftijd.csv

22. Huiberts AJ, de Gier B, Hoeve CE, et al. Vaccine effectiveness of primary and booster COVID-19 vaccinations against SARS-CoV-2 infection in the Netherlands from July 12, 2021 to June 6, 2022: A prospective cohort study. *Int J Infect Dis*. Aug 2023;133:36-42. doi:10.1016/j.ijid.2023.04.401

23. Cook RD. Influential observations in linear regression. *Journal of the American Statistical Association*. 1979;74(365):169-174. doi:10.1080/01621459.1979.10481634

24. Daggegevens van het weer in Nederland. Royal Netherlands Meteorological Institute (KNMI). 2022. https://www.knmi.nl/nederland-nu/klimatologie/daggegevens

25. Wu P, Goldstein E, Ho LM, et al. Excess mortality associated with influenza A and B virus in Hong Kong, 1998-2009. *J Infect Dis*. Dec 15 2012;206(12):1862-71. doi:10.1093/infdis/jis628

26. Covid-19 Community Mobility Reports. Google. 2022. https://www.google.com/covid19/mobility/

27. Het NVP tijdens COVID-19. Nederlands Verplaatsingspanel. 2023. https://www.goudappel.nl/nl/expertises/data-en-it-oplossingen/nederlands-verplaatsingspanel/het-nvp-tijdens-covid-19

28. Hale T, Angrist N, Goldszmidt R, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav*. Apr 2021;5(4):529-538. doi:10.1038/s41562-021-01079-8

29. Wood SN. *Generalized additive models : an introduction with R*. Second edition. ed. Chapman & Hall/CRC texts in statistical science. CRC Press/Taylor & Francis Group; 2017:xx, 476 pages.

30. Burnham KP, Anderson DR, Burnham KP. *Model selection and multimodel inference : a practical information-theoretic approach*. 2nd ed. Springer; 2002:xxvi, 488 p.