

Analysis COVID-seasonality: results for publication

Don Klinkenberg

2024-11-28

Introduction

This document contains the code for the analyses in the publication “Seasonal variation in SARS-CoV-2 transmission in the Netherlands, 2020-2022: statistical evidence for a negative association with temperature.” by Klinkenberg et al.

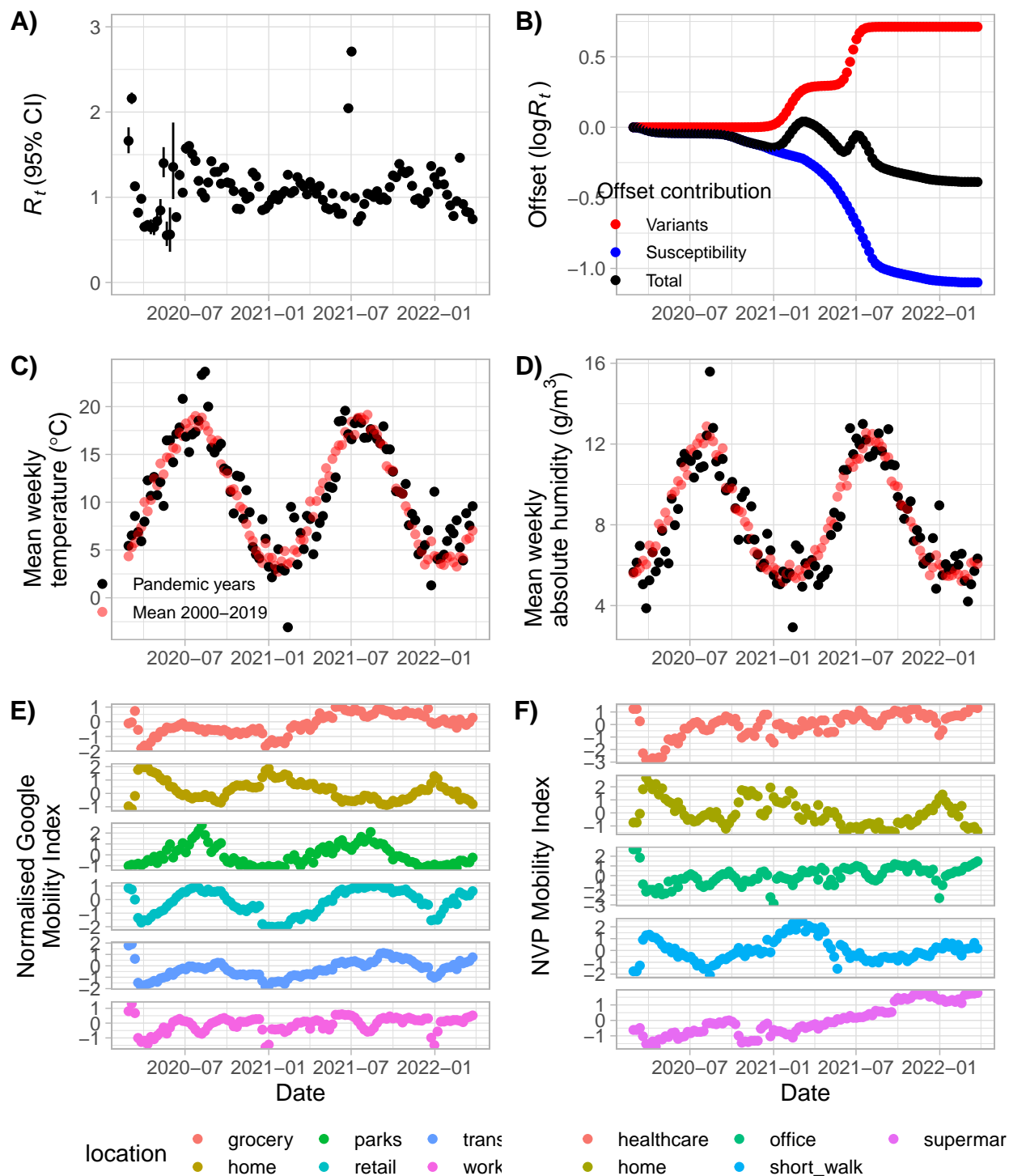
Loading and processing the data

Data that is read and processed:

- population size 2019.
- logRt and variance per day (file dated April 4, 2022). The inverse ($1/\text{variance}$) is included as a weight in the regression. Both of these are set back in time five days, to go from the Rt on the day of symptom onset to the Rt on the day of infection, an average incubation period back.
- cumulative incidence per day, estimated from hospital admissions (file dated July 3, 2021). To calculate the offset $\log(\text{proportion_susceptible})$.
- relative transmissibility based on variants per day, estimated from genomic surveillance (file dated October 5, 2021), to calculate the offset $\log(\text{relativeRt_variants})$.
- vaccine uptake per week, age and dose (file dated July 10, 2023). To calculate the offset $\log(\text{proportion_susceptible})$. For this purpose, the age-independent vaccination rate of the completed primary series is used, and this is multiplied by 0.75 due to the estimated VE against Delta (<https://www.sciencedirect.com/science/article/pii/S1201971223005349>).
- intervention periods: 41 periods between January 1, 2020 and April 1, 2023. After March 2020, there were four periods longer than a month: March 28 2020-May 10 2020; June 1 2020-July 5 2020; July 5 2020-Aug 30 2020; February 28 2021-Apr 18 2021.
- KNMI data (file dated July 11, 2023) containing daily average temperature and relative (!) humidity per day in De Bilt, from January 1, 2000. These are used to calculate the absolute humidity. With this data, five alternative (temp,ah) time series were created, by taking the data from four years earlier, eight years earlier, etc. (so that the leap day is correct). The years 2000-2019 were used for the multi-year average.
- Google mobility reports (files from 2020, 2021, 2022), containing a daily (relative) score in six categories: work, grocery, residential, transit, retail, and parks.
- NVP data (file dated September 15, 2023), containing a daily (relative score) in five categories from March 2, 2020: people staying at home, short walk, supermarket, office, care.

