eAppendix 1 for:

Daylight savings time transitions and the incidence rate of unipolar depressive episodes

Contents

1. Description of transfer function intervention analysis

This paper followed the approach to intervention analysis given in Cryer and Chan (1) and originally introduced by Box and Tiao (2). This approach models the time series being studied as a combination of an autoregressive integrated moving average model process (ARIMA) and a transfer function of a binary variable that estimates how the interventions – here, the time transitions – act upon this process.

The generic (non-seasonal) intervention model can be written as follows:*

$$y_t = \mu + \frac{\theta(B)}{\phi(B)}\varepsilon_t + \frac{\omega(B)}{\delta(B)}B^k x_t$$

Where:

y is the time series in question or a difference of it. In our case the first difference of the series of weekly incidence rates of unipolar depression or bipolar disorder in Denmark.

t indexes time.

μ is a constant.

B is the backshift operator, defined such that: $B^k y_t = y_{t-k}$.

$\phi(B)$ is the autoregressive parameter, which consists of a polynomial of order p in the backshift operator such that: $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 \ldots \phi_p B^p$.

$\theta(B)$ is the moving average parameter, which consists of a polynomial of order q in the backshift operator such that: $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 \ldots - \theta_q B^q$.

ε is a white noise disturbance term (i.e. $\varepsilon \sim i.i.d.(0, \sigma^2)$).

x is the intervention variable.

$\omega(B)$ is the numerator polynomial of the transfer function for the intervention variable, which consists of a polynomial in the backshift operator of order f such that: $\omega(B) = \omega_0 + \omega_1 B + \omega_2 B^2 \ldots + \omega_f B^f$.

* The exposition extends directly to the seasonal case.

$\delta(B)$ is the denominator polynomial of the transfer function for the intervention variable, which consists of a polynomial in the backshift operator of order g such that: $\delta(B) = 1 - \delta_1 B - \delta_2 B^2 \ldots - \delta_g B^g$.

This expression can be written more succinctly as:

$$y_t = \mu + v(B)x_t + N_t$$

Where $v(B)$ is the transfer function and N_t is the ARIMA noise process. The latter term can be thought of as the unperturbed process; the path of the time series had there been no intervention. It serves as the counterfactual for the analysis and given the exogeneity of x, the deviation from this process, $v(B)$, yields the causal effect of the intervention.

To estimate $v(B)$ for the effect of the transitions on the time series for the incidence rates of depression and bipolar disorder in Denmark, we went through the following steps:

1) Identified and estimated a tentative ARIMA model for N_t (including any seasonal components) using all weeks in the time series.

In this step, we went through the standard Box-Jenkins procedure to identify tentative values of p,d, and q (see e.g. Box and Jenkins (3) or Cryer and Chan (1)). We did this both manually and through the algorithm developed by Hyndman and Khandakar (4). We then inspected the time series for seasonality visually and with both the Canova-Hansen and the Osborn-Chui-Smith-Birchenhall tests. (5). Normally, one would restrict the time series being used in this step to periods before an intervention occurs since the deviation from the unperturbed process is zero in these periods, and they can therefore be used to specify the model for N_t. However, because the incidence rates in our data were repeatedly being affected by the biannual time transitions, we had no clear pre-intervention period on which to estimate a model for the unperturbed series. Instead, we opted to estimate this using the entire time series. This likely meant that our estimate of N_t was biased in the direction of the effect of the time transitions, which in turn meant that our estimates of the deviations from this process were biased towards zero. Because of this, the reported estimates of $v(B)$ probably represent a lower bound of the true effect.[**]

2) Specified intervention time and transfer function.

[**] An analogy would be a randomized experiment where, instead of comparing the mean of the treatment group to the mean of the control group, one compared the mean of the treatment group to the mean of all subjects in the study, both treatment and control. If a difference was found between the treatment group and the pooled group, this would mean that the difference between the treatment group and the control group would be greater.

In this next step, we specified a shape for the transfer function $v(B)$. We hypothesized that the effect of the time transitions on the incident rates would take the form of an immediate increase with a gradual decay back to the pre-transition level. This lead us to specify the transfer function as an autoregressive moving average model (ARMA) of order (1,0) with a pulse input. The pulse input variable took on a value of 1 in the first week after each transitions and a value of 0 in all other weeks. After estimation we inspected the residuals through the same Box-Jenkins methodology as in 1) in order to check whether the ARIMA model for N_t needed to be modified. This was not the case for any of the models.

3) Tested alternative transfer functions

After estimating the parameters of the specified transfer function we tried expanding it with further terms to determine if a better fit for the intervention effect existed. We evaluated specifications according to the AIC values for the model as a whole and the significance of the parameter estimates of the transfer function. A few of the more advanced specifications provided slightly lower AIC values than the (1,0) model, but their estimates of the immediate effect of the time transitions were very similar. Out of concerns for parsimony we therefore maintained the simple specification.

4) Calculated and graphed intervention effect

Using the estimates from the final model of step 3), we calculated the estimated effect of the intervention upon the values of y (see Table 2 in the manuscript). These were then graphed (see Figure 2 in the manuscript).

5) Carried out placebo tests

The last step in our approach went beyond what is proposed in Cryer and Chan (1). If the standard procedure yielded an intervention effect different from zero we carried out "placebo" tests for the time series in question. This was only the case for the transitions from summer time to standard time and the incidence rates of unipolar depression.

