The Effect of Electrical Load Shedding on Pediatric Hospital Admissions in South Africa

Christian Gehringer,[a,b] *Heinz Rode,*[b] *and Michael Schomaker*[c]

Background: South Africa faced repeated episodes of temporary power shutdowns, or load shedding, in 2014/2015. The effect of load shedding on children's health is unknown.

Methods: We determined periods of load shedding using Twitter, Facebook, and data from the City of Cape Town. We obtained the number of unscheduled hospital admissions between June 2014 and May 2015 from Red Cross Children's Hospital, Cape Town, and weather data from the South African Weather Service. We used quasi-Poisson regression models to explore the relationship between number of hospital admissions and load shedding, adjusted for season, weather, and past admissions. Based on assumptions about the causal process leading to hospital admissions, we estimated the average treatment effect, that is, the difference in expected number of admissions per day had there been load shedding each day or on any of the preceding 2 days compared with if there had not been any load shedding.

Results: We found a 10% increase (95% confidence interval: 4%, 15%) in hospital admissions for days where load shedding was experienced on the same day, or no more than 2 days prior, compared with when there was no load shedding in the past 2 days. The increase was more pronounced during weekdays (12% [7%, 18%] vs. 1% [-9%, 11%]), and for specific diagnoses (e.g., respiratory system: 14% [2%, 26%]). The average treatment effect was estimated as 6.50 (5.12, 7.87) highlighting that about 6 additional admissions a day could be attributed to load shedding.

Conclusions: The association we measured is consistent with our hypothesis that failures of the power infrastructure increase risk to children's health. See video abstract at, http://links.lww.com/EDE/B409.

Submitted December 6, 2017; accepted July 26, 2018.

From the [a]Division of Internal Medicine, University Hospital Basel, University of Basel, Basel, Switzerland; [b]Red Cross War Memorial Children's Hospital, Cape Town, South Africa; and [c]Centre for Infectious Disease Epidemiology and Research, University of Cape Town, Cape Town, South Africa.

Supported by the US National Institute of Allergy and Infectious Diseases through the International epidemiological Databases to Evaluate AIDS, Southern Africa (IeDEA-SA), grant 5U01AI069924-05.

The authors report no conflicts of interest.

Availability of code and data: The R-code for this analysis is part of the supplementary material; http://links.lww.com/EDE/B389. The data are available from the authors upon request to verify the results.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Michael Schomaker, Centre for Infectious Disease Epidemiology and Research, University of Cape Town, Anzio Road, Observatory 7925, Cape Town, South Africa. E-mail: michael.schomaker@uct.ac.za.

ISSN: 1044-3983/18/2906-0841
DOI: 10.1097/EDE.0000000000000905

Keywords: Load shedding, Power failure, Pediatrics, Causal inference, TMLE

(*Epidemiology* 2018;29: 841–847)

The Republic of South Africa faced repeated episodes of temporary power shutdowns in 2014–2015. Owing to its inability to satisfy the power demand (because of loss of power generation) and to prevent uncontrolled blackouts, the monopoly power supplier ESKOM implemented this practice, which is also known as rotational load shedding, for several hours a day in most of the country. Load shedding is an intervention of last resort when power demand exceeds supply: times and areas affected by load shedding have been communicated by ESKOM to the public on short notice, for example, via schedules published on Twitter and different dedicated homepages.

Even though electricity is the main power source for heating, cooking, and lighting throughout South Africa,[1] the consequences of load shedding are predominately discussed with regard to its economic implications, probably because of South Africa's challenging economic situation.[2] Unfortunately, surprisingly little information can be found about health-related implications and costs. This is worrying, as case reports from hospitals suggest a direct link between blackouts and health outcomes, such as an increased burden on already overworked staff, for example, during surgeries.[3]

Failures of the electrical infrastructures are known to have increased hospital admissions, health-related complications, and mortality during both the "Northeast blackout" of 2003 in the United States and Canada and a power blackout in Italy, Europe, the same year.[4–9] Reasons for increased admissions included carbon-monoxide intoxications because of the use of portable generators,[7] more emergencies owing to failure of electrical medical devices,[6] more domestic accidents,[5] and a higher rate of food poisoning.[7] Other studies investigated natural disasters and extreme weather conditions, which were accompanied by power failures and affected the health of the respective population by an increased number of emergency presentations, carbon monoxide poisoning, among other reasons.[10–13] Although spontaneous power shutdowns as described above, and repeated power shutdowns during load shedding in South Africa have different causes, the implications, that is,

lack of electricity are very much the same. One may therefore assume similar underlying etiological mechanisms. Nevertheless, to the best of our knowledge, no study has yet investigated the health effects of load shedding, which differs from unexpected power failures in the sense that people can partly adapt their lives around load shedding schedules and face shorter durations of no electricity supply. The effect of load shedding on health outcomes is of particular interest in a developing country like South Africa, where resources are scarce, electricity is the main source of heating and cooking, and health care facilities are often overburdened.[14]

This study analyzes the effect of load shedding on admissions to the Red Cross War Memorial Children's Hospital (RCCH), a 300-bed tertiary pediatric hospital in Cape Town.

METHODS

Study Design

This is a retrospective, single-centre observational study analyzing the relationship between unscheduled hospital admissions to the RCCH and load shedding in the period between 1 June 2014 and 31 May 2015.

Study Population and Number of Admissions

The RCCH catchment area is the city of Cape Town except for specialty consultations, for example, burns, from the whole Western Cape and beyond.

The number of direct admissions of children up to 13 years of age, excluding internal transfers, was used as primary outcome in this analysis. Planned admissions, that is, patients with appointments, were excluded from the analysis. We thus considered unplanned admissions—for example, owing to emergencies, external transfers for specialized care (e.g., burns, surgeries, and intensive care), and presentations because of proximity or personal experience—as the main quantity of interest. We used International Statistical Classification of Diseases and Related Health Problems (ICD) codes to group patients according to their leading diagnosis, and the responsible speciality, that is, medicine or surgery.

Load Shedding as Documented on Twitter and Facebook

Our primary exposure variable is binary and indicates whether load shedding was implemented on the respective day (or preceding days, see below). In secondary analyses, which are descriptive and spatial in nature, the exposure relates to the event that load shedding was implemented in a specific area.

The authors identified days of load shedding from the available Twitter (San Francisco, CA) tweets from the ESKOM account @Eskom_SA (accessed on 12/05/2015 and 14/06/2015) and cross-checked them with Facebook (Menlo Park, CA) entries (https://www.facebook.com/EskomSouthAfrica/), documenting the respective day of an event but neglecting the exact time span (usually between 6 AM and 10 PM) and the outage severity because those details were only inconsistently reported and information quality differed by area and time. This information was then validated, and also updated for 9 days, by using data from the electricity generation and distribution department of the City of Cape Town, which also provided information on the length and area of load shedding (for those areas which were under direct control of the city).

Weather and Other Potential Confounders

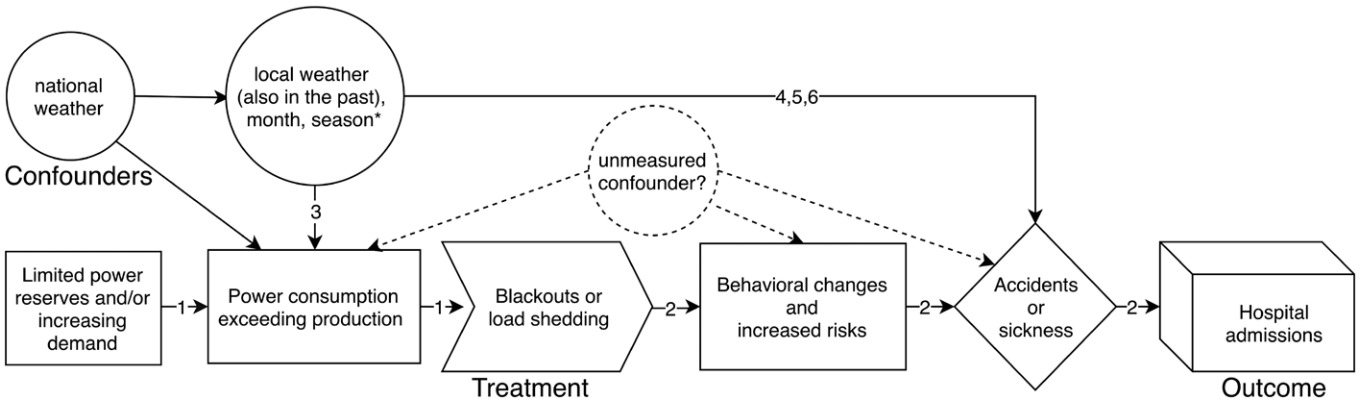
Weather data, identified as a possible confounder (see below), was obtained from the South African Weather Service. Relative humidity (in %), pressure (in hectopascal), precipitation (rainfall in mm), temperature (in degrees Celsius), and wind speeds (in meter/second) were obtained for the five weather stations in Cape Town: Cape Town airport, South African Astronomical Observatory, Royal Yacht Club, Molteno Reservoir, and Kirstenbosch. Sunshine (hours per day) was only measured at Cape Town airport. We defined the arithmetic mean of the measurements of Cape Town Airport, the Observatory, and Molteno Reservoir as our weather indicators, based on the proximity to the catchment area of the RCCH. Kirstenbosch was excluded because it lies on the slopes of a mountain and has weather conditions that are not representative for the rest of Cape Town (see eFigure 1; http://links.lww.com/EDE/B390 for smoothed weather data for different stations). The Yacht Club was excluded as well because humidity and temperature measurements were missing for 119 consecutive days and the sea climate may not perfectly resemble the weather conditions of the study population.

