medRxiv preprint doi: https://doi.org/10.1101/2022.08.03.22278371; this version posted August 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Sample size bias in the empirical assessment of the acute risks associated with Daylight Saving Time transitions*

José María Martín-Olalla

Universidad de Sevilla,
Facultad de Física,
Departamento de Física de la Materia Condensada,
ES41012 Sevilla,
Spain†

Jorge Mira

Universidade de Santiago de Compostela,
Facultade de Física,
Departamento de Física Aplicada and iMATUS,
ES15782 Santiago de Compostela,
Spain‡

(Dated: August 3, 2022)

The assessment of the acute impact of Daylight Saving Time (DST) transitions is a question of great interest for an understanding of the benefits and inconveniences of a practice that is now under public scrutiny in Europe and America.
Here we report a thorough analysis of a record of twelve well-known research studies that reported increased risks associated with DST transitions in health issues —acute myocardial infarction, ischemic strokes— and in societal issues —accidents, traffic accidents and fatal motor vehicle accidents—.
We found that a 5 % increase of the risks suffices to understand the reported increased risks associated with the spring transition. Reported values above this threshold are impacted by the sample size of the study.
In the case of the autumn transition, no increase of the risks is found.

Keywords: DST; summer time; latitude; sleep deprivation; spring transition; Europe; insolation; season; motor vehicle accidents

Sample size N is the best-known study limitation of any research based on a sampled observation. Any finite sample size translates in a limited knowledge of the parent population that may have originated the observation. This limitation is essentially contained in the size of the confidence interval associated with the reported point estimate[1]. Researchers push hard to study large samples in the hope of a better determination of the effect. Nonetheless it is the case that the same phenomenon is analyzed by vividly different sample sized studies, which gives rise to a distribution of sampled results. Because every research study is analyzed on a solo basis, it is infrequent that the sample of reported point estimates is analyzed consistently in relation to the distribution of the sample sizes.

We bring here a comprehensive analysis of the research studies on the acute effects that Daylight Saving Time (DST) transitions may cause on public health and on societal issues. DST is the seasonal biannual changing of the clocks, a long-standing practice by which modern, extratropical societies —chiefly in America and Europe— adapt the phase of their daily rhythms to the early summer sunrises and to the late winter sunrises. DST transitions set a natural experiment in which every individual participates. Yet when it comes to a research study that assesses the increased risks associated with the practice, things are further limited: it is only one health —myocardial infarction[2; 3; 4; 5; 6; 7; 8] or ischemic strokes[9]— or one societal —accidents[10; 11; 12; 13]— issue that is analyzed in a limited region —a country[2; 3; 7; 9; 10; 11; 12; 13], some region[6; 8], or some hospital[4; 5]— and during a limited period of time —one year[9; 10; 11], a few years[4; 5; 6; 7], one or two decades[2; 3; 8; 12; 13]—.

All these factors impact the sample size of the study that ranges from five to 14 000 counts, whereas the point estimate of the effect ranges from a factor 0.8 to a factor 1.9 —where no effect is associated with 1—. A selection bias usually gives more importance to larger estimations since they might unveil larger hazards[14]. Eventually, and

* CC BY NC ND

† olalla@us.es; https://orcid.org/0000-0002-3750-9113; https://ror.org/03yxnpp24
‡ jorge.mira@usc.es; https://orcid.org/0000-0002-6024-6294; https://ror.org/030eybx10

Typeset by REVTEX
NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.03.22278371; this version posted August 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

in view of these results, governments in Europe and America are considering discontinuing this practice in the fear that its impact on human health may be more severe than previously thought.

Here our comprehensive analysis shows that the parent population that may have originated the distribution of incidence ratios in this set of twelve well-known research studies shows an increase of the risk around 5 %. Any reported IR above this threshold is impacted by the sample size of the study.

METHODS

The point estimate of the acute impact of DST transitions on human health is primordially assessed by the incidence ratio $\mathrm{IR} = O/N$: the ratio of the observed counts O in a societal issue after a transition —also known as the study group— to the expected, contra-factual, counts N in the same societal issue —the control group, and hereafter the sample size of the study—.

We consider seven acute myocardial infarction (AMI) studies which reported O in the first seven days following the spring and autumn transitions and deduced N from various estimates[2; 3; 4; 5; 6; 7] or reported IR for the first seven days after a transition[8]. We also considered one ischemic stroke study[9] reporting IR in the first seven days after a transition. Finally we considered a series of four accident-related studies: accidental deaths[10], traffic accidents[11], road accidents[12] and fatal motor vehicle accidents[13]. All of them reported O and N or only IR in the first few days after a transition.