The reasoning behind these tests was simple. If the estimated intervention effect was truly due to the time transitions, we should not be able to find a larger number of similar effects if we placed the intervention times at different, randomly selected dates when no time transitions had occurred. Of course, because of pure chance, we should expect to find some significant effects elsewhere in the series, but the more numerous such false positives were, the less trustworthy the estimate of the effect of the true intervention would become.

Had conditions for these placebo interventions been controlled, we would have known exactly how many false positives we should on average expect to see. However, for some of these unrelated periods there might have been other,

occurrences that could have produced effects similar to those of the time transitions. Thus we could not say precisely how many false positives were too many, but we could say that the fewer there were the more certain we could be that the estimated effects for the true interventions did in fact represent its actual impact on the incidence rates.

The placebo tests were carried out by repeatedly estimating the parameters of the exact same transfer function identified above, but each time with different dates set for the interventions. For each run, we created a new vector for x_t where the 1's that corresponded to the occurrences of the time transitions were move a number of periods backward or forward in time. We started off with the 1's being placed 45 weeks before the actual transitions and then moved them five weeks forward every time we re-estimated the model. In the last run, the placebo interventions were located 45 weeks after the corresponding time transitions. We then graphed the estimated parameters along with their confidence intervals against the timing of the placebo interventions.

2. Modelling seasonality

The Canova-Hansen and the Osborn-Chui-Smith-Birchenhall tests showed no signs of seasonality for either of the time series. However, visual inspection of the series for unipolar depressive episodes indicated a slight correlation with the 52nd lag, consistent with weak yearly seasonality in depression. Since the DST transitions occured at the same times every year, they were themselves inherently seasonal phenomena, and ss a precaution, we therefore reran the models for depression while including a number of Fourier series[1] that served to capture the remaining seasonality (see e.g. http://robjhyndman.com/hyndsight/longseasonality/).

It is not clear from the literature how many pairs of Fourier terms should ideally be included. Here, we estimate models with one through four, supplemented by a model with the number of terms that result in the lowest possible AIC up to the maximum of half the seasonal length, i. e. $52.18/2 \approx 26$ (see e.g. http://robjhyndman.com/hyndsight/forecasting-weekly-data/). The results are reported in eTable 1.

[1] A Fourier series is essentially a sum of k pairs of cosine and sine functions. These functions have as their input the periodicity (52.18 in this case), time, and an integer $j \leq k$ that determines the frequency of the sine and cosine waves.
Each of these are terms are then multiplied by a coefficient. A Fourier series can be written as:

$$\sum_{j=1}^{k} \alpha_j sin\left(\frac{2\pi jt}{P}\right) + \beta_j cos\left(\frac{2\pi jt}{P}\right)$$

Where k is the total number of pairs, j is an integer determining frequency, t is time, P is the periodicity of the time series one seeks to model, and α and β are coefficients.

eTable 1: The effect of daylight saving time transitions, including *k* Fourier terms

	Dependent variable: log(unipolar depressive episodes)					
	k = 0 (original) ARIMA(1,1,1) (1)	*k* = 1 ARIMA(1,1,1) (1a)	*k* = 2 ARIMA(1,1,1) (1b)	*k* = 3 ARIMA(1,1,1) (1c)	*k* = 4 ARIMA(1,1,1) (1d)	*k* = 26 (lowest AIC) ARIMA(1,1,1) (1de)
Summertime (AR1)	0.006	0.027	0.769	0.065	0.042	-0.060
	(-2.214, 2.226)	(-2.438, 2.491)	(0.446, 1.091)	.	(-1.966, 2.049)	(-3.728, 3.608)
Summertime (MA0)	0.017	0.017	-0.049	-0.0003	0.013	0.005
	(-0.039, 0.072)	(-0.040, 0.074)	(-0.098, 0.001)	(-0.013, 0.013)	(-0.044, 0.070)	(-0.088, 0.098)
Wintertime (AR1)	0.756	0.709	0.869	0.997	0.998	0.997
	(0.637, 0.875)	(0.573, 0.845)	(0.615, 1.122)	(0.995, 0.999)	(0.996, 1.000)	(0.995, 1.000)
Wintertime(MA0)	0.108	0.098	0.073	0.112	0.094	0.108
	(0.065, 0.152)	(0.051, 0.145)	(0.025, 0.120)	(0.065, 0.160)	(0.045, 0.144)	(0.045, 0.170)
Obs.	939	939	939	939	939	939
Log-likelihood	660.05	662.34	683.9	690.04	692.34	843.22
AIC	-1308.10	-1308.69	-1347.79	-1356.08	-1356.69	-1570.44

The table contains the coefficients (and 95 % confidence intervals) of the moving average (MA0) and autoregressive model (AR1) terms from a series of transfer function models of order (1,0) with autoregressive integrated moving average (ARIMA) specifications of the unperturbed processes and a Fourier series of annual periodicity. Each MA0 term multiplied by 100% provides an approximate estimate of the immediate increase in the incidence rate after the time transitions. The AR1 term shows how much of the estimated increase in each week that persisted into the following week. The order of the ARIMA model for each series is given in the column header and the coefficients are omitted from the table body. The first number in parentheses following "ARIMA" gives the number of AR terms, the second gives the number of differences taken, and the third gives the number of MA terms. For each order of differencing an observation in the beginning of the time series is lost. The integer *k* gives the number of terms in each Fourier series. Note that for $k = 3$ the standard error for the AR1 term of the Summertime transitions could not be estimated.