We further identified seasonal trends as another potential measured confounder.

Statistical Analysis

We used kernel density plots to look at the distribution of hospital admissions depending on whether load shedding occurred and if it was a weekday. We used quasi-Poisson regression models to explore the relationship between the number of hospital admissions and load shedding. We considered month, weather, last week of the month (pay week), a weekly trend modeled with sine and cosine terms, past weather indicators up to a lag of 2 days, and past admissions up to a lag of 28 days to be potential confounders or relevant to model the admission process. The quasi-Poisson model can be interpreted as any other Poisson model, but allows the variance of the model to be different from the mean, and can therefore deal with overdispersion, that is, greater variability in the data than expected under the specified model. All continuous variables were included nonlinearly in the model using p-splines.[15] In the main model, load shedding is a binary variable, which indicates if there was a load shedding event on the same day or up to 2 days prior of the day of interest, as hospital referral or admission may not occur immediately after load shedding.

Models used for sensitivity analyses used different definitions and different model classes (negative binomial model, linear model, INGARCH model, see eText1). In secondary



1) ESKOM (CEO Dames, B), Power System Emergency, 03/2014 (accessed on 03/2017, http://www.eskom.co.za/OurCompany/MediaRoom/Documents/poweremergency6march.pdf)
2) Bateman, C. (2008). "Eskom debacle: health care risks, frustrations climb." S Afr Med J 98(3): 171-173.
3) Sigauke, C. et al.(2010)."Daily peak electricity load forecasting in South Africa using a multivariate non-parametric regression approach." ORiON 26(2).
4) Hopp, S., et al. (2018). "Medical diagnoses of heat wave-related hospital admissions in older adults." Prev Med 110: 81-85.
5) Ravljen, M., et al. (2018). "Immediate, lag and time window effects of meteorological factors on ST-elevation myocardial infarction incidence." Chronobiol Int 35(1): 63-71.
6) de Pablo, F., et al. (2009). "Winter circulation weather types and hospital admissions for cardiovascular, respiratory and digestive diseases in Salamanca, Spain." Int J Climatol 29(11):1692-1703.
* Changes in weather conditions to more extremes like heat or cold lead to increasing demand by air conditioning or electric heaters and might directly influence health states.

FIGURE 1. DAG for our assumptions about the relationship between load shedding, hospital admissions, and weather. DAG indicates directed acyclic graph.

analyses, we looked at alternative exposure variables, that is, (1) the event of load shedding, on the same day or any of the two preceding days, in one of Cape Town's 16 load shedding areas; and (2) the exposure in interaction with length of load shedding, modeled nonlinearly with p-splines, and defined as the number of minutes of load shedding per day averaged over the respective areas. For both secondary exposures data on areas that were not under direct control of the city, which include both populated and unpopulated (mountainous) areas (in total 59% of Cape Town's official size), were not available and were thus not part of the calculations.

We used frequentist model averaging[16,17] to estimate the importance of the inclusion of different lags, that is, to what degree load shedding on the same day versus previous days is important to describe hospital admissions in the above Quasi-Poisson models. Briefly, frequentist model averaging means calculating Akaike's Information Criterion for all possible models. Then, a higher weight is given to models that are more plausible according to Akaike's Information Criterion. The sum of the weights of those models that include the variable of interest are used as a variable importance (VI) measure ($0 \leq VI \leq 1$). We used $VI > 0.5$ as a rule to include a lag variable.

We also used frequentist model averaging to determine the inclusion of past admissions, weather indicators, and complexity of the weekly trend. More details on the final model, and more methodological background, are given in eText1.

The directed acyclic graph (DAG) in Figure 1 represents our assumptions about the causal process leading to LS and hospital admissions.

Because local weather conditions in Cape Town may affect hospital admissions, such as viral infections and weather-related accidents, and electricity demand and therefore the probability of experiencing load shedding, local weather may be a confounder.[18] National weather conditions may affect the implementation of load shedding but is likely unrelated to admissions at RCCH. Under the assumptions represented in the DAG, the causal effect of load shedding on hospital admissions can be estimated by adjusting for local weather and seasonal patterns using appropriate methodology, for instance targeted maximum likelihood estimation (TMLE).[19,20] We estimated the average treatment effect, that is, the difference in expected number of admissions per day had there been a load shedding event each day or on any of the preceding 2 days, during the whole year, compared with if there had not been any load shedding, using TMLE with super learning.[21] We refer the reader to eText1, and the references therein, for a more technical background. Briefly, TMLE first standardizes the data with respect to the confounders presented in Figure 1. In a second targeted step, estimation of the average treatment effect as defined above is potentially improved by utilizing information from the treatment assignment mechanism, which is the probability of load shedding conditional on the potential confounders.

All analyses were conducted in R,[22] using packages "SuperLearner" and "tmle"[23] for the causal inference analysis, package "MuMIn"for model averaging, and packages "MASS" and "tscount" to fit the negative binomial and INGARCH model, respectively. We obtained ethical approval from University of Cape Town's Human Research Ethics Committee for this study (Ref#: 901/2016).

RESULTS

During the study period between June 2014 and May 2015, Cape Town experienced 72 days of load shedding, 48

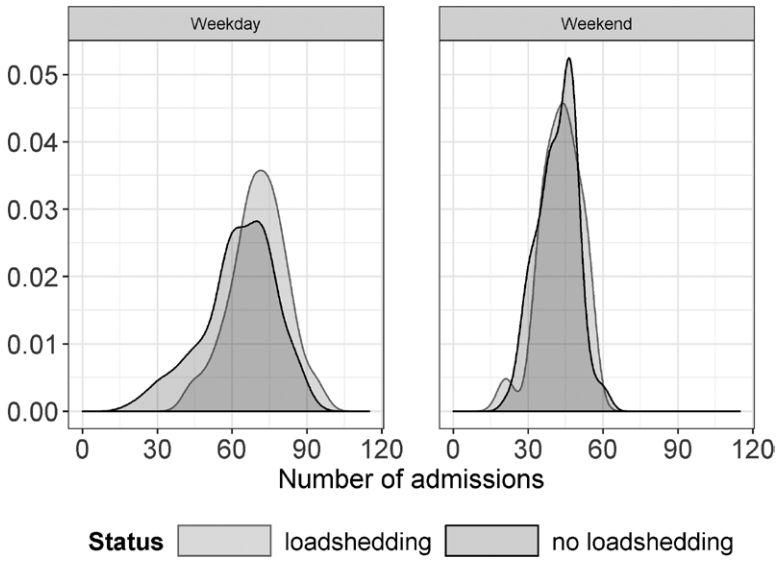


FIGURE 2. Kernel density plots for the distribution of "number of admissions"; stratified by weekday vs. weekend and load shedding vs. no load shedding. Figure is available in color online.

during the week and 24 on the weekend. Load shedding started as soon as 11 June 2014, but many events (38) occurred in April/May 2015. Figure 2 shows that there are more hospital admissions during days of load shedding, but typically during weekdays. The mean number of unscheduled admissions during the study period was about 57. On days of load shedding there were on average 61.3 admissions a day, and on days without there were about 56.7 admissions.

As speculated in the DAG, weather conditions, such as precipitation, were associated with both the probability of load shedding, and the rate of hospital admissions, supporting our initial assumption that weather may be a confounder (eFigure 2; http://links.lww.com/EDE/B390).

The Table shows that load shedding leads to a 10% increase (95% confidence interval [CI]: 4%, 15%) in hospital admissions, after adjustment for weather indicators, month, week of payment, seasonal trends, and past admissions. Similar results are obtained under a standard Poisson model or a linear regression model, but these models violated certain assumptions, including overdispersion and normality of residuals.

Using a negative binomial regression model led to an estimate of 10% (5%, 15%), an INGARCH model to 6% (2%, 10%), though not all assumptions were met for the latter model (eFigure 6; http://links.lww.com/EDE/B390). Using another definition of a LS event (same day, only 1 day prior, only 2 days prior, only 3 days prior) led to incidence rate ratios (IRRs) which suggest an increase in hospital admissions between 5% and 9% (Table). The VI measure obtained from frequentist model averaging suggests that inclusion of a 3- or 4-day lag period does not add much (VI ≤ 0.4) information. This demonstrates the usefulness of evaluating the effect of LS for events that occurred up to 2 days prior of admission.

The IRRs, estimated for occurrence of load shedding in each of the city's official 16 load shedding areas, are visualized in Figure 3.

LS events in the southern peninsula of Cape Town and in the residential areas of Durbanville produced the lowest rate ratios. High IRRs were found for the township of Philippi, the satellite town of Atlantis, and areas close to Red Cross hospital (Hanover Park, Lansdown, Observatory, Rondebosch, Newlands) and areas of mixed population and income groups (Hout Bay, southern suburbs, Parow, Goodwood). These areas were sometimes, but not always, located close to areas of lower median household income (eFigure 4; http://links.lww.com/EDE/B390). There are associations of varying strength between the implementation of load shedding in different areas, as this followed a schedule, highlighting the complex spatial dependence structure (eFigure 5; http://links.lww.com/EDE/B390).