We analyze three sets of IR: day IR associated with any of the first seven days after a transition —there are $n = 64$ values of this kind—, week IR for the whole first week after a transition —$n = 9$— and a set of stratified IR[3; 4; 7; 8] associated with patient characteristics and co–diagnoses —among others: gender, age group, cholesterol; $n = 78$— in the first week.

Supporting information shows a summary of week IR (when available) and day IR (otherwise) and their corresponding sample size.

On the other hand, exact confidence intervals (CI) for Poisson distributed events can be obtained as a function of N from the quantiles of the χ^2 distribution for $2N \cdot \mathrm{IR}_t$ degrees of freedom (low bound) and $2 \cdot (N \cdot \mathrm{IR}_t + 1)$ degrees of freedom (high bound)[15]. Here IR_t is the incidence ratio of the parent population that may have originated the distribution of reported IRs. It is an adjustable parameter and we search for the IR_t whose 95 %CI contains the 95 % of the reported IRs, given the sample size of the study.

RESULTS

Fig. 1 shows the reported IRs versus the sample size of the study. The left panel shows the results for the spring transition; the right panel, those of the autumn transition. Generally speaking day IRs (intermediate ink) have smaller sample sizes than week IRs (darkest ink) and show a much larger variability.

Fig. 1 left shows in broken lines the 95 %CI for $\mathrm{IR}_t^{\mathrm{spring}} = 1.00$, Table 1 summarizes the occurrences of reported IR inside the 95 %CI. In the first row of the table we see 59 (stratum, 75.6 %), 47 (day, 73.4 %) and 5 (week, 55.6 %) reported IRs inside the 95 %CI for $\mathrm{IR}_t^{\mathrm{spring}} = 1.00$. Therefore, the null hypothesis $\mathrm{IR}_t^{\mathrm{spring}} = 1.00$ does not sustain at the standard level of significance (5 %).

In contrast the null hypothesis $\mathrm{IR}_t^{\mathrm{spring}} = 1.05$ —whose 95 %CI is shown by solid lines— yields the largest occurrences inside the CI: 72 (92.3 %), 60 (93.8 %) and 9 (100.0 %) of the stratum, day and week observations (see second row in Table 1). Therefore the null sustains at the standard level of significance. A similar analysis for the autumn transition shows that the null hypothesis $\mathrm{IR}_t^{\mathrm{autumn}} = 1.00$ and $\mathrm{IR}_t^{\mathrm{autumn}} = 0.99$ sustains with similar scores.

As a result we understand that for the spring transition the incidence ratio is close to 1.05 or 5 % increase of the risks, while the autumn transition poses negligible small risks. Therefore, reported IRs above $\mathrm{IR}_t^{\mathrm{spring}} = 1.05$ —that often raises the alarm in decision makers, the public opinion, and researchers— are likely impacted by the limitation that the sample size always brings.

DISCUSSION

To understand the size of the spring effect we were able to estimate the relative sample standard deviation of the record in four studies[4; 6; 12; 13]. Figure 2 in [6] shows the residuals of the model on a daily basis. Visual perception suggests that the standard error of the model is at least some 20 % of the daily AMIs, the largest reported IR is

medRxiv preprint doi: https://doi.org/10.1101/2022.08.03.22278371; this version posted August 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