In comparison with the original specification reported as Model 1 here and in Table 2 in the paper it is evident that neither of the five different Fourier series have much of an impact on the MA0 coefficients of the DST transitions to wintertime. Even when *k* equals 26 and AIC reaches its lowest, the coefficients remain virtually unchanged. The AR1 coefficients, estimating the decay, actually increase with the addition of the Fourier series, indicating a longer-lasting effect.

3. Meteorological conditions

While previous studies indicate that meteorological patterns are not associated with hospital admissions for bipolar disorder or depression (McWilliams et al. 2014), we re-estimate our models with the average daily level of precipitation, the average daily temperature, and the average daily minutes of sunlight as controls as an additional test of potential confounding from seasonal factors. Results are presented in eTable 2.

eTable 2: Controlling for weather and minutes of daylight

	Dependent variable: log(unipolar depressive episodes)			
	Original specification ARIMA(1,1,1) (1)	Precipitation ARIMA(1,1,1) (1a)	Temperature ARIMA(1,1,1) (1b)	Min. of daylight ARIMA(1,1,1) (1c)
Conditions	-	-0.003	-0.001	-0.0001
		(-0.007, 0.002)	(-0.002, 0.001)	(-0.0001, 0.00004)
Summertime (AR1)	0.006	0.004	-0.001	0.021
	(-2.214, 2.226)	(-2.219, 2.226)	(-2.487, 2.485)	(-2.366, 2.408)
Summertime (MA0)	0.017	0.015	0.014	0.017
	(-0.039, 0.072)	(-0.040, 0.070)	(-0.040, 0.069)	(-0.039, 0.074)
Wintertime (AR1)	0.756	0.754	0.734	0.709
	(0.637, 0.875)	(0.636, 0.871)	(0.610, 0.858)	(0.570, 0.848)
Wintertime(MA0)	0.108	0.111	0.107	0.098
	(0.065, 0.152)	(0.067, 0.155)	(0.062, 0.151)	(0.052, 0.145)
Obs.	939	939	939	939
Log-likelihood	660.05	660.8	650.15	662.08
AIC	-1308.10	-1307.59	-1286.29	-1310.17

The table contains the coefficients (and 95 % confidence intervals) of the moving average (MA0) and autoregressive model (AR1) terms from a series of transfer function models of order (1,0) with autoregressive integrated moving average (ARIMA) specifications of the unperturbed processes. Each MA0 term multiplied by 100% provides an approximate estimate of the immediate increase in the incidence rate after the time transitions. The AR1 term shows how much of the estimated increase in each week that persisted into the following week. The order of the ARIMA model for each series is given in the column header and the coefficients are omitted from the table body. The first number in parentheses following "ARIMA" gives the number of AR terms, the second gives the number of differences taken, and the third gives the number of MA terms. For each order of differencing an observation in the beginning of the time series is lost.

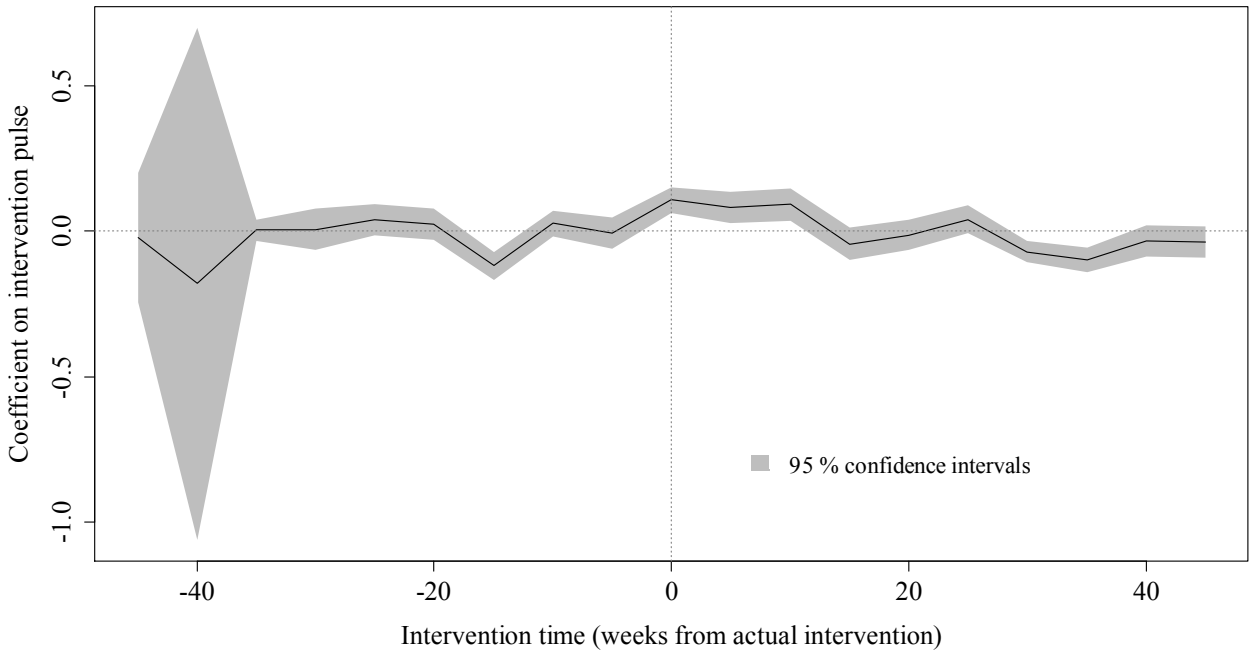
The "condtions" variable refers to either the average daily level of precipitation in millimeters, the average midday temperature in Celsius, or the average minutes of daylight.