Hospital admission rates did not differ substantially when comparing overall surgical with medical specialties, but results differed with respect to the different diagnoses based on ICD-10 code chapters (Table). The highest IRR was observed for diseases of the eye and ear (12% [-16%, 48%]), the digestive system (11% [-8%, 33%]) and the respiratory system (14% [2%, 26%]). No relevant changes in admission were observed for intoxications or infections not defined in other ICD-10 chapters.

In exploratory analyses, we found that the increase in admissions occurred primarily during the week as shown by inclusion of an interaction with weekend/weekday in the model (12% [7%,18%] vs. 1% [-9%, 11%]). Moreover, we could not find evidence that length of load shedding affected the rate of admissions (eFigure 3; http://links.lww.com/EDE/B390).

TABLE. Incidence Rate Ratios (IRR), Obtained from a Quasi-Poisson Model, That Is, Ratio of Admissions for Days Where Load Shedding (LS) was Experienced on the Same Day, or No More than 2 Days Prior, Compared with When There Was No LS in the Past 2 Days. Main Results Are Highlighted in Bold.

Intervention	IRR[a]	95% CI	VI[b]
Main result 1:			
LS: same day or up to 2 days prior	**1.10**	**1.04;1.15**	
Interaction:			
Weekday	1.12	1.07, 1.18	
Weekend	1.01	0.91, 1.11	
Interaction: length of load shedding see eFigure3			
Other models:			
LS: same day	1.05	1.00, 1.11	0.30
LS: 1 day prior	1.09	1.04, 1.15	0.85
LS: 2 days prior	1.07	1.01, 1.13	0.69
LS: 3 days prior	1.07	1.01, 1.13	0.40
LS: 4 days prior	0.98	0.93, 1.04	0.34
By specialty:			
Surgical cases	1.08	1.00, 1.16	
Medical cases	1.11	1.05, 1.18	
By ICD-10 codes (code range):			
Certain Infections (A00-B99)	1.04	0.92, 1.17	
Eye & Ear (H00-H95)	1.12	0.84, 1.48	
Respiratory system (J00-J99)	1.14	1.02, 1.26	
Digestive system (K00-K93)	1.11	0.92, 1.33	
Skin (L00-L99)	1.06	0.86, 1.31	
Injuries, Poisoning (S00-Y98)	0.97	0.83, 1.13	
Other	1.11	1.03, 1.20	
	ATE	**95% CI**	
Main result 2: ATE estimated by TMLE[c]	**6.50**	**5.12, 7.87**	
Naïve linear regression, adjusted for confounders	5.04	2.29, 7.80	

The reported average treatment effect (ATE), estimated with targeted maximum likelihood estimation (TMLE), estimates the difference in expected number of admissions per day had there been a LS event each day or on any of the preceding 2 days, during the whole year, compared with if there had not been any LS.

[a]All estimates are adjusted for weather history, month, week of payment, admission history, and a seasonal weekly trend, see eText 1 for the exact model specification.

[b]Variable importance (VI) obtained from frequentist model averaging via quasi-Akaike Information Criterion weights (between 0 = unimportant and 1 = very important), see eText 1 for more details.

[c]Targeted maximum likelihood estimates have been obtained using super learning, see eText1 for the list of learners.

ATE indicates average treatment effect; CI, confidence interval; IRR, incidence rate ratio; VI, variable importance.

Using TMLE, we estimated the average treatment effect as 6.50 (95% CI: 5.12, 7.87), or in other words that about six additional admissions a day could be attributed to load shedding.

Average treatment effects and IRRs for individual diagnoses are listed in eTable 1; http://links.lww.com/EDE/B390. Most of these estimates are not precise enough to conclude that specific diagnoses would occur more often in days following load shedding events.

DISCUSSION

Our analyses demonstrate that load shedding as implemented in South Africa is associated with a substantial increase in hospital admissions of children, on the same day and up to 2 days following the power interruption. Under the assumptions that we have identified and included all confounding variables in our analysis, that is, that the DAG from Figure 1 is correct, and that the modeling approach discussed in eText1 is appropriate, this effect is causally interpretable. Note that this applies to the average treatment effect estimated by TMLE, as the IRRs estimated with quasi-Poisson regression require stronger assumptions to be causally interpretable, for example, a constant treatment effect across all covariate strata.

A strength of our study is our rigorous approach of data collection using Twitter, Facebook, and the City of Cape Town after local authorities, including ESKOM, did not support our request for data sharing. Moreover, contacted radio stations only kept short-term records of less than 2 weeks. This approach of data collection may serve as a future model for surveys where data that normatively should be publicly available are withheld. Moreover, we have clearly communicated our assumptions under which our effect estimates are causally interpretable and used state-of-the-art methodology to facilitate this analysis.

Our study has some limitations. First, we did not have access to individual patient folders and higher numbers of specific diagnoses to further expand our hypotheses on what are the biggest risks of load shedding. Our results on individual diagnoses are imprecise. Moreover, while we had been able to identify the areas of load shedding events, these areas are large and often cover populations of different household income groups and ethnicities. Our results may indicate that poorer areas could be affected more heavily by load shedding than wealthier areas. However, wherever load shedding was not directly controlled by the city but ESKOM, data were unavailable. This includes townships such as Khayelitsha and Nyanga, and small residential areas in Cape Town's north. There is also a complex spatial-temporal relationship with respect to the implementation of load shedding in different areas. We may not have been able to model this in all detail and it may therefore be advisable to interpret our spatial analysis with care. In addition, there is the possibility of migration between different areas, though this may be negligible as the study period is very short.

In general, the methods we use (quasi-Poisson regression, TMLE with super learning) require that observations be independent conditional on the covariates; this assumption is needed for the validity of the likelihood functions used and for the application of super learning. Although we have tried to include seasonal trends, past admissions and past weather indicators to meet this assumption, we cannot exclude the possibility that there remains dependence which we have not been able to model. In this case, inference may be affected, and CIs may not be correct. Apart from missed seasonal trends,

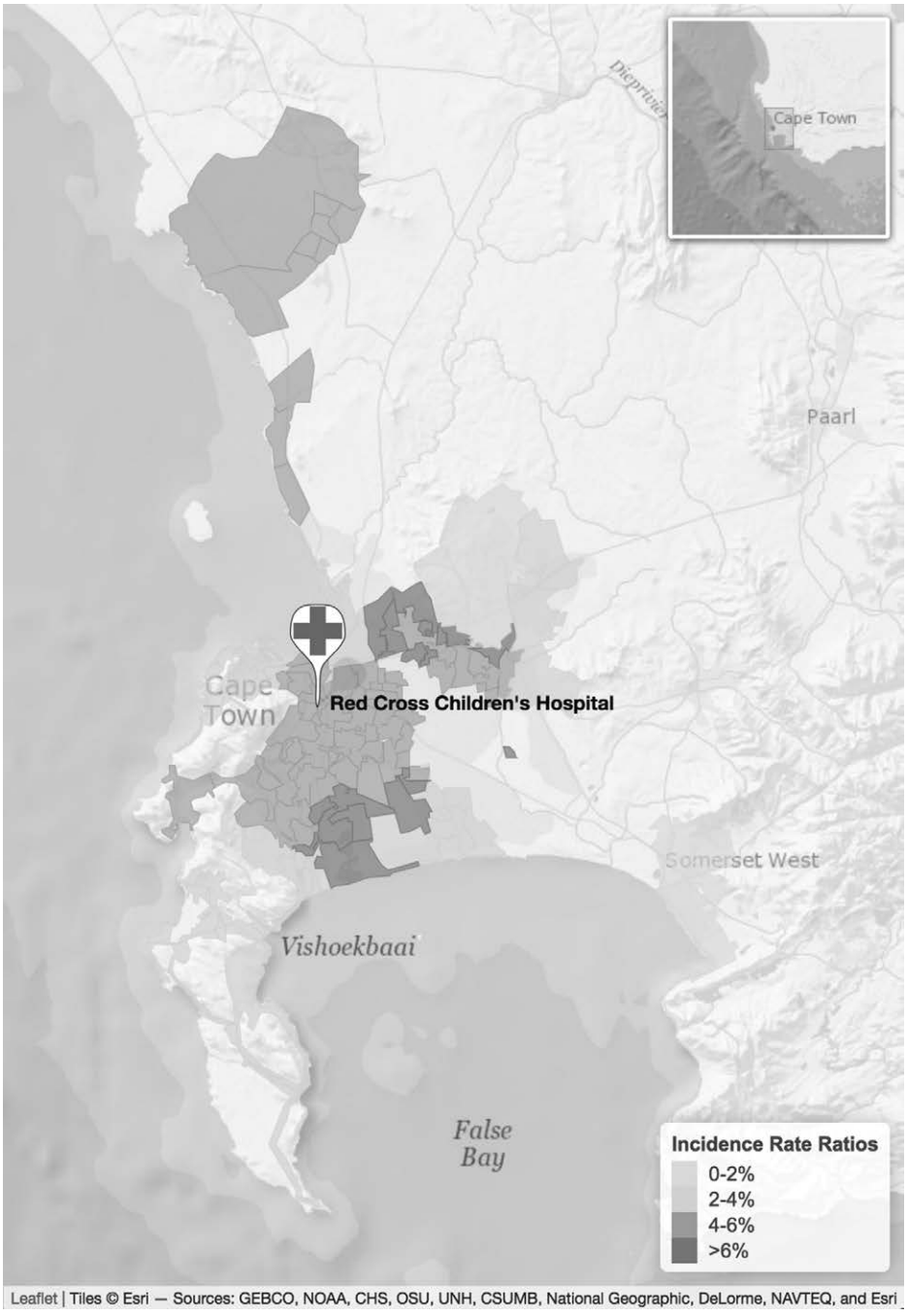


FIGURE 3. Incidence rate ratios for number of hospital admissions depending on area of load shedding, calculated with an adjusted Quasi-Poisson regression model, see eText 1 for details. Those areas which were not under the city's control (but ESKOM's control) are excluded because of data unavailability. Figure is available in color online.

there could be other confounders we have not measured: however, these could only be variables that cause an unusually high energy demand in the country, such as international sporting events, although we are unaware of any such events in the relevant time period. Last, it is important to note that our results may not be generalizable to high-income countries or settings where living conditions differ greatly from those in South Africa.