All these data are combined into a complete dataset ‘alldata_daily’. A complete dataset ‘alldata_weekly’ is also created from this, in which a running average over 7 days is made of all climate and Google data, and to which the NVP data is added. The final datasets run from February 22, 2020 to March 29, 2023.

Time series data



Autocorrelation and influence (results in supplement)

We start by looking at autocorrelation, in analyses with daily data and weekly data, and influence of individual datapoints in analyses with weekly data. We perform a total of eight regression analyses with the model

$$\log(R_t) \sim \text{offsets} + \text{period} [+ \text{dayinweek}] [+ \text{delayed logRt}] [+ \text{temp} + \text{ah}]$$

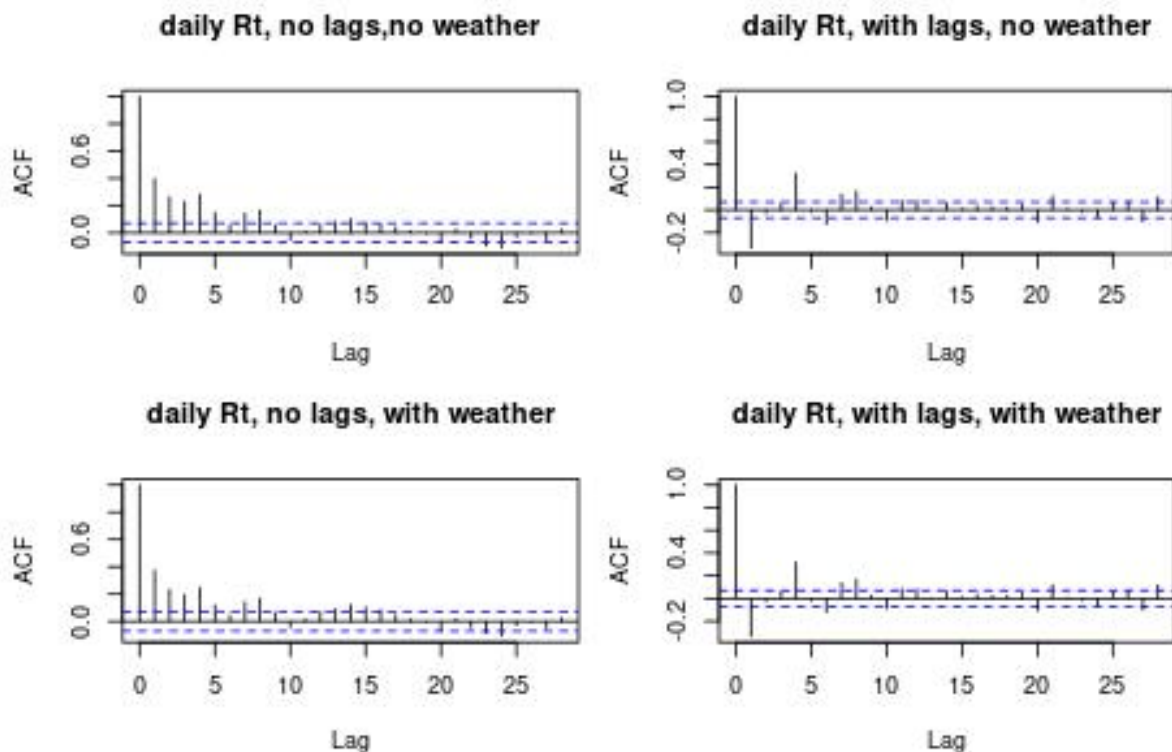
The eight analyses vary by

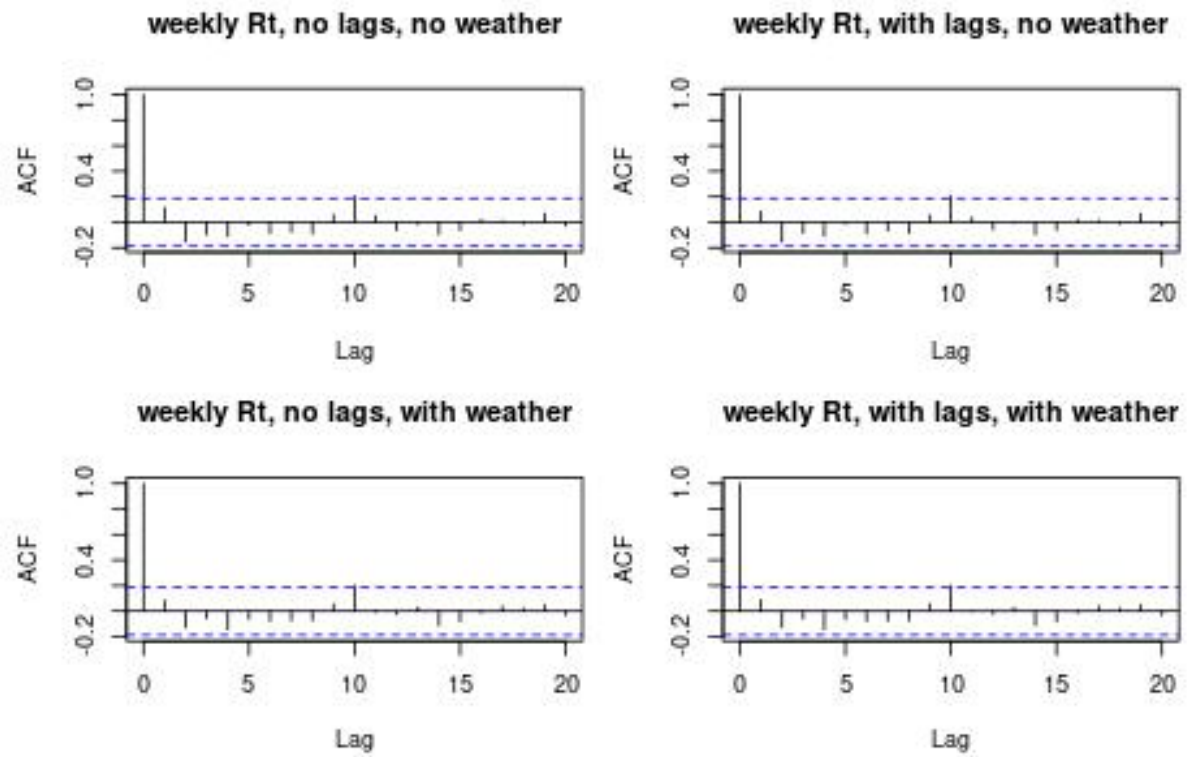
- daily data (with variable dayinweek) or weekly data (from Fridays).
- with or without delay terms: $\log R(t-1) + \log R(t-2) + \dots + \log R(t-7)$ for daily, and $\log R(t-1)$ for weekly.
- with or without temperature and humidity.

The daily analyses also contain the variable dayinweek. The weekly analyses are done with Friday data as default, because these are estimated on Wednesdays when the least disruptions in reports are expected due to weekend effects.