As the table shows, neither the inclusion of precipitation levels, temperature, nor minutes of daylight affect the main results; both the immediate increase (MA0) and the decay (AR1) remain at their original levels. Further, none of the meteorological conditions are significantly related to the incidence rates of unipolar depressive episodes in themselves.

4. Results of placebo tests

Below, we plotted the coefficient of the MA0 term of the transfer function along with its 95% confidence interval against the timing of the interventions. The more of the confidence intervals for the placebo interventions that overlap zero, the more credible a causal interpretation of the coefficient of the true interventions is.

eFigure 1: Placebo Interventions



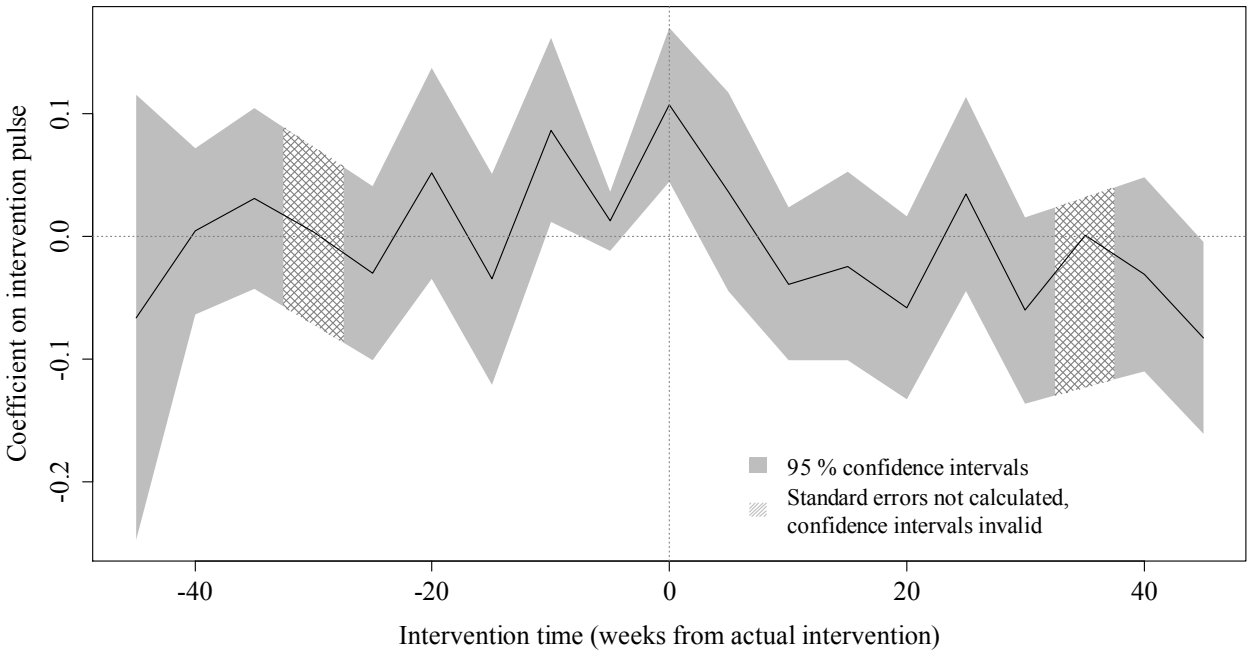
The figure plots the estimated MA0 coefficients from a series of transfer function intervention analysis models of the incidence rates of unipolar depression in Denmark between 1995 and 2012. The timing of the given vector of placebo interventions relative to the true time transitions is shown on the $\boldsymbol{x}$-axis. The coefficient for the true interventions is plotted at week zero. The grey polygon gives the 95% confidence intervals of the MA0 coefficient; if this overlaps the horizontal line at zero the coefficient in question is insignificant at the 0.05 level. For all models, the placebo test uses the same ARIMA specification for $\boldsymbol{N_t}$ as identified for the model of the true interventions.

The MA0 coefficient for the actual time transitions from summer time to standard time are plotted at week zero and the remaining points show the same coefficient for the placebo interventions. It is clear from the plot that only a few of the placebo interventions came out significant. Of these, just two were positive and none were of greater magnitude than the coefficient for the true intervention. The AR1 terms are not shown in the graph, but for the significant placebo interventions they were either very small or indistinguishable from zero, indicating that these jumps did not last very long.

The two positive placebo interventions fell five and ten weeks after the true intervention; in early December and early January, respectively. Even though the series were all differenced, meaning that they measured changes from week to

week, the significant increases in changes here could be consistent with seasonal patterns in depression. As noted above, the Osborn-Chui-Smith-Birchenhall and Canova-Hansen tests showed no signs of a seasonal unit root, but as a precaution we also reran the placebo tests with all 26 pairs of Fourier terms from the seasonal robustness test above. We plotted the results of these tests below:

efigure 2: Placebo Interventions, 26 Fourier terms



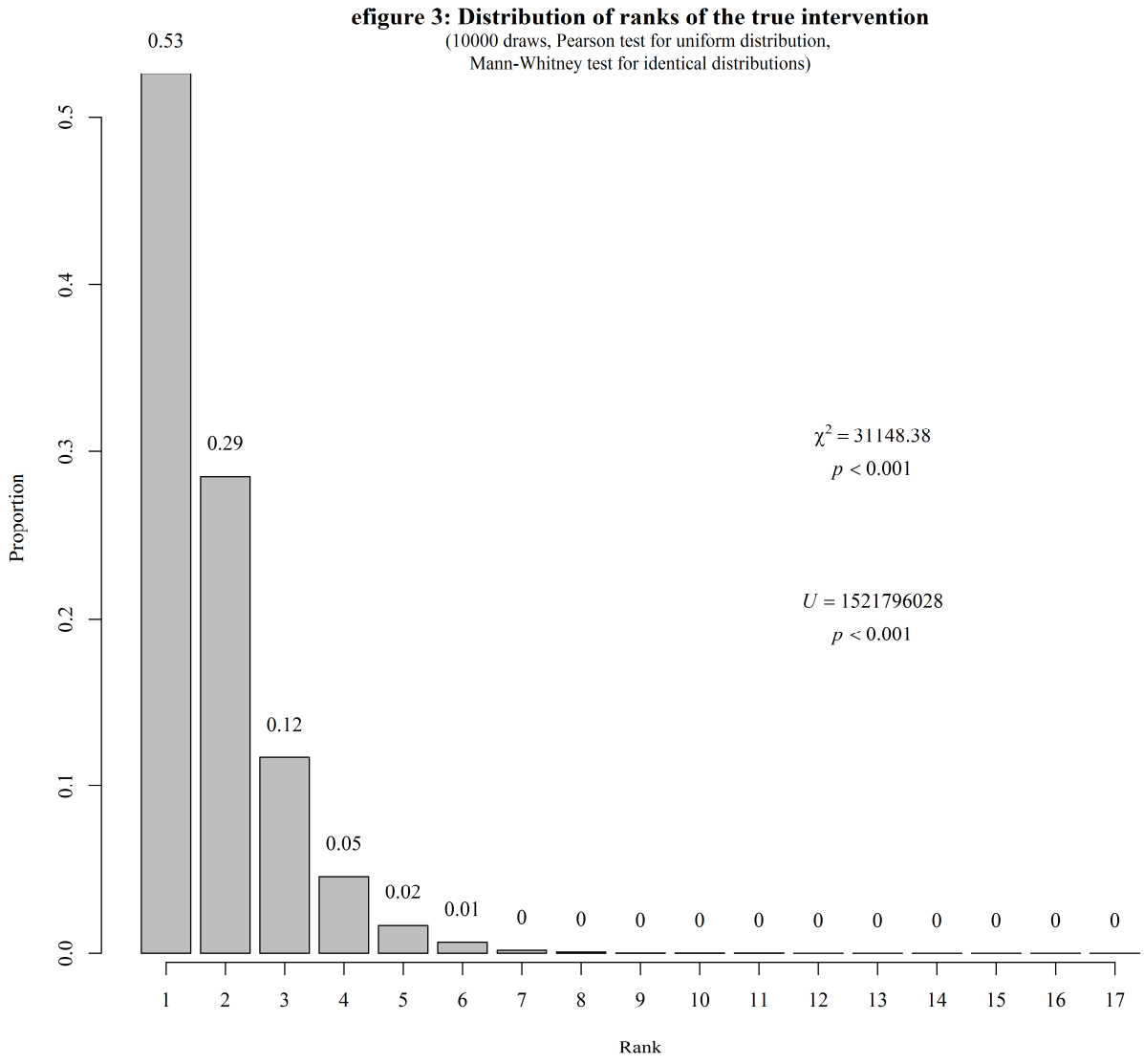
The figure plots the estimated MA0 coefficients from a series of transfer function intervention analysis models of the incidence rates of unipolar depressive episodes in Denmark between 1995 and 2012. The timing of the given vector of placebo interventions relative to the true time transitions is shown on the x-axis. The coefficient for the true interventions is plotted at week zero. The grey polygon gives the 95% confidence intervals of the MA0 coefficient; if this overlaps the horizontal line at zero the coefficient in question is insignificant at the 0.05 level. For all models, the placebo test uses the same ARIMA specification for N_t as identified for the model of the true interventions and includes 26 pairs of Fourier terms.

With the full set of Fourier terms in the model, the two placebo interventions five and ten weeks after the true intervention could no longer be distinguished from zero, while the true intervention remained at its previous level. This strongly indicated that these placebo interventions in the beginning of Winter were indeed a product of seasonality in the time series for unipolar depressive episodes, whereas the effect of the DST transitions was not.

Importantly, now only one placebo intervention resulted in a significant increase in the incidence rate, while another resulted in a barely significant decrease. Given that other events, taking place at other points in time over the course of the year, may also influence incidence rates, it is not clear how many significant interventions one should expect.