The effect of load shedding on health may be best explained with case reports from RCCH. For example, there was a patient with a skin burn because of handling candles during load shedding and the father admitted the patient after the skin got infected 2 days later. Another admission was related to injuries caused by a pan containing hot fat, which had been placed near an outdoor fire since the electric kitchen stove could not be used. These two cases happened at night and at home, where accidents may happen most often. Since Capetonians have long daily commutes,[24] often more than 2 hours one-way, it may well be that accidents (related to limited lighting and heating options) happen after they return home from

work. This could explain why, by our estimates, load shedding affected admissions primarily during weekdays, where people arrive home late. Besides such obvious relationships, there might be less obvious ones that may not be immediately clear to treating doctors: for example, inhalation of fumes from an improvised stove, or the ingestion of food from an interrupted cold chain. Increased admissions owing to eye and ear, lung, and digestive system suggest external noxious influences, for example, owing to combustion, as a possible trigger.

We did not find ICD-10 codes of the mixed chapter "injuries, poisoning, and other external causes" to be contributing to the increased number of admissions as described for other blackouts. Since we deal with children only, it may well be that access to toxins such as gasoline is limited. Furthermore, the diverse diagnoses covered in this chapter might make an observable effect less likely.

Moreover, infections are partly covered by the respective organ specific ICD-10 chapter, for example, for the respiratory system. This explains the missing relationship of the ICD-10 chapter "Certain infectious and parasitic diseases" with load shedding and more generally highlights the challenges of the interpretation of grouped disease categories. In-depth analyses implied trends of higher incidences of burns, traumatic fractures, meningitis, and other individual diagnoses, but the number of cases per diagnosis were too small for reliable interpretation. Bigger studies are needed to enhance our understanding of (indirect) causal relationships and, most importantly, to prevent casualties in situations when power failures occur.

An increased number of hospital admissions during load shedding leads to an increased burden of already overwhelmed health care facilities. Additional resources are not necessarily available, and it remains unclear what the consequences of the additional costs are. This consideration is relevant in the current and very lively debate on South Africa's future energy mix. While costs of generating energy, political considerations, and CO_2 emissions are certainly relevant aspects of this discussion, security of an uninterrupted power supply should remain a priority not only from an economic perspective, but also from a public health point of view. As we have shown, the above measured association is consistent with our hypothesis that failures of the power infrastructure increase risk to children's health.

ACKNOWLEDGMENTS

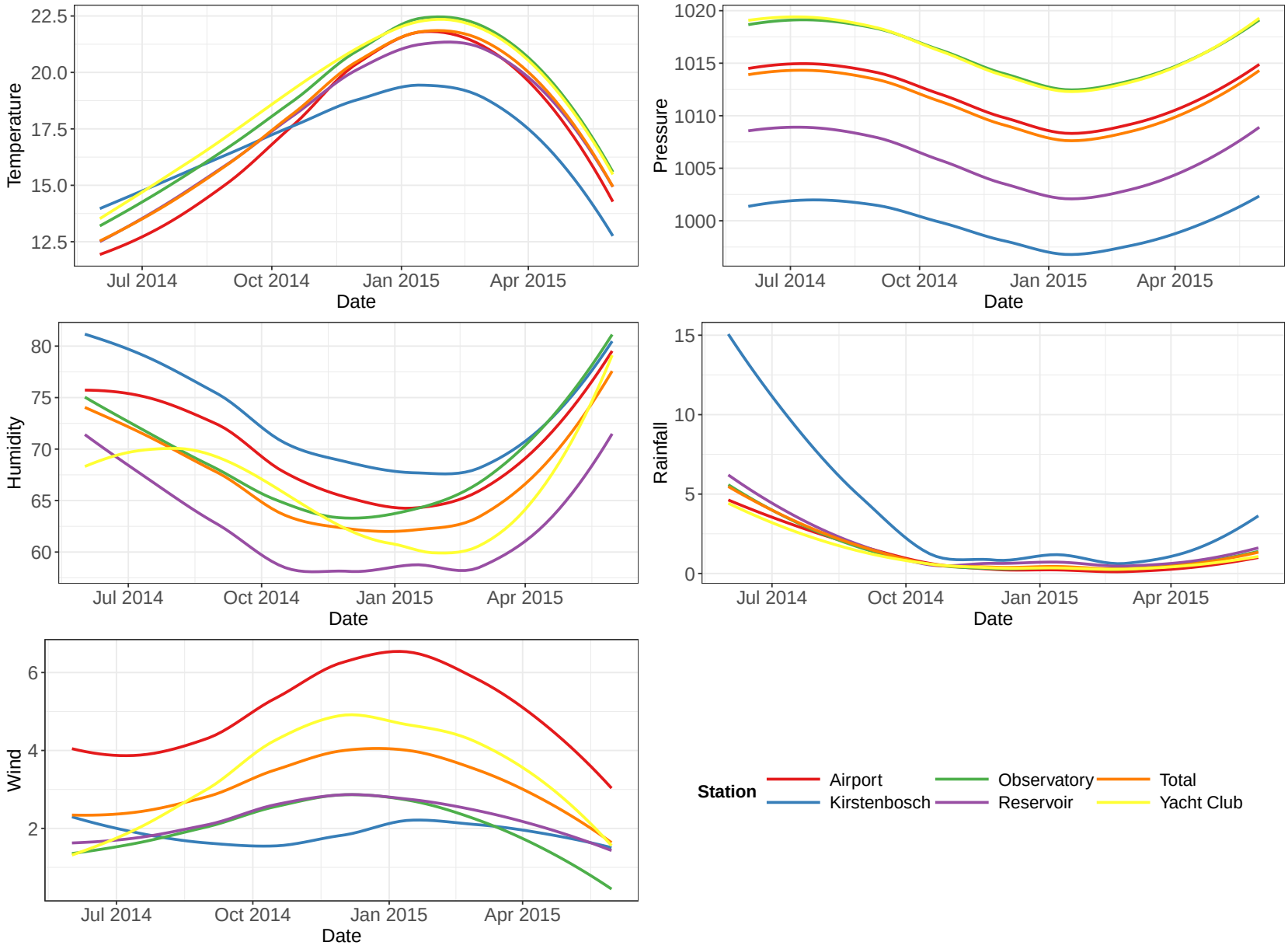
We greatly acknowledge the help from the staff of the electricity generation and distribution department of the City of Cape Town. We would like to particularly thank Peter Jaeger in his help of administering and interpreting load shedding data. We are also immensely grateful to Mary-Ann Davies and Kenneth Sinclair-Smith for their committed support of this project and their active engagement in the logistics of it. We further thank Elsa DeJager from the South African Weather Service, and Mark Seiderman from the National Centers for Environmental Information, for their help in acquiring our weather data. We also thank Annibale Cois for his insights on this topic; and Daniela Bandic, Thomas Derungs and Sabine Bélard for constructive support throughout this project.