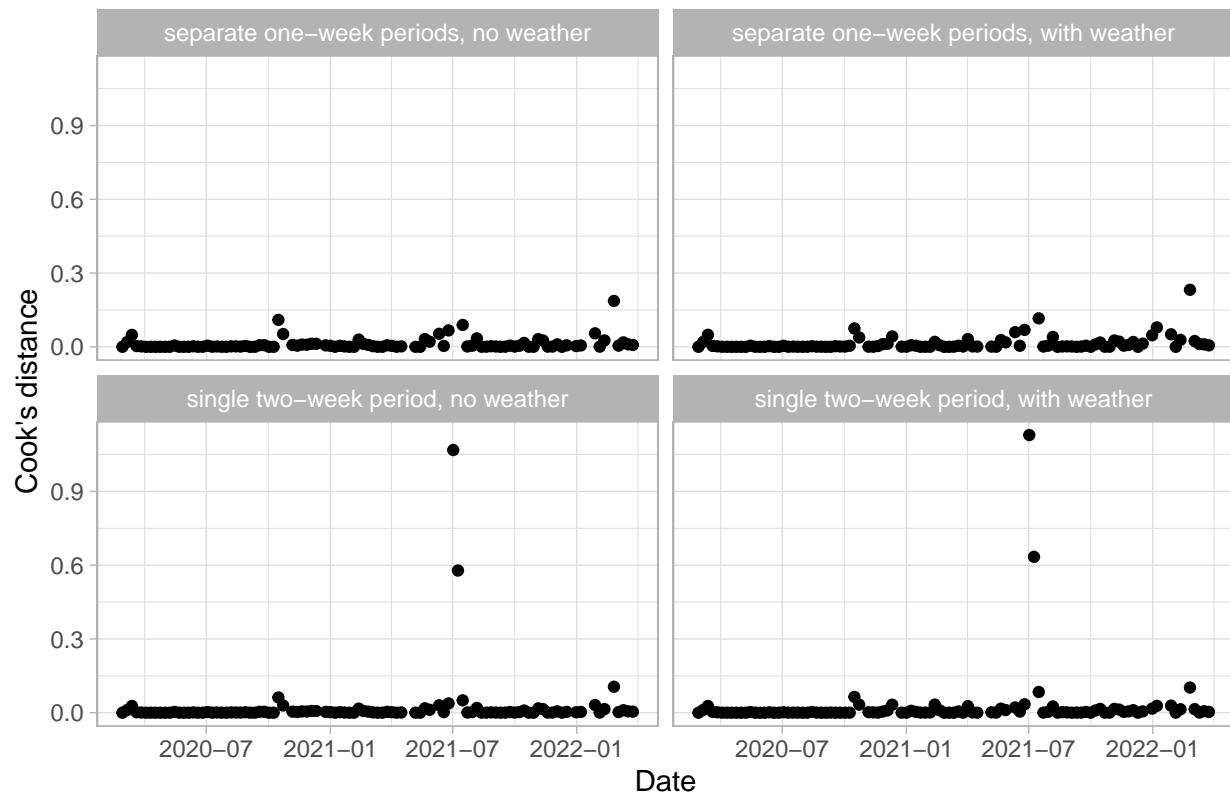
Table 1: autocorrelation tests

modelid	DelayTerms	DailyData	Seasonality	BreuschGodfrey_p
1	FALSE	FALSE	FALSE	0.783
2	FALSE	FALSE	TRUE	0.171
3	TRUE	FALSE	FALSE	0.805
4	TRUE	FALSE	TRUE	0.377
5	FALSE	TRUE	FALSE	0.000
6	FALSE	TRUE	TRUE	0.000
7	TRUE	TRUE	FALSE	0.120
8	TRUE	TRUE	TRUE	0.063





Influence of datapoints on regression results



Main analysis: weather variables

In the main analysis, we look at models in which temperature and humidity are included as linear terms in the model or as splines, both those of the week itself and those of the week before. Furthermore, the model only contains the offset and period as explanatory variables. This gives thirteen different regression analyses:

$$\log(R_t) \sim \text{offsets} + \text{period} [+ s(\text{temp}_t)] [+ s(\text{ah}_t)] [+ s(\text{temp}_{t-1})] [+ s(\text{ah}_{t-1})]$$

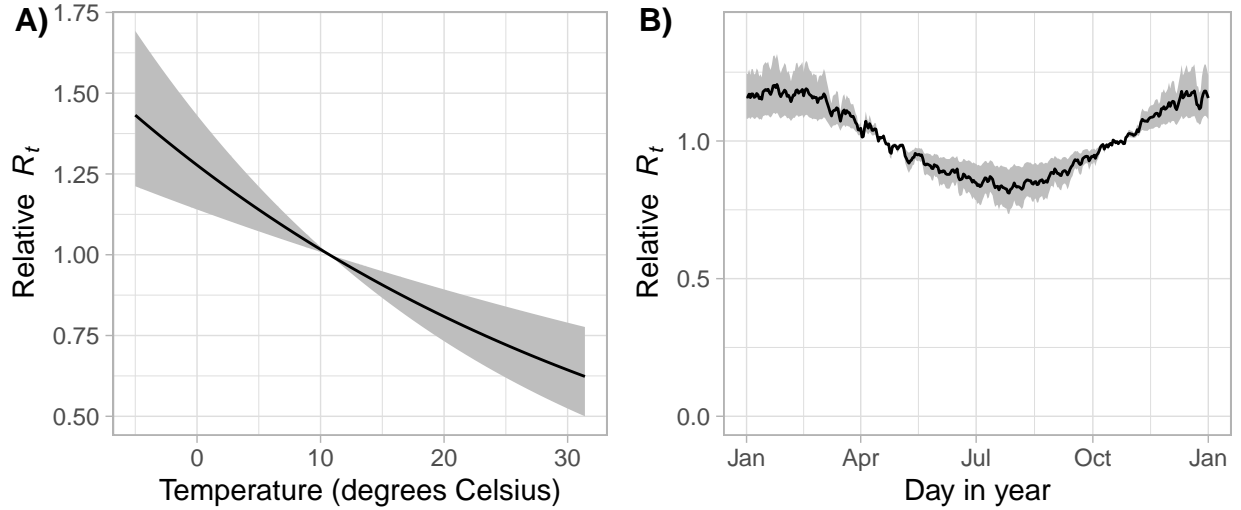
Table 2: Main analysis, linear models

Delay terms	Humidity	Temperature	Rsqr	dAICc	Rtmax/Rtmin
FALSE	FALSE	TRUE	0.74	-19.20	1.49 (1.23;1.8)
FALSE	TRUE	TRUE	0.74	-14.58	1.44 (1.21;1.79)
TRUE	FALSE	TRUE	0.73	-13.57	1.47 (1.15;1.91)
FALSE	TRUE	FALSE	0.72	-10.46	1.5 (1.17;1.92)
TRUE	TRUE	TRUE	0.74	-6.87	1.41 (1.2;1.77)
TRUE	TRUE	FALSE	0.71	-6.19	1.36 (1.09;1.88)
FALSE	FALSE	FALSE	0.68	0.00	1 (1;1)
TRUE	FALSE	FALSE	0.68	0.00	1 (1;1)