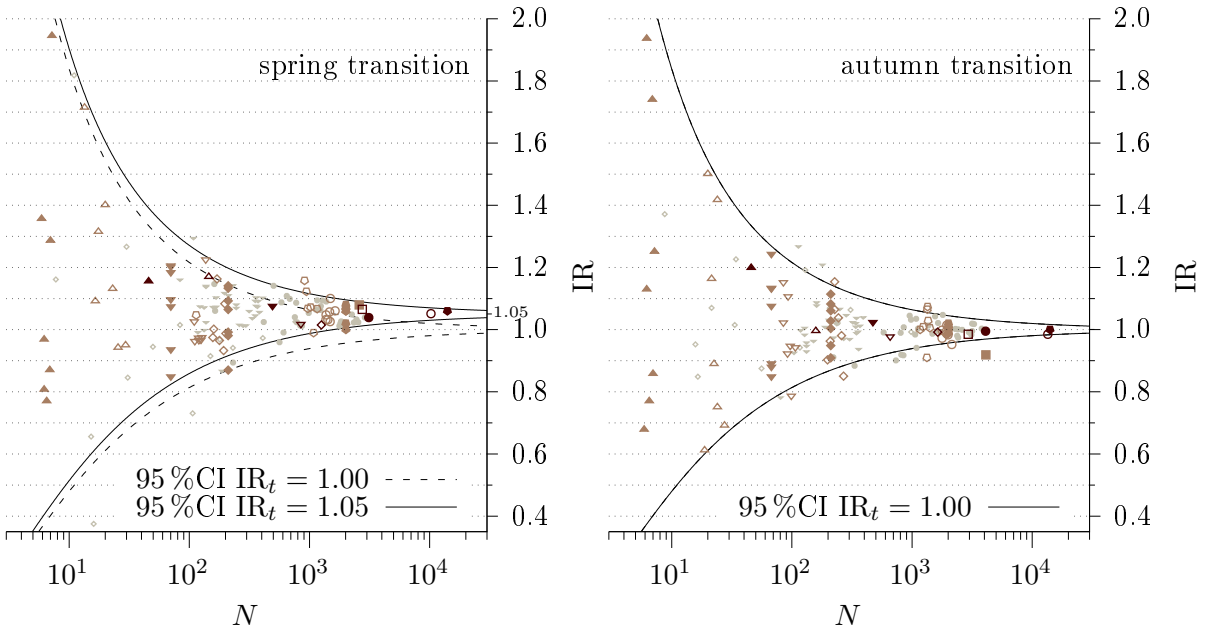

Figure 1 Scatter plot for stratified (lightest ink), day (intermediate ink) and week (darkest ink) IR associated with the spring transition (left panel) and the autumn transition (right panel). The 95 % confidence intervals for selected IR$_t$ are shown. Legend: open circles[2], solid circles[3], open up triangles[4], solid up triangles[5], open down triangles[6], open diamonds[7], solid down triangles[8], solid diamonds[9], open squares[10], solid squares[11], open pentagons[12], solid pentagons[13]. See Table 1 for a breakdown of occurrences inside the 95 % confidence intervals.

in line with this number. Table 2b in [12] provides the standard error of the model as 0.069 in the $\log_{10}$ space or $10^{0.069} = 1.17$ in the linear space: their largest reported IR is in line with this number. Finally Figure 1 in [13] shows the sample standard deviation of the weekly observation in two periods of eleven years each. From this we infer a relative sample standard deviation $\sim 8\,\%$ in the week previous to the spring transition. Their reported IR is below this threshold. Table 1 in [4] contains N_i and O_i in the seven years of the record, from which we found a relative standard deviation of the sample $\sim 40\,\%$ for one week. In contrast [4] reported a week IR = 1.17 or 17 % increase in the week after the spring transition.

From these four studies —$N \in (150, 14\,000)$— we understand that, while the spring transition shows a systematic increase of the risks, the size of the excursion is below the sample relative standard deviation of O. Therefore the impact of the transitions is below the myriad of things that populates the variability of the tested quantity, which explains why the assessment of the impact of the clock transitions on health issues is elusive.

CONCLUDING REMARKS

We do not conclude from our analysis that DST transitions do not impact on public health. Instead, we bring attention to the fact that its impact might be as mild as previously thought. Decision makers and researchers should understand that the assessment of the IR is only the starting point of the balance of the risks associated with the practice. They should be balanced out against the risks of canceling a practice. A retrospective analysis showed that DST has reduced the public health risks associated with the onset of human activity in the early hours of the winter mornings.[16]

ACKNOWLEDGMENTS

The authors thank to Prof. Mercedes Conde Amboage from Universidade de Santiago de Compostela for fruitful discussion.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.03.22278371; this version posted August 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

IR_t	**Stratified**	**Day**	**Week**
		spring	
1.00	59/78 (75.6 %)	47/64 (73.4 %)	5/9 (55.6 %)
1.05	72/78 (92.3 %)	60/64 (93.8 %)	9/9 (100.0 %)
		autumn	
1.00	67/78 (85.9 %)	56/64 (87.5 %)	9/9 (100.0 %)

Table 1 Occurrences inside the 95 % confidence interval for selected IR_t (see Fig. 1 for a graphical representation). First the number of occurrences relative to the total number of observations; then, in parentheses, the percentage score. The null hypothesis $\mathrm{IR} = \mathrm{IR}_t$ sustains at the standard level of significance for $\mathrm{IR}_t^{\mathrm{spring}} = 1.05$ and $\mathrm{IR}_t^{\mathrm{autumn}} = 1.00$.