REFERENCES

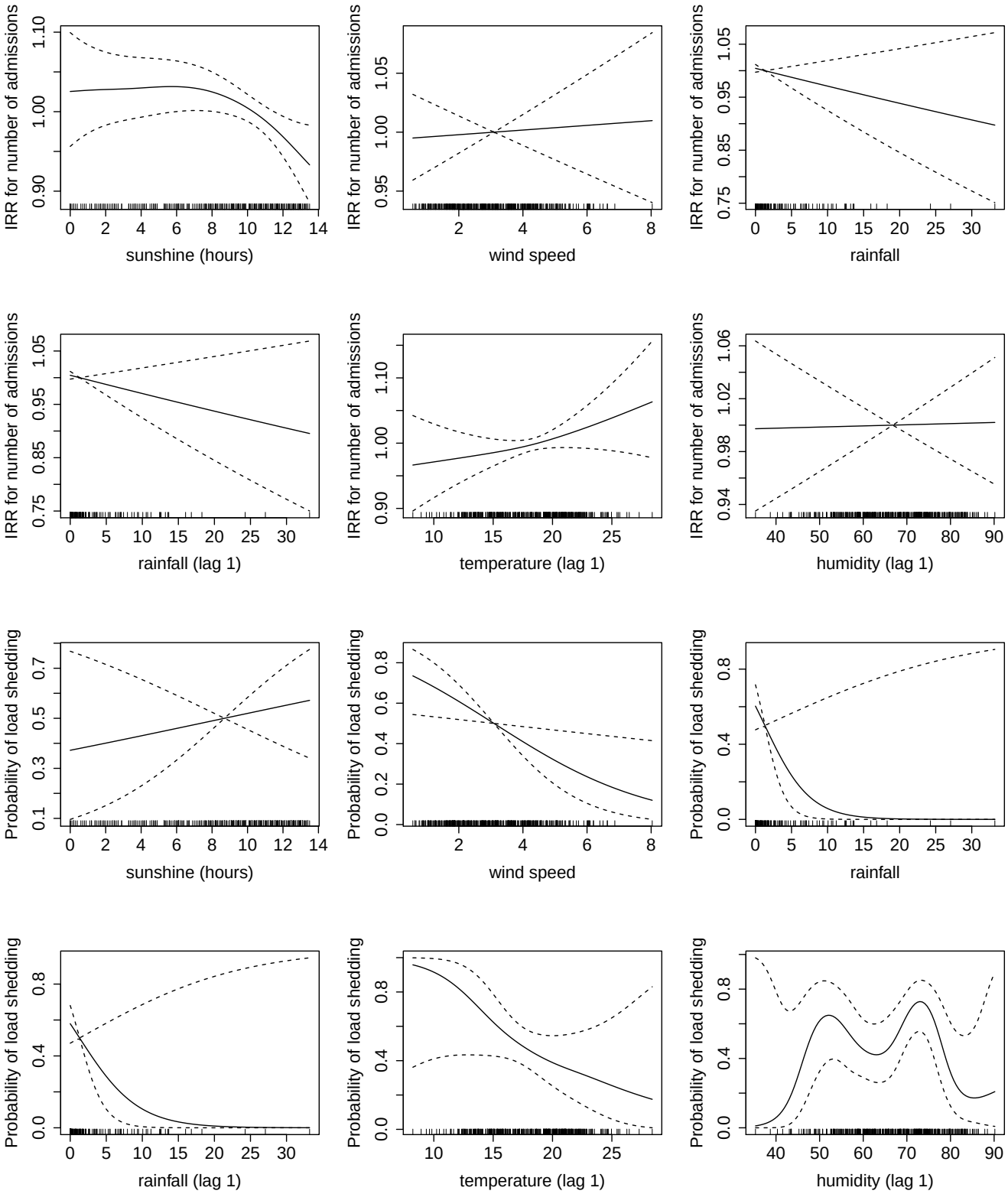
1. General household survey 2015. Available at: http://www.statssa.gov.za. Statistics South Africa (STATS SA). Accessed August 2018.
2. Inglesi R, Pouris A. Forecasting electricity demand in South Africa: a critique of Eskom's projections. *South African J Sci*. 2010;106:50–53.
3. Bateman C. Eskom debacle: health care risks, frustrations climb. *S Afr Med J*. 2008;98:171–173.
4. Anderson GB, Bell ML. Lights out: impact of the August 2003 power outage on mortality in New York, NY. *Epidemiology*. 2012;23:189–193.
5. Eachempati SR, Mick S, Barie PS. The impact of the 2003 blackout on a level 1 trauma center: lessons learned and implications for injury prevention. *J Trauma*. 2004;57:1127–1131.
6. Greenwald PW, Rutherford AF, Green RA, et al. Emergency department visits for home medical device failure during the 2003 North America blackout. *Acad Emerg Med*. 2004;11:786–789.
7. Klein KR, Herzog P, Smolinske S, et al. Demand for poison control center services "surged" during the 2003 blackout. *Clin Toxicol (Phila)*. 2007;45:248–254.
8. Lin S, Fletcher BA, Luo M, Chinery R, et al. Health impact in New York City during the Northeastern blackout of 2003. *Public Health Rep*. 2011;126:384–393.
9. Pironi L, Spinucci G, Paganelli F. Effects of the September 28 2003 blackout in Italy in patients on home parenteral nutrition (HPN). *Clin Nutr*. 2004;23:133.
10. Abir M, Jan S, Jubelt L, et al. The impact of a large-scale power outage on hemodialysis center operations. *Prehosp Disaster Med*. 2013;28:543–546.
11. Johnson-Arbor KK, Quental AS, Li D. A comparison of carbon monoxide exposures after snowstorms and power outages. *Am J Prev Med*. 2014;46:481–486.
12. Kelman J, Finne K, Bogdanov A, et al. Dialysis care and death following Hurricane Sandy. *Am J Kidney Dis*. 2015;65:109–115.
13. Nakayama T, Tanaka S, Uematsu M, et al. Effect of a blackout in pediatric patients with home medical devices during the 2011 eastern Japan earthquake. *Brain Dev*. 2014;36:143–147.
14. Luque-Fernandez MA, Van Cutsem G, Goemaere E, et al. Effectiveness of patient adherence groups as a model of care for stable patients on antiretroviral therapy in Khayelitsha, Cape Town, South Africa. *PLoS One*. 2013;8:e56088.
15. Wood SN. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC; 2006.
16. Burnham K, Anderson D. *Model Selection and Multimodel Inference. A Practical Information-theoretic Approach*. New York: Springer; 2002.
17. Schomaker M, Heumann C. Model selection and model averaging after multiple imputation. *Computational Stat Data Anal*. 2014;71:758–770.
18. Sigauke C, Chikobvu D. Daily peak electricity load forecasting in South Africa using a multivariate non-parametric regression approach. *ORiON (Operations Research Society of South Africa (ORSSA))*. 2010;26:16.
19. Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge University Press; 2009.
20. Van der Laan M, Rose S. *Targeted Learning*. Springer; 2011.
21. Van der Laan M, Polley E, Hubbard A. Super learner. *Stat Appl Genet Mol Biol*. 2008;6:Article 25.
22. R-Core-Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.R-project.org/. 2017.
23. Gruber S, van der Laan MJ. tmle: An R Package for targeted maximum likelihood estimation. *J Stat Software*. 2012;51:1–35.
24. Sinclair-Smith K, Turok I. The changing spatial economy of cities: an exploratory analysis of Cape Town. *Dev Southern Africa*. 2012;29:391–417.

Supplementary Material for "Christian Gehringer, Heinz Rode, Michael Schomaker: *The effect of load shedding on pediatric hospital admissions in South Africa*"

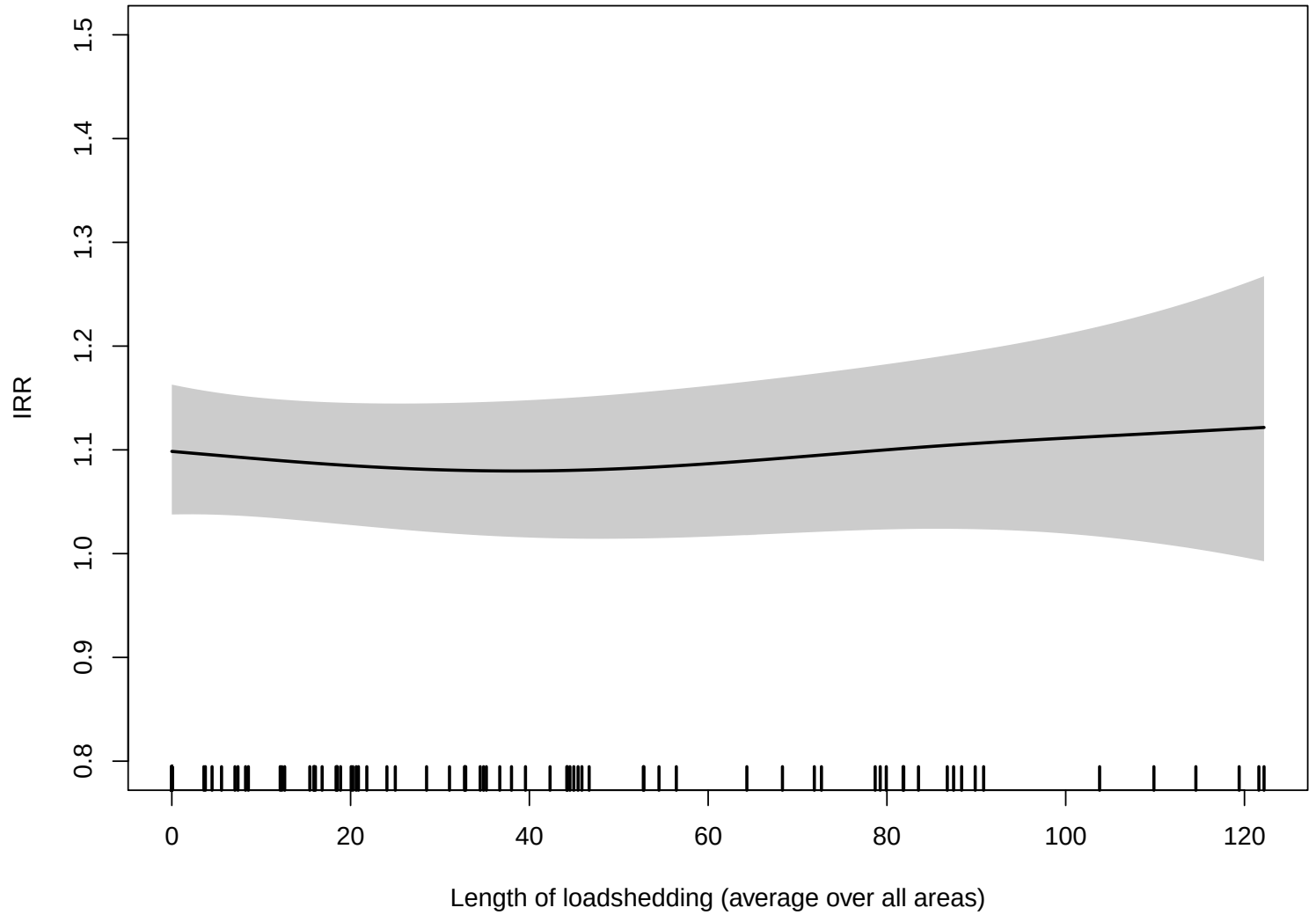
eFigure 1. Smoothed weather data for different weather stations in the Cape Town area: relative humidity (in %), pressure (in hectopascal), rainfall (in mm), temperature (in degrees Celsius), and wind speeds (in meter/second).