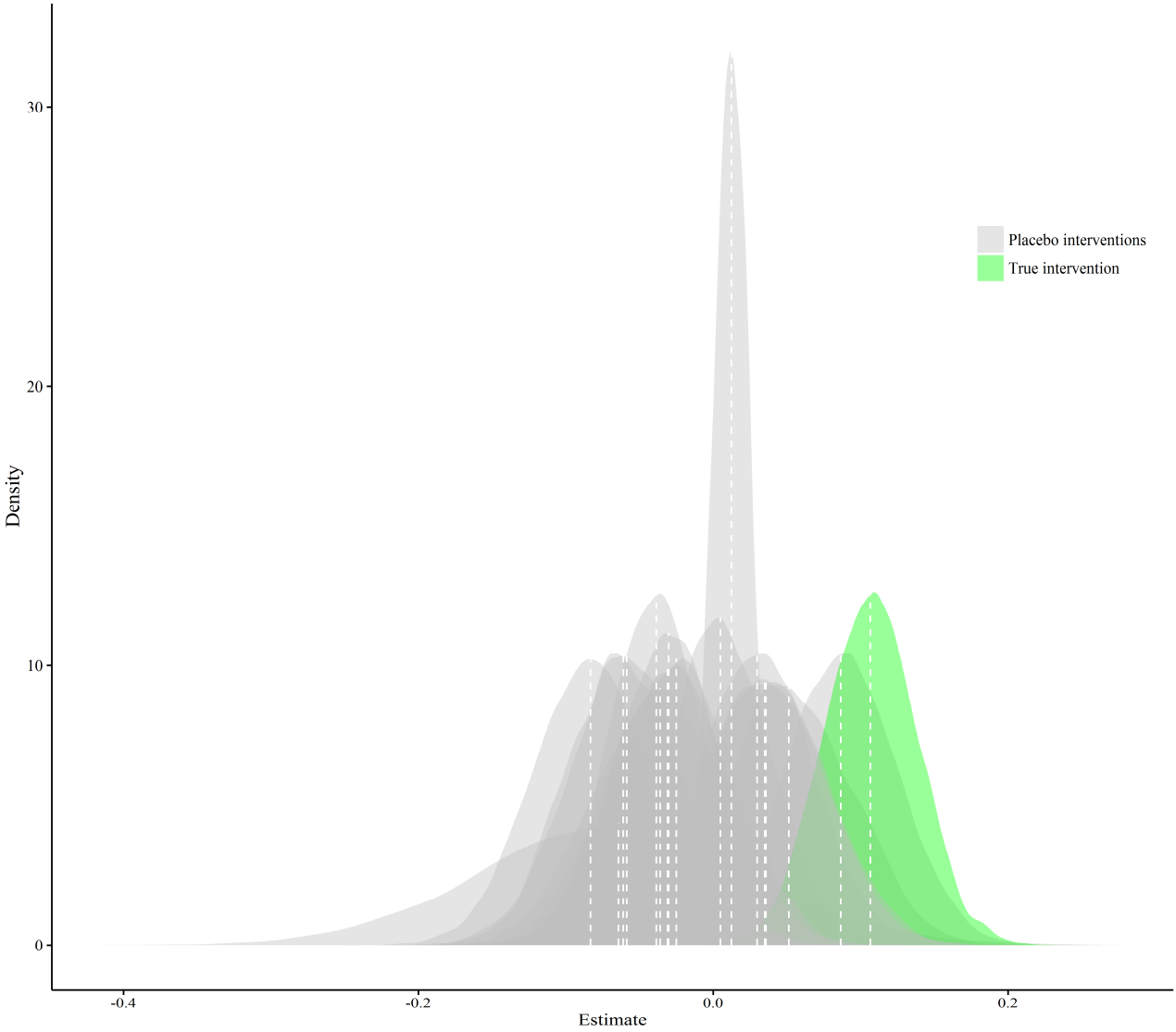
However, between one and two is very close to what we could expect due to pure chance ($1/18 \approx 0.056$ and $2/18 \approx 0.111$). This strongly indicated that the observe increases were indeed due to the DST transitions.

Above, we tested separately whether each of the MA0 estimates were different from zero. To get a better sense of how likely it was that this pattern had arisen due to chance, we also implemented a single test for the whole set of results. The test consisted of the following procedure: For each of the 17 interventions from efigure 2 for which standard errors could be calculated, we created a sampling distribution from 10,000 random draws from a Gaussian distribution with mean equal to the given MA0 estimate and standard deviation equal to its standard error. For each of the 10,000 draws, we then ranked the realizations of the estimates for the 17 interventions. Under the null hypotheses that an increase similar to the one observed at the true intervention was equally likely throughout the time series, the rank of the MA0 estimate for the true intervention should follow a uniform distribution, meaning that it should be largest estimate 1 out of 17 times, the second largest 1 out of 17 times, the third largest 1 out of 17 times, and so on. Below, we plotted the distribution of these ranks from the 10,000 draws .



As is evident from the plot, the ranks of the true intervention throughout the 10,000 draws did not at all follow a uniform distribution. We also tested this formally through a Pearson χ^2-test and a Mann-Whitney *U*-test. In the former, the null hypothesis stated that the observed frequency distribution was uniform and the alternative stated that it was not. In the latter, the null hypotheses stated that the probability that a draw from the sampling distribution for the true intervention was greater than a draw from the other sampling distributions was equal to 0.5, and the one-sided alternative stated that this probability exceeded 0.5. The p-values obtained for both tests were < 0.001. This could also be seen by plotting the simulated distributions for each of the interventions on top of each other.

efigure4: Simulated estimate distributions



The distribution for the true intervention, highlighted in green, was clearly the rightmost one, and a substantial part of it did not overlap at all with any of the distributions for the placebo interventions. Where it did overlap, it was primarily with the largest placebo intervention or the extreme right tail of some of the smaller placebo interventions. Consistent with the almost exclusively insignificant placebo interventions shown above, the plot also revealed that most of the distributions for the placebo interventions were clustered around zero.