REFERENCES

[1] V. Amrhein, S. Greenland, and B. McShane, Nature **567**, 305 (2019).
[2] I. Janszky and R. Ljung, New England Journal of Medicine **359**, 1966 (2009).
[3] I. Janszky, S. Ahnve, R. Ljung, K. J. Mukamal, S. Gautam, L. Wallentin, and U. Stenestrand, Sleep Medicine **13**, 237 (2012).
[4] M. R. Jiddou, M. Pica, J. Boura, L. Qu, and B. A. Franklin, The American Journal of Cardiology **111**, 631 (2013).
[5] V. Čulić, Chronobiology International **30**, 662 (2013).
[6] A. Sandhu, M. Seth, and H. S. Gurm, Open Heart **1**, e000019 (2014).
[7] J. O. Sipilä, P. Rautava, and V. Kytö, Annals of Medicine **48**, 10 (2015).
[8] I. Kirchberger, K. Wolf, M. Heier, B. Kuch, W. Von Scheidt, A. Peters, and C. Meisinger, BMC Public Health **15**, 1 (2015).
[9] J. O. Sipilä, J. O. Ruuskanen, P. Rautava, and V. Kytö, Sleep Medicine **27-28**, 20 (2016).
[10] S. Coren, Perceptual and Motor Skills **85**, 921 (1996).
[11] S. Coren, New England Journal of Medicine **334**, 924 (1996).
[12] D. Robb and T. Barnes, Accident Analysis and Prevention **111**, 193 (2018).
[13] J. Fritz, T. VoPham, K. P. Wright, and C. Vetter, Current Biology **30**, 729 (2020).
[14] S. Marek, B. Tervo-Clemmens, F. J. Calabro, D. F. Montez, B. P. Kay, A. S. Hatoum, M. R. Donohue, W. Foran, R. L. Miller, T. J. Hendrickson, S. M. Malone, S. Kandala, E. Feczko, O. Miranda-Dominguez, A. M. Graham, E. A. Earl, A. J. Perrone, M. Cordova, O. Doyle, L. A. Moore, G. M. Conan, J. Uriarte, K. Snider, B. J. Lynch, J. C. Wilgenbusch, T. Pengo, A. Tam, J. Chen, D. J. Newbold, A. Zheng, N. A. Seider, A. N. Van, A. Metoki, R. J. Chauvin, T. O. Laumann, D. J. Greene, S. E. Petersen, H. Garavan, W. K. Thompson, T. E. Nichols, B. T. Yeo, D. M. Barch, B. Luna, D. A. Fair, and N. U. Dosenbach, Nature **603**, 654 (2022).
[15] J. Sun, Y. Ono, and Y. Takeuchi, Journal of Occupational Health **38**, 196 (1996).
[16] J. M. Martín-Olalla, Chronobiology International **39**, 1 (2022).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.03.22278371; this version posted August 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

SUPPORTING INFORMATION

Supporting Information

Study	Region	Spring			Autumn		
		Years	N	IR [95 %CI]	Years	N	IR [95 %CI]
Acute myocardial infarction							
Ref. [2]	Sweden	15	10 251.0	1.05 [1.03, 1.07]	20	13 492.0	0.99 [0.97, 1.00]
Ref. [3]	Sweden	10	3115.5	1.04 [1.00, 1.08]	13	4095.5	0.99 [0.96, 1.03]
Ref. [4]	Two hospitals (Michigan, USA)	6	145.4	1.17 [1.00, 1.36]	6	158.3	0.99 [0.85, 1.16]
Ref. [5]	One hospital (Croatia)	6	45.9	1.15 [0.86, 1.51]	6	45.9	1.20 [0.90, 1.56]
Ref. [6]	Michigan (USA)	4	849.0	1.02 [0.95, 1.09]	4	658.0	0.98 [0.90, 1.06]
Ref. [7]	Finland	7	1257.2	1.01 [0.96, 1.07]	9	1633.2	0.99 [0.94, 1.04]
Ref. [8]‡	Ausburg (Germany)	26	491.9	1.08 [0.97, 1.25]	25	473.0	1.02 [0.93, 1.13]
Ischemic strokes							
Ref. [9]†‡	Finland	1	210.7	1.14 [1.00, 1.30]	1	210.7	1.11 [0.98, 1.25]
Accidents							
Ref. [10]	USA	1	2750.0	1.07 [1.03, 1.10]	1	2938.5	0.98 [0.95, 1.02]
Ref. [11]†	Canada	1	2594.0	1.08 [1.04, 1.12]	1	4111.0	0.92 [0.89, 0.95]
Ref. [12]†	New Zealand	11	910.0	1.16 [1.02, 1.28]	12	1346.0	1.07 [0.94, 1.16]
Ref. [13]‡	USA	22	14 044.8	1.06 [1.03, 1.09]	22	14 044.8	1.00 [0.98, 1.03]

Table 2 The sample size N —the total number of events in the control week, the expected total number of events from a model or the average number of events in a week— and the incidence ratio IR associated with the week after the spring transition and the autumn transition. Values were obtained for the week after a transition, except when marked with †, in which case the largest IR reported in the first week after a transition is shown. Studies marked with ‡ did not provide N, we deduced them from an average of the record. Blue ink annotates $p-$values below the standard level of significance $\alpha = 0.05$. According to their N, the reported values can be located inside the 95 %CI for $\mathrm{IR}_t = 1.05$ (spring), $\mathrm{IR}_t = 1$ (autumn), see Figure 1 and Table 1.