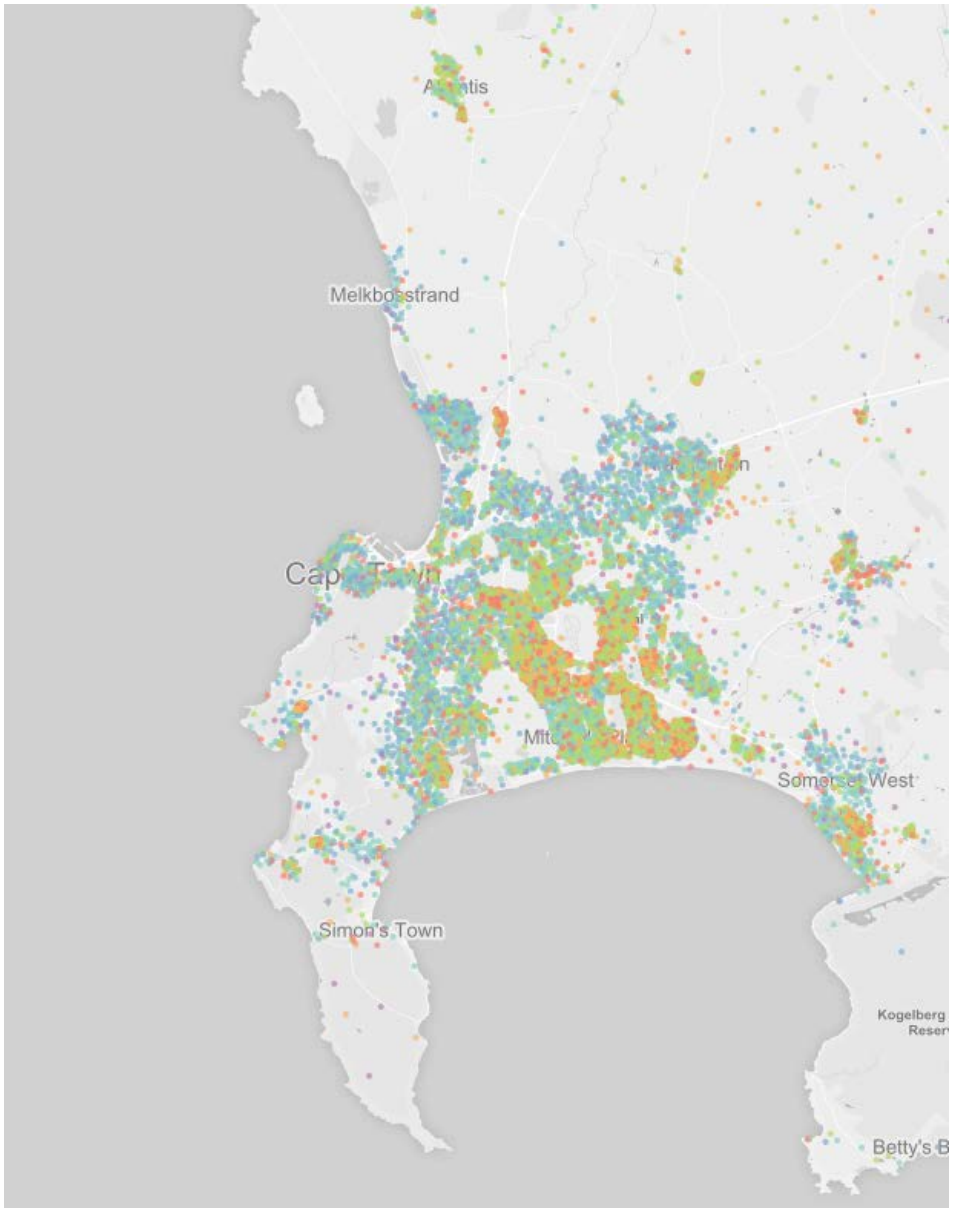
eFigure 2. Probability of load shedding (bottom 6 panels), and incidence rate ratios for number of hospital admissions (top 6 panels), depending on weather indicators calculated with an additive (covariate adjusted) logistic regression model. The reported weather indicators are sunshine (in hours), wind speed (in meter/second), rainfall (in mm), temperature (in degrees Celsius), and relative humidity (in %). Each panel shows the respective estimated penalized spline, conditional on other weather indicators and month.



eFigure 3. Non-linear interaction between load shedding (same day or up to 2 days prior) and length of load shedding (average time, measured in minutes, per area and per day), modelled via penalized splines in an additive Quasi-Poisson model.



eFigure 4. Median yearly household income, as estimated in the national census from 2011, stratified by area. The dot map was produced by the tool of Adrian Firth, available at https://dotmap.adrianfrith.com/.

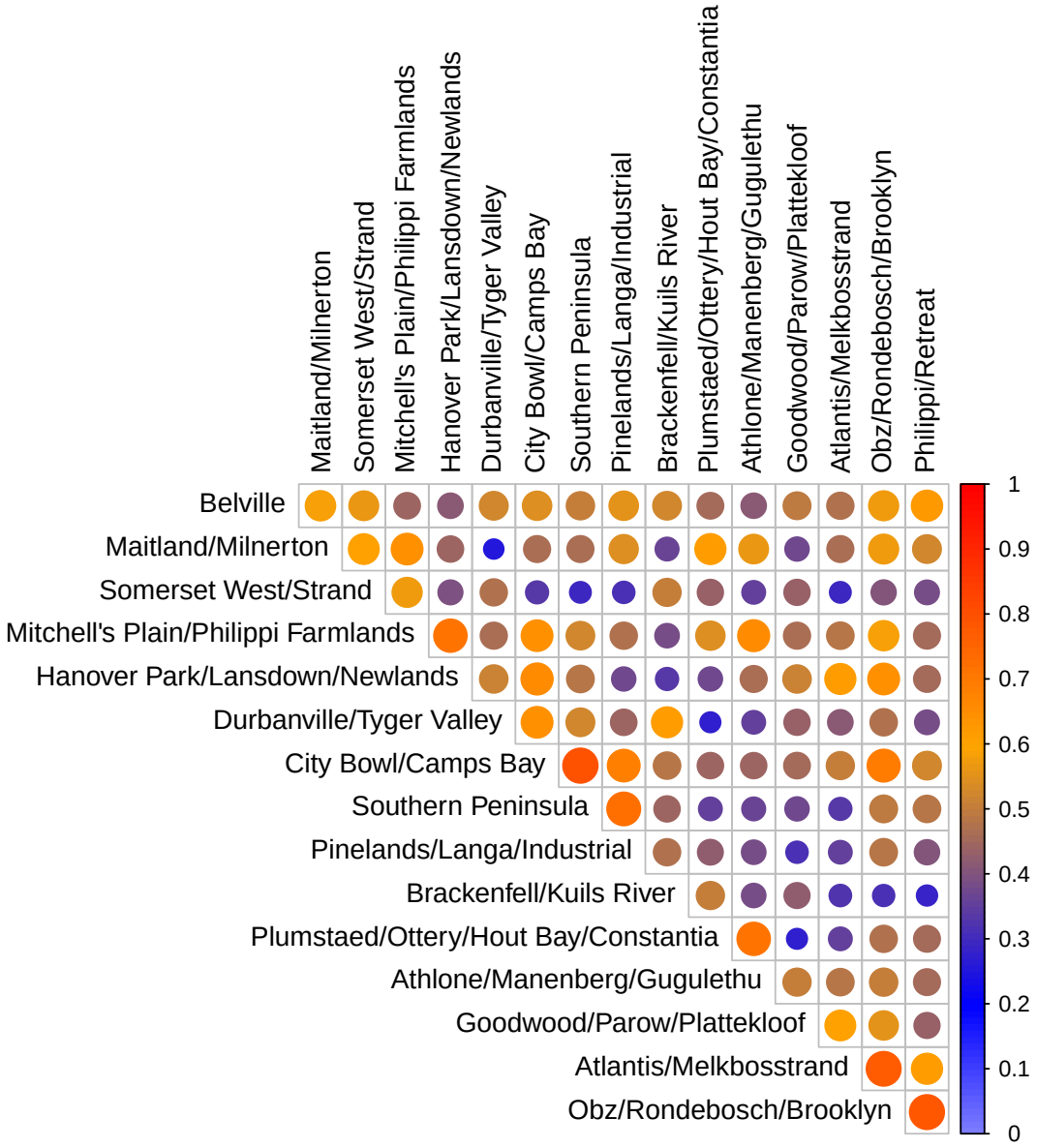


One dot = 100 households

R0 –	R4 800
R4 801 –	R19 600
R19 601 –	R76 400
R76 401 –	R307 600
R307 601 –	R1 228 800
R1 228 800 –	

Annual household income from Census 2011. Note that some 800 000 people living in institutional settings are not included in household statistics.

eFigure 5. Association between load shedding events in different areas measured by Cramer's V. The closer Cramer's V is to 1, i.e. the more red and the bigger the circle, and the stronger the association between two areas in terms of load shedding events taking place on the same day.



eTable 1. Average Treatment Effect (ATE) and incidence rate ratios (IRR) for diagnoses where the number of reported cases exceeds 25.