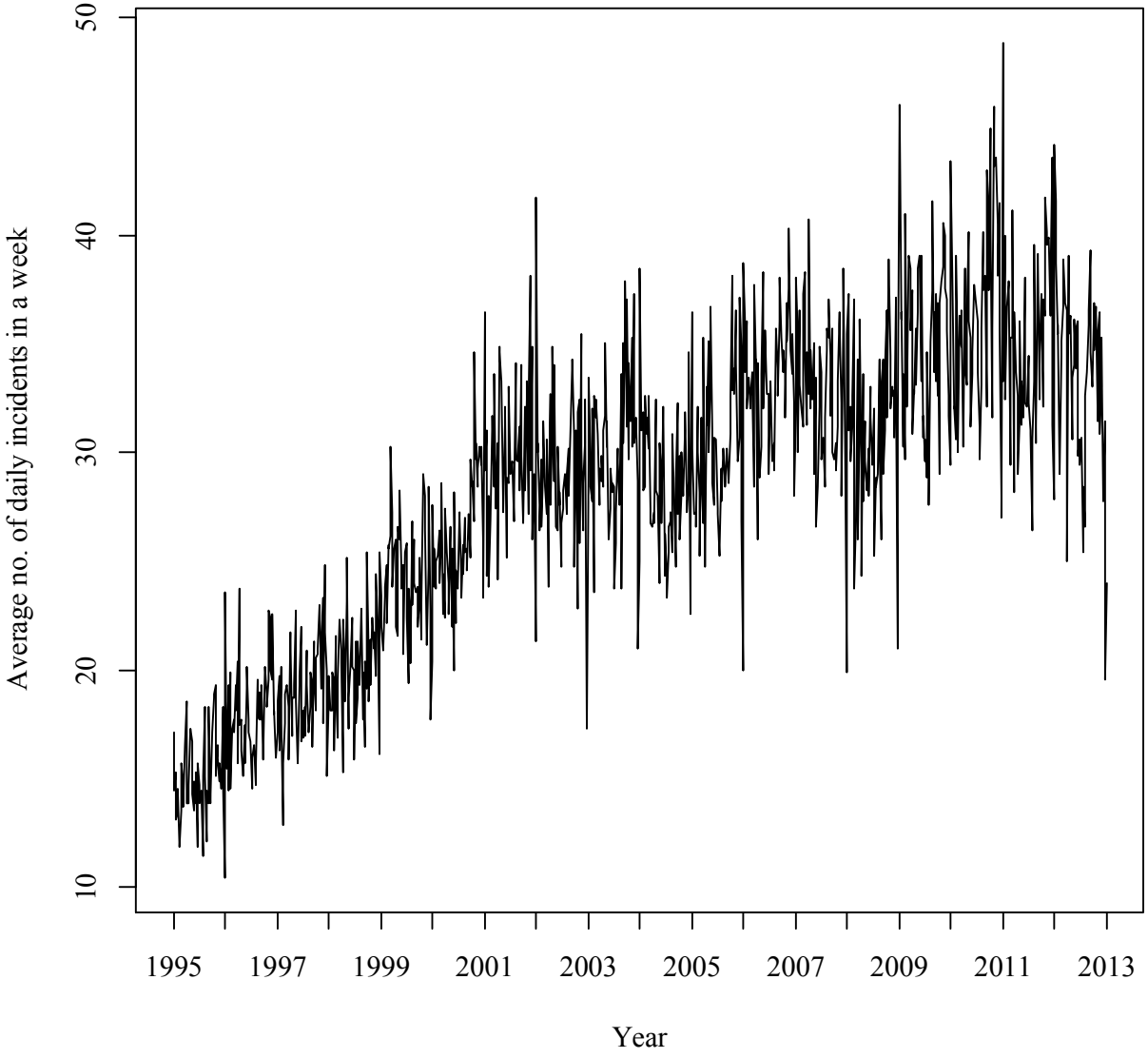
5. Dates of the DST transitions

eTable 3: DST transitions, Denmark, 1995-2012

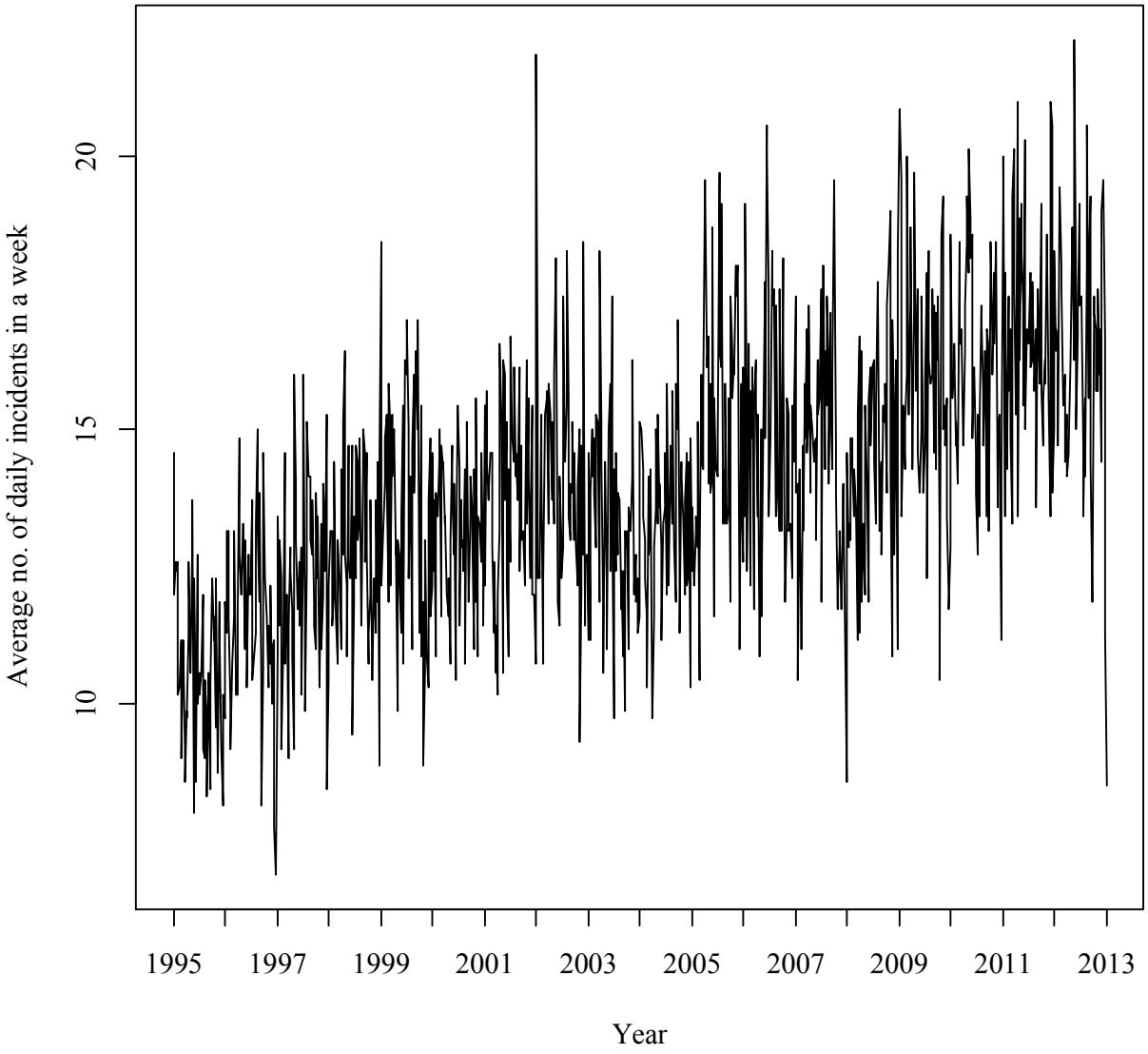
Year	DST	Standard time
1995	26. Mar. 02:00	24. Sep. 03:00
1996	31. Mar. 02:00	27. Oct. 03:00
1997	30. Mar. 02:00	26. Oct. 03:00
1998	29. Mar. 02:00	25. Oct. 03:00
1999	28. Mar. 02:00	31. Oct. 03:00
2000	26. Mar. 02:00	29. Oct. 03:00
2001	25. Mar. 02:00	28. Oct. 03:00
2002	24. Mar. 02:00	27. Oct. 03:00
2003	30. Mar. 02:00	26. Oct. 03:00
2004	28. Mar. 02:00	31. Oct. 03:00
2005	27. Mar. 02:00	30. Oct. 03:00
2006	26. Mar. 02:00	29. Oct. 03:00
2007	25. Mar. 02:00	28. Oct. 03:00
2008	30. Mar. 02:00	26. Oct. 03:00
2009	29. Mar. 02:00	25. Oct. 03:00
2010	28. Mar. 02:00	31. Oct. 03:00
2011	27. Mar. 02:00	30. Oct. 03:00
2012	25. Mar. 02:00	28. Oct. 03:00

6. Plots of the time series for the incidence rates of unipolar depressive episodes and bipolar disorder

eFigure 3: Incidence rate for major depressive episodes, 1995-2012



eFigure 4: Incidence rate for bipolar disorder, 1995-2012



7. Data, software and code

Replication data (the weekly incidents rates) as well as replication code is available in eAppendix 2 and eAppendix 3 respectively. The analyses were done using R version 3.2.2 (2015-08-14)

Variables in the replication dataset:

week: The week number. The count starts in the first week of January, 1995.

depress_count: The average daily number of contacts in the given week that resulted in a diagnosis of a unipolar depressive episode.