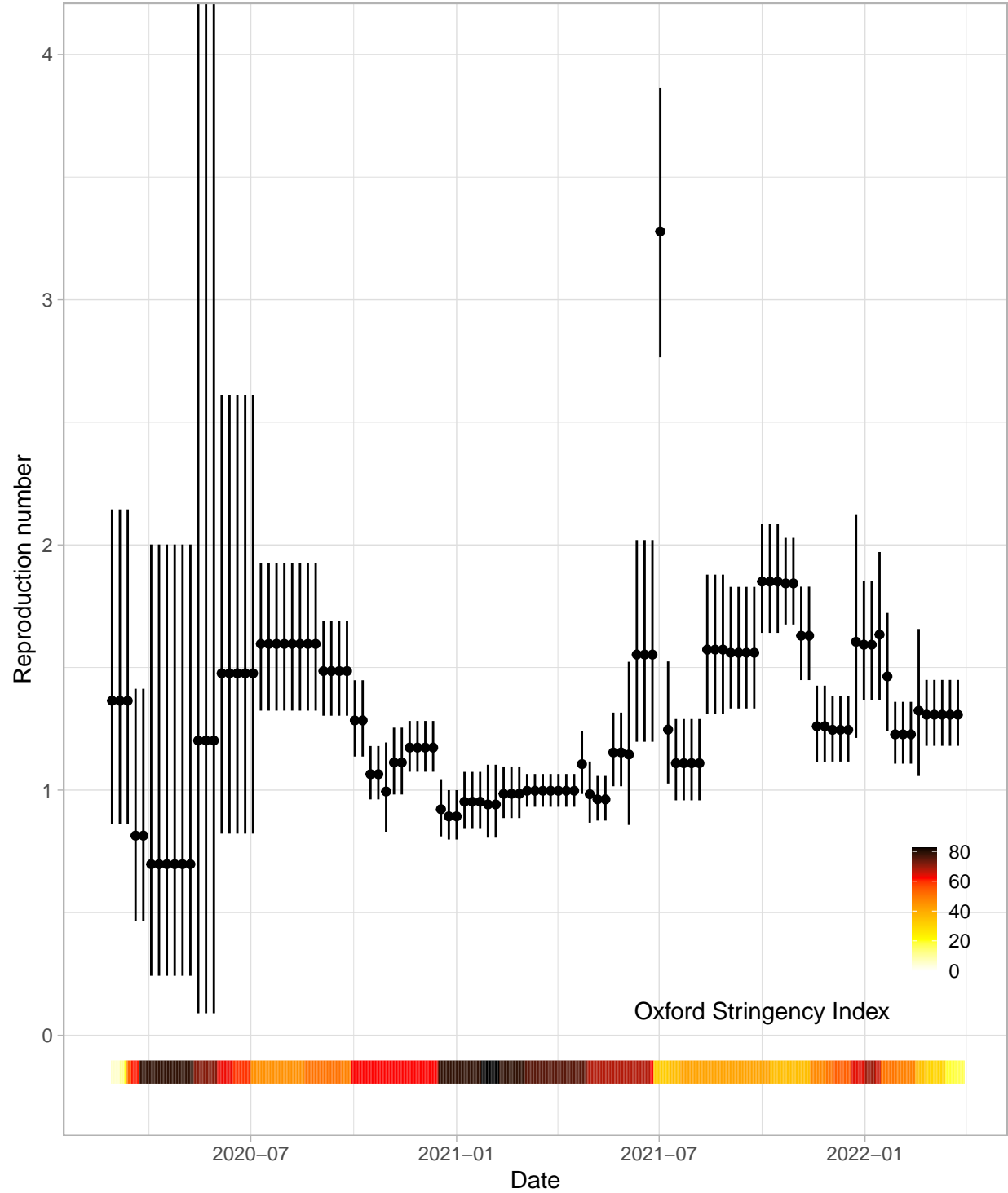
Table 3: Main analysis, spline models

Delay terms	Humidity	Temperature	Rsqr	dAICc	Rtmax/Rtmin
FALSE	FALSE	TRUE	0.75	-15.12	1.84 (1.4;2.46)
FALSE	TRUE	FALSE	0.72	-10.46	1.5 (1.17;1.92)
FALSE	TRUE	TRUE	0.76	-9.35	1.84 (1.4;2.45)
TRUE	FALSE	TRUE	0.75	-9.21	1.83 (1.32;2.57)
TRUE	TRUE	FALSE	0.71	-6.19	1.36 (1.09;1.9)
FALSE	FALSE	FALSE	0.68	0.00	1 (1;1)
TRUE	FALSE	FALSE	0.68	0.00	1 (1;1)
TRUE	TRUE	TRUE	0.77	2.51	1.89 (1.4;2.68)