ICD-10	Diagnosis	IRR	95%	CI	ATE	95%	CI
J05	Acute obstructive laryngitis	1.74	0.53	5.73	0.05	0.01	0.10
T22	Burn & corrosion of shoulder & upper limb	1.47	0.45	4.85	0.29	0.21	0.37
T24	Burn & corrosion of lower limb	1.45	0.57	3.73	0.06	0.01	0.11
T23	Burn & corrosion of wrist and hand	1.37	0.49	3.83	0.01	-0.04	0.06
A41	Sepsis	1.34	0.62	2.91	0.00	-0.06	0.06
J44	Chronic obstructive pulmonary disease	1.29	0.58	2.88	-0.08	-0.16	-0.00
Q21	Congenital malformations of cardiac septa	1.27	0.37	4.34	0.61	0.49	0.73
J18	Pneumonia	1.24	0.74	2.07	0.00	-0.12	0.13
S09	Unspecified injuries of head	1.08	0.75	1.57	0.05	-0.08	0.18
G40	Epilepsy	1.08	0.54	2.16	-0.13	-0.23	-0.03
Z00	General examination	1.07	0.36	3.16	0.18	0.10	0.27
J21	Acute bronchiolitis	1.01	0.56	1.80	-0.24	-0.40	-0.07
T20	Burn & corrosion of head, face, and neck	0.95	0.42	2.14	-0.10	-0.17	-0.03
J45	Asthma	0.90	0.40	2.03	-0.07	-0.14	-0.01
T29	Burns of multiple regions	0.87	0.41	1.86	-0.13	-0.20	-0.06
A87	Viral meningitis	0.87	0.25	3.04	0.23	0.13	0.33
T74	Child abuse, neglect & other maltreatment	0.81	0.33	2.02	0.02	-0.02	0.07
A09	Infectious gastroenteritis	0.79	0.52	1.19	-0.12	-0.36	0.12
S42	Fracture of shoulder and upper arm	0.79	0.34	1.82	-0.04	-0.10	0.01
T21	Burn & corrosion of trunk	0.69	0.31	1.55	0.00	-0.07	0.08
S72	Fracture of femur	0.62	0.30	1.30	-0.09	-0.16	-0.02
S52	Fracture of forearm	0.61	0.30	1.23	-0.12	-0.19	-0.05
S06	Intracranial injury	0.59	0.16	2.17	-0.07	-0.11	-0.03
S82	Fracture of lower leg	0.59	0.25	1.40	-0.10	-0.15	-0.04
G03	Meningitis	0.56	0.13	2.50	-0.05	-0.10	-0.01
L02	Cutaneous abscess	0.52	0.24	1.15	-0.10	-0.18	-0.02
J22	Acute lower respiratory infection	0.37	0.14	0.99	-0.11	-0.17	-0.04

eText 1. Technical details about the statistical approaches used in the paper.

Content:

Notation and Variables. We are interested in the number of admissions at Red Cross hospital Y_i, measured on days $i = 1 \ldots, 365$. Then, $E(Y)$ refers to the expected number of admissions per day. The (main) interventions of interest are the binary variable A indicating whether load shedding has been implemented on day i or not, and $\bar{A}^2$ which denotes whether load shedding occurred on the same or any of the two preceding days. The matrix of potential confounders $\mathbf{L}$ includes hours of sunshine (L^1), wind speed (L^2), humidity (L^3), pressure (L^4), precipitation (L^5), temperature (L^6), month (L_7), last week of the month (week of payment, L_8), and a seasonal (weekly) trend modeled via sine and cosine terms, i.e.

$$\cos(\omega_k t) \quad \text{and} \quad \sin(\omega_k t) \quad \text{with} \quad \omega_k = \frac{2k\pi}{T}$$

with $T = 7$ days. More specifically we consider $\cos(\omega_1 t) = L^9$, $\sin(\omega_1 t) = L^{10}$, $\cos(\omega_2 t) = L^{11}$, $\sin(\omega_2 t) = L^{12}$, $\cos(\omega_3 t) = L^{13}$, $\sin(\omega_3 t) = L^{14}$, $\cos(\omega_4 t) = L^{15}$, $\sin(\omega_4 t) = L^{16}$ to be potential confounders. Moreover, in most of the below described models, we also consider past values of the weather indicators up to a lag of 2 days as well as past admissions (i.e. $Y_{i-\text{lag}}$, $\text{lag} \in \{1, 2, \ldots, 13, 14, 21, 28\}$) to be of interest.

Frequentist Model Averaging. Since the potential set of adjustment variables is large, mostly due to the number of lagged and seasonal variables, we need to reduce the number of variables, so that the implementation of the below specified models is computationally feasible at all (i.e. for successful fitting of the splines and the causal inference procedures). A classic approach would be model selection with hypothesis testing, or using a criterion such as Akaike's Information Criterion (AIC). However, data-driven model selection has the disadvantage that model selection uncertainty is not being taken into account, i.e. different samples of data may yield different conclusions with respect to variable inclusion (Chatfield, 1995). An alternative to model selection is model averaging.

With model averaging, one calculates a weighted average $\hat{\bar{\beta}} = \sum_\kappa w_\kappa \hat{\beta}_\kappa$ from the k estimators $\hat{\beta}_\kappa$ ($\kappa = 1, \ldots, k$) of the set of candidate (regression) models $\mathcal{M}$ where the weights are calculated in a way such that 'better' models receive a higher weight. A popular weight choice would be based on the exponential AIC,

$$w_\kappa^{\text{AIC}} \quad = \quad \frac{\exp(-\frac{1}{2}\text{AIC}_\kappa)}{\sum_{\kappa=1}^{k} \exp(-\frac{1}{2}\text{AIC}_\kappa)},$$

where AIC_κ is the AIC value related to model $M_\kappa \in \mathcal{M}$ (Buckland et al., 1997). Note that the weights sum up to one. The k point and k variance estimates can then be combined into a single model, based on the weights. In our analysis we are not interested in this combined estimate but rather into a variable importance measure obtained from the model averaging weights. With this, the sum of the weights of those models which include the variable of interest are simply being added up. So, if there is a single best model (which has by far the lowest AIC), then this model receives weight 1 and all variables contained in it receive a variable importance measure of 1 (and the others 0). However, if there are say 10 very good models (measured by AIC), and only 4 of those contain the respective variable, then the sum of the weights of those 4 models would determine the variable's importance. We use a 50% support, i.e. a variable importance of 0.5 or higher, as a rule to include a variable into our final model. More details are given below. Readers who are interested into model averaging may consult the following references: Burnham and Anderson (2002), Hjort and Claeskens (2003), Hoeting et al. (1999).

Quasi-Poisson Model: Model Specification. Our considerations start with a Poisson model which considers all potential variables of interest:

$$E(Y) = \exp(\beta_0 + \beta_1 A + \mathbf{L}\boldsymbol{\beta_2}) \quad (1)$$

where $\mathbf{L}$ includes $L^1, \ldots, L^{16}$, $L^1_{i-1}, \ldots, L^6_{i-1}$, $L^1_{i-2}, \ldots, L^6_{i-2}$, and Y_{i-l}, $l \in \{1, 2, \ldots, 13, 14, 21, 28\}$. Note that a standard Poisson model assumes $E(Y) = Var(Y) = \mu$ which may not be met (and in fact is not met in our data[1]). For this reason we use below only Quasi-Poisson models which allow $Var(Y) = \phi E(Y)$, where ϕ is a parameter which is constant across all $i = 1, \ldots, n$. The parameter ϕ can be estimated from the data using the χ^2-statistic, and can also be used to correct the standard maximum likelihood estimate $\widehat{\text{Var}}(\hat{\beta}_{i,ML})$, i.e. one can use $\hat{\phi} \times \widehat{\text{Var}}(\hat{\beta}_{i,ML})$ to construct confidence intervals for $\hat{\beta}_i$. With this approach we still use the model equation (1), but with more flexible variance estimation (McCullagh and Nelder, 1989).