manic_bipolar_count: The average daily number of contacts in the given week that resulted in a diagnosis of bipolar disorder.

from_clock_change_week: A binary variable that takes on a value of 1 in the first week after each transition to summertime.

to_clock_change_week: A binary variable that takes on a value of 1 in the first week after each transition to wintertime.

any_clock_change_week: A binary variable that takes on a value of 1 if either *from_clock_change_week* or *to_clock_change_week* is equal to 1.

precipitation: The average daily level of precipitation in the given week measured in millimeters.

temperature: The average daily temperature in the given week measured in degrees Celsius.

daylight: The average daily minutes of daylight in the given week.

8. References

1. Cryer J, Chan K. Time Series Analysis: With Applications in R. New York, NY: Springer; 2010.

2. Box GE, Tiao GC. Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical association 1975;70(349):70-79.

3. Box GEP, Jenkins GM. Time series analysis, forecasting and control. Holden-Day, San Francisco 1970.

4. Hyndman, RJ. Forecasting with long seasonal periods. 2010. Retrieved February 26, 2016, from http://robjhyndman.com/hyndsight/longseasonality.

5. Hyndman, RJ. Forecasting weekly data. 2014. Retrieved February 23, 2016, from http://robjhyndman.com/hyndsight/forecasting-weekly-data/.

6. Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 2008;27(3).

7. Canova F, Hansen BE. Are seasonal patterns constant over time? A test for seasonal stability. Journal of Business & Economic Statistics 1995;13(3):237-252.