Reproduction number relative to annual mean

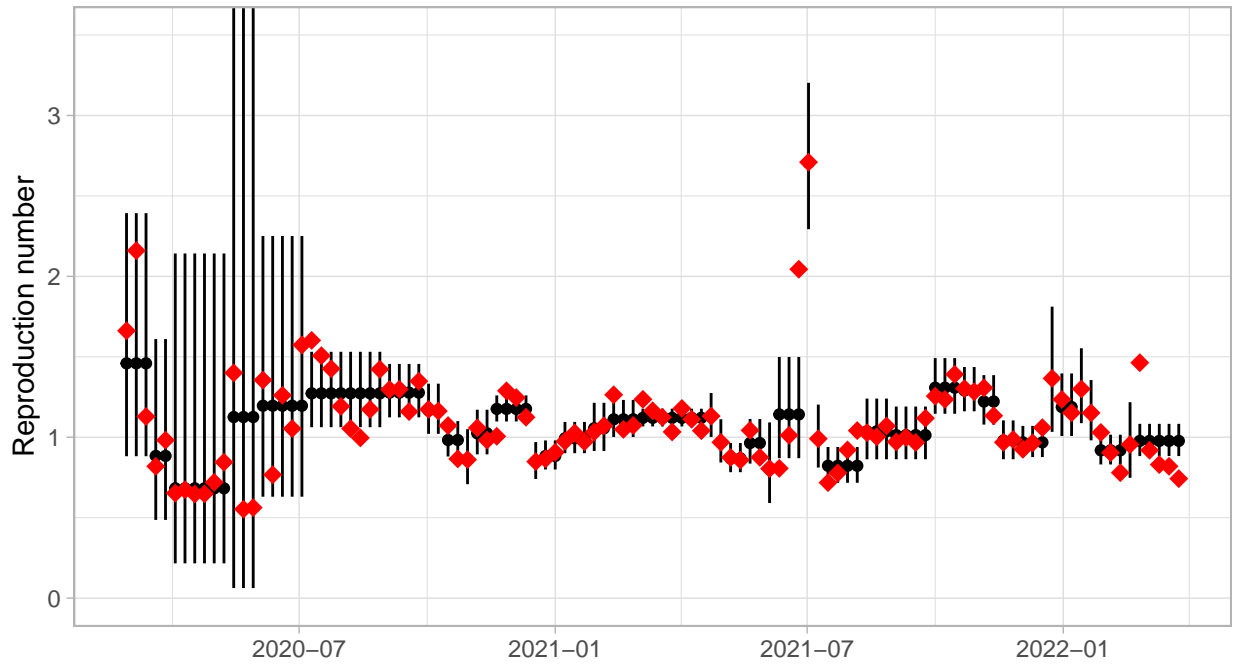


Predicted reproduction number with only period effects

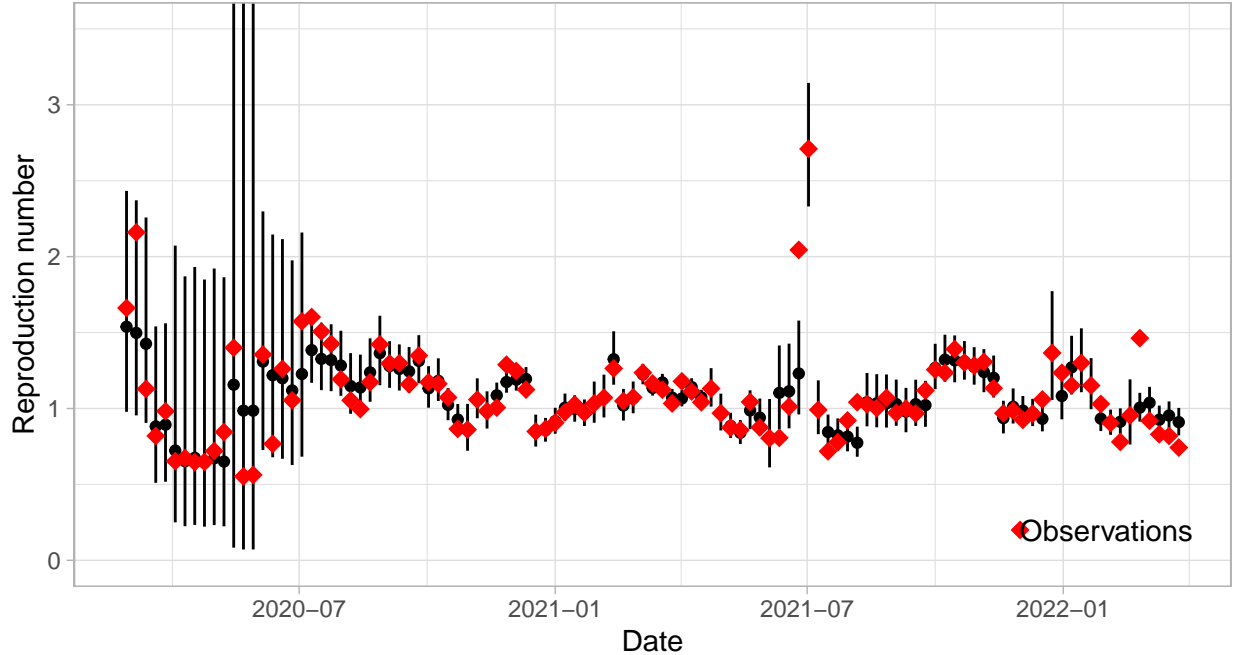


Effect of including temperature on the prediction of R_t

A) Model without temperature



B) Selected model (with temperature)



Additional analysis: alternative temperature and humidity time series

These analyses were meant to see if the relation with temperature is due to the confounder day-of-year, potentially giving rise to a correlation between two independent seasonally varying variables.

Table 4: dAICc weather time series

Years	only ah	only T	T and ah
average	4.09	4.92	9.21
2020-2022	-19.20	-10.46	-14.58
2016-2018	-4.92	1.11	-10.95
2012-2014	4.50	5.18	8.77
2008-2010	4.68	5.49	0.62
2004-2006	1.89	-0.16	5.09
2000-2002	5.16	4.72	9.71

Additional analyses: Google mobility en Netherlands Translocation Panel (NVP)

These analyses were meant to see if the relation is mediated by mobility as measured by Google and by NVP. None of the NVP models explained the data better than the selected model of the main analysis.

Table 5: Sensitivity Google mobility

Delay terms	retail	grocery	parks	transit	work	residential	Rsqr	dAICc	Rtmax/Rtmin
FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	0.77	-27.54	1.5 (1.3;1.8)
FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	0.78	-26.35	1.6 (1.3;1.9)
TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	0.80	-24.75	1.6 (1.3;1.9)
FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	0.78	-24.02	1.5 (1.3;1.8)
FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	0.77	-23.20	1.6 (1.3;1.9)
FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	0.77	-23.19	1.6 (1.3;1.9)
FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	0.76	-22.89	1.6 (1.3;2)
FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	0.76	-22.44	1.5 (1.2;1.8)
FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	0.76	-22.25	1.6 (1.3;1.9)
FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	0.77	-22.07	1.6 (1.3;2)
TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	0.76	-21.87	1.5 (1.3;1.9)
FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	0.78	-21.24	1.6 (1.3;1.9)
TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	0.78	-21.17	1.5 (1.3;1.8)
FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	0.78	-20.84	1.6 (1.3;2)
FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	0.78	-20.51	1.6 (1.3;1.9)
TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	0.78	-19.92	1.5 (1.2;1.8)
FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	0.74	-19.20	1.5 (1.2;1.8)
TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	0.74	-19.20	1.5 (1.2;1.8)

Sensitivity analyses

Treating the two-week period 26 June 2021 - 9 July 2021 as one period

This is around “Dansen met Janssen”, when nightlife was opened for a brief period of time, leading to an explosion of new cases.

Table 6: Single two-week period, linear models

Delay terms	Humidity	Temperature	dAICc	Rtmax/Rtmin
FALSE	TRUE	TRUE	-12.71	1.5 (1.2;1.9)

Delay terms	Humidity	Temperature	dAICc	Rtmax/Rtmin
FALSE	FALSE	TRUE	-12.01	1.6 (1.2;2)
TRUE	FALSE	TRUE	-6.57	1.5 (1.1;2.2)
FALSE	TRUE	FALSE	-1.57	1.4 (1.1;2)
TRUE	TRUE	TRUE	-1.31	1.5 (1.2;2.1)
FALSE	FALSE	FALSE	0.00	1 (1;1)
TRUE	FALSE	FALSE	0.00	1 (1;1)
TRUE	TRUE	FALSE	3.85	1.4 (1.1;2.2)

Table 7: Single two-week period, spline models

Delay terms	Humidity	Temperature	dAICc	Rtmax/Rtmin
FALSE	TRUE	TRUE	-18.96	2.5 (1.7;3.7)
FALSE	FALSE	TRUE	-12.47	2.6 (1.7;3.9)
TRUE	FALSE	TRUE	-6.60	2.5 (1.6;4.1)
FALSE	TRUE	FALSE	-1.57	1.4 (1.1;2)
TRUE	TRUE	TRUE	-1.42	2.6 (1.7;4)
FALSE	FALSE	FALSE	0.00	1 (1;1)
TRUE	FALSE	FALSE	0.00	1 (1;1)
TRUE	TRUE	FALSE	7.94	1.5 (1.1;2.5)

Day of week

Using logRt of all days-of-week one by one, instead of Friday as in the main analysis.

Table 8: dAICc, day of week, linear models

temp	ah	Xdelay	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
FALSE	FALSE	FALSE	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TRUE	FALSE	FALSE	-19.20	-20.99	-29.19	-27.39	-10.82	-5.74	-14.58
FALSE	TRUE	FALSE	-10.46	-10.36	-16.68	-16.07	-1.50	0.83	-9.68
TRUE	TRUE	FALSE	-14.58	-17.53	-24.67	-22.64	-10.52	-4.27	-9.13
TRUE	FALSE	TRUE	-13.57	-15.98	-25.35	-24.46	-5.30	-0.74	-9.27
FALSE	TRUE	TRUE	-6.19	-5.43	-12.53	-12.44	4.09	6.14	-4.03
TRUE	TRUE	TRUE	-6.87	-6.78	-14.92	-13.96	0.76	6.56	0.62

Table 9: dAICc, day of week, spline models

temp	ah	Xdelay	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
FALSE	FALSE	FALSE	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TRUE	FALSE	FALSE	-15.12	-17.14	-24.56	-19.89	-8.08	-4.80	-15.79
FALSE	TRUE	FALSE	-10.46	-10.36	-16.68	-16.06	1.42	-1.49	-3.62
TRUE	TRUE	FALSE	-9.35	-11.69	-21.32	-15.90	-1.48	-6.45	-12.70
TRUE	FALSE	TRUE	-9.21	-12.09	-20.74	-17.19	-2.23	5.17	-9.91
FALSE	TRUE	TRUE	-6.19	-1.86	-9.49	-12.44	7.61	3.08	2.28
TRUE	TRUE	TRUE	2.51	3.95	-6.25	-7.85	10.75	6.36	-1.14

Table 10: Rtmax/Rtmin, day of week, linear models

temp	ah	Xdelay	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
FALSE	FALSE	FALSE	1 (1;1)	1 (1;1)	1 (1;1)	1 (1;1)	1 (1;1)	1 (1;1)	1 (1;1)
TRUE	FALSE	FALSE	1.5 (1.2;1.8)	1.5 (1.2;1.7)	1.5 (1.3;1.8)	1.4 (1.2;1.6)	1.2 (1.1;1.4)	1.2 (1.1;1.4)	1.3 (1.1;1.5)
FALSE	TRUE	FALSE	1.5 (1.2;1.9)	1.5 (1.2;1.8)	1.6 (1.3;1.9)	1.4 (1.2;1.6)	1.2 (1;1.4)	1.2 (1;1.4)	1.3 (1.1;1.6)
TRUE	TRUE	FALSE	1.4 (1.2;1.8)	1.4 (1.2;1.7)	1.5 (1.3;1.8)	1.4 (1.2;1.6)	1.2 (1.1;1.4)	1.2 (1.1;1.4)	1.3 (1.1;1.5)
TRUE	FALSE	TRUE	1.5 (1.1;1.9)	1.4 (1.1;1.8)	1.4 (1.2;1.8)	1.3 (1.1;1.5)	1.2 (1.1;1.5)	1.3 (1.1;1.6)	1.4 (1.1;1.7)
FALSE	TRUE	TRUE	1.4 (1.1;1.9)	1.4 (1.1;1.9)	1.4 (1.1;1.9)	1.3 (1.1;1.6)	1.2 (1;1.5)	1.2 (1;1.5)	1.3 (1.1;1.7)
TRUE	TRUE	TRUE	1.4 (1.2;1.8)	1.4 (1.2;1.8)	1.4 (1.2;1.8)	1.3 (1.1;1.5)	1.3 (1.1;1.5)	1.3 (1.1;1.6)	1.3 (1.1;1.6)

Table 11: Rtmax/Rtmin, day of week, spline models

temp	ah	Xdelay	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
FALSE	FALSE	FALSE	1 (1;1)	1 (1;1)	1 (1;1)	1 (1;1)	1 (1;1)	1 (1;1)	1 (1;1)
TRUE	FALSE	FALSE	1.9 (1.4;2.5)	1.6 (1.3;2)	1.7 (1.4;2.1)	1.5 (1.2;1.8)	1.3 (1.1;1.7)	1.5 (1.2;1.9)	1.7 (1.3;2.2)
FALSE	TRUE	FALSE	1.5 (1.2;1.9)	1.5 (1.2;1.8)	1.6 (1.2;1.9)	1.4 (1.2;1.6)	1.2 (1.1;1.4)	1.3 (1.1;1.6)	1.4 (1.1;1.8)
TRUE	TRUE	FALSE	1.8 (1.4;2.5)	1.6 (1.3;2)	1.7 (1.3;2.1)	1.5 (1.2;1.8)	1.4 (1.2;1.7)	1.6 (1.3;2)	1.7 (1.4;2.2)
TRUE	FALSE	TRUE	1.8 (1.3;2.6)	1.6 (1.2;2.1)	1.6 (1.2;2)	1.4 (1.1;1.7)	1.3 (1.1;1.7)	1.5 (1.2;2)	1.8 (1.3;2.3)
FALSE	TRUE	TRUE	1.4 (1.1;1.9)	1.5 (1.2;2.1)	1.6 (1.2;2.1)	1.3 (1.1;1.6)	1.2 (1.1;1.5)	1.4 (1.2;1.8)	1.4 (1.1;1.9)
TRUE	TRUE	TRUE	1.9 (1.4;2.7)	1.7 (1.3;2.3)	1.6 (1.3;2.2)	1.4 (1.2;1.8)	1.4 (1.2;1.9)	1.7 (1.4;2.2)	1.7 (1.3;2.3)

Offset

Not including an offset in the analyses (models have already been run).

Table 12: No offset, linear models

Delay terms	Humidity	Temperature	Rsqr	dAICc	Rtmax/Rtmin
FALSE	FALSE	TRUE	0.74	-16.05	1.4 (1.2;1.7)
FALSE	TRUE	TRUE	0.74	-12.57	1.4 (1.2;1.7)
TRUE	FALSE	TRUE	0.73	-10.56	1.4 (1.1;1.8)
FALSE	TRUE	FALSE	0.71	-6.86	1.4 (1.1;1.8)
TRUE	TRUE	TRUE	0.74	-6.63	1.4 (1.2;1.7)
TRUE	TRUE	FALSE	0.71	-3.62	1.3 (1.1;1.7)
FALSE	FALSE	FALSE	0.68	0.00	1 (1;1)
TRUE	FALSE	FALSE	0.68	0.00	1 (1;1)

Table 13: No offset, spline models

Delay terms	Humidity	Temperature	Rsqr	dAICc	Rtmax/Rtmin
FALSE	FALSE	TRUE	0.75	-12.01	1.8 (1.3;2.4)
FALSE	TRUE	TRUE	0.76	-7.70	1.8 (1.4;2.4)
FALSE	TRUE	FALSE	0.71	-6.86	1.4 (1.1;1.8)
TRUE	FALSE	TRUE	0.75	-6.24	1.8 (1.3;2.4)
TRUE	TRUE	FALSE	0.71	-3.62	1.3 (1.1;1.7)
FALSE	FALSE	FALSE	0.68	0.00	1 (1;1)
TRUE	FALSE	FALSE	0.68	0.00	1 (1;1)
TRUE	TRUE	TRUE	0.78	2.15	1.8 (1.4;2.6)

Estimated period effect in relation to the offset