To reduce the number of variables we consider different models in which a number of variables are being held fixed and others vary (i.e. a group of related variables, say lagged variables) so that model averaging and the above introduced variable importance measure can be calculated[2].

i) To determine the lags needed for $Y_{i-\text{lag}}$ we implement model averaging based on the following full Quasi-Poisson model:

$$\begin{aligned} E(Y) = \exp(&\beta_0 + \beta_1 \bar{A}^2 + \beta_2 Y_{i-1} + \beta_3 Y_{i-2} + \beta_4 Y_{i-3} + \beta_5 Y_{i-4} + \beta_6 Y_{i-5} + \beta_7 Y_{i-6} + \beta_8 Y_{i-7} + \beta_9 Y_{i-8} + \beta_{10} Y_{i-9} \\ &+\beta_{11} Y_{i-10} + \beta_{12} Y_{i-11} + \beta_{13} Y_{i-12} + \beta_{14} Y_{i-13} + \beta_{15} Y_{i-14} + \beta_{16} Y_{i-21} + \beta_{17} Y_{i-28} + \beta_{18} L^9 + \beta_{19} L^{10} + \beta_{20} L^{11} \\ &+\beta_{21} L^{12} + \beta 22 L^7 + \beta_{23} L^8 + \beta_{24} L^1 + \beta_{24} L^2 + \beta_{24} L^5 + \beta_{25} L^3_{i-1} + \beta_{26} L^5_{i-1} + \beta_{27} L^6_{i-1} + \beta_{28} L^2_{i-2}) \end{aligned}$$

This analysis yields variable importance measures > 0.5 for the following lags of Y: $1, 3, 7, 9$.

ii) To determine the amplitude of the seasonal component, i.e. the k needed for $\cos(\omega_k t)$ and $\sin(\omega_k t)$, we implement model averaging based on the following full Quasi-Poisson model:

$$\begin{aligned} E(Y) = \exp(&\beta_0 + \beta_1 \bar{A}^2 + \beta_2 Y_{i-1} + \beta_3 Y_{i-3} + \beta_4 Y_{i-7} + \beta_5 Y_{i-9} \\ &+\beta 6 L^7 + \beta_7 L^8 + \beta_8 L^1 + \beta_9 L^2 + \beta_{10} L^5 + \beta_{11} L^3_{i-1} + \beta_{12} L^5_{i-1} + \beta_{12} L^6_{i-1} + \beta_{14} L^2_{i-2} \\ &\beta_{15} L^9 + \beta_{16} L^{10} + \beta_{17} L^{11} + \beta_{18} L^{12} + \beta_{19} L^{13} + \beta_{20} L^{14} + \beta_{21} L^{15} + \beta_{22} L^{16}) \end{aligned}$$

Cosine and sine terms (i.e. $L^9 - L^{16}$) that are supported by a variable importance of > 0.5 are $k = 1$ and $k = 2$.

iii) To determine the importance of the potential confounder weather we implement model averaging for the following two Quasi-Poisson and logistic regression models:

$$\begin{aligned} E(Y) = \exp(&\beta_0 + \beta_1 \bar{A}^2 + \beta_2 Y_{i-1} + \beta_3 Y_{i-3} + \beta_4 Y_{i-7} + \beta_5 Y_{i-9} + \beta 6 L^7 + \beta_7 L^8 + \beta_8 L^9 + \beta_9 L^{10} + \beta_{10} L^{11} + \beta_{11} L^{12} \\ &\beta_{12} L^1 + \beta_{13} L^2 + \beta_{14} L^3 + \beta_{15} L^4 + \beta_{16} L^5 + \beta_{17} L^6 + \beta_{18} L^1_{i-1} + \beta_{19} L^2_{i-1} + \beta_{20} L^3_{i-1} + \beta_{21} L^4_{i-1} + \beta_{22} L^5_{i-1} \\ &+\beta_{23} L^6_{i-1} + \beta_{24} L^1_{i-2} + \beta_{25} L^2_{i-2} + \beta_{26} L^5_{i-2} + \beta_{27} L^6_{i-2}) \end{aligned}$$

$$\begin{aligned} P(A = 1) = &\beta_0 + \beta_1 A_{i-1} + \beta_2 A_{i-2} + \beta 3 L^7 + \beta_4 L^8 + \beta_5 L^1 + \beta_6 L^2 + \beta_7 L^3 + \beta_8 L^4 + \beta_9 L^5 + \beta_{10} L^6 + \beta_{11} L^1_{i-1} \\ &+\beta_{12} L^2_{i-1} + \beta_{13} L^3_{i-1} + \beta_{14} L^4_{i-1} + \beta_{15} L^5_{i-1} + \beta_{16} L^6_{i-1} + \beta_{17} L^1_{i-2} + \beta_{18} L^2_{i-2} + \beta_{19} L^3_{i-2} + \beta_{20} L^4_{i-2} \\ &+\beta_{21} L^5_{i-2} + \beta_{22} L^6_{i-2} \end{aligned}$$

Variables with an importance > 0.5 in either of the two models are $L^1, L^2, L^5, L^3_{i-1}, L^5_{i-1}, L^6_{i-1}, L^2_{i-2}$.

The above considerations yield to the following final (additive) Quasi-Poisson model which has been used to obtain the main results in Table 1:

$$\begin{aligned} E(Y) = \exp(&\beta_0 + \beta_1 \bar{A}^2 + \mathbf{L}^* \boldsymbol{\beta}_2 + f_1(Y_{i-1}) + f_2(Y_{i-3}) + f_3(Y_{i-7}) + f_4(Y_{i-9}) + f_5(L^1) + f_6(L^2) + f_7(L^5) \\ &+f_8(L^3_{i-1}) + f_9(L^5_{i-1}) + f_{10}(L^6_{i-1}) + f_{11}(L^2_{i-2})) \end{aligned} \quad (2)$$

In the above equation, $\mathbf{L}^* = (L^7, L^8, L^9, L^{10}, L^{11}, L^{12})$. With $f(\cdot)$ we refer to unspecified smooth functions which we fit with penalized splines, as implemented in the *R*-package `mgcv` (Wood, 2017). The estimate $\exp(\hat{\beta}_1)$ is the one reported in Table 1.

[1] We have overdispersion, i.e. the ratio of the deviance to the degrees of freedom is about 1.4. The estimated dispersion parameter for quasi poisson family is 1.39935.

[2] with *R*-package `MuMIn` (Barton, 2017)

Targeted Maximum Likelihood Estimation (in our analysis). From a causal perspective we are interested in the counterfactual outcome $Y_i^{A=a}$ which refers to the hypothetical outcome that would have been observed if at day i there had been, possibly contrary to the fact, the intervention $A = a$, i.e. load shedding or not. More generally we would like to know the expected number of admissions per day for a particular intervention, $E(Y^a)$. Since A is binary, a sensible target quantity is

$$\psi \quad = \quad E(Y^1) - E(Y^0),$$

that is the average treatment effect meaning that we are interested in the difference in expected number of admissions per day had load shedding being implemented during the whole year compared to if this had not been the case. If we look at $\bar{A}^2$ rather than A we can interpret ψ as the difference in expected number of admissions per day had there been a load shedding event each day or on any of the preceding two days, during the whole year, compared to if this had not been the case. This causal parameter can be identified under the following assumptions:

1. Consistency: if A is binary, then $Y_i = A_i Y_i^1 + (1 - A_i)Y_i^0$.
2. Positivity: $P(A = a|L = l) > 0$ for $\forall l$ with $P(L = l) \neq 0$.
3. Conditional exchangeability: $Y^a \coprod A|L$ for $\forall A = a, L = l$.

Consistency can be interpreted as having a well-defined intervention, that can't be interpreted/implemented in multiple ways. This assumption could be met in our data as implementation of load shedding as a temporary power-shutdown is well-defined in the sense that the consequence of load shedding is the unavailability of electricity. Positivity requires a positive probability of treatment assignment in all confounder strata, i.e. a positive probability of load shedding occurrence no matter what season and weather. There are no practical or theoretical considerations which would point towards violation of this assumption, however there may be practical positivity violations in the sample data because of the small sample size. An indication of possible positivity violations would be very small ($<< 0.01$) estimated probabilities of treatment assignment used in the TMLE procedure (see below). In our data probabilities varied between 0.01781 and 0.89374 showing no signs of severe practical positivity violations. Conditional exchangeability would be violated if there are unmeasured confounders. Our DAG (Figure 1, main manuscript) explains why we believe we have measured the main confounders, though we can't exclude the possibility of unmeasured confounders. Moreover, we consider the seasonal trend to be modeled correctly; if this is not the case both conditional exchangeability and the (conditional) independence assumption needed for the modeling (i.e. the validity of the likelihood functions used and application of super learning, see below) could be violated and this may introduce bias.

The theory and application of TMLE has been described elsewhere[3]. Briefly, TMLE requires the fitting of both the conditional expectation of the outcome $E(Y|A, \mathbf{L})$ and the treatment mechanism $P(A = 1|\mathbf{L})$. Modeling the conditional outcome enables standardization with respect to the confounders, i.e. integrating $\mathbf{L}$ out, which equates to the (parametric) g-formula:

$$E(Y^a) \quad = \quad \int_l E(Y|A = a, L = l)\, dF_l(l)$$

where $F_L(\cdot)$ is the cumulative distribution function with respect to L. Thus, the average treatment effect ψ can be obtained by using the g-formula to calculate $E(Y^1)$ and $E(Y^0)$. Practically this equates to fitting an appropriate regression model and predicting the outcome for the whole sample for both $A = 1$ and $A = 0$[4]. TMLE adds an additional *targeted* step whereby the initial estimate is updated by a regression model which contains i) a fixed intercept/offset, which is the initial estimate of the ATE (from the g-formula) and ii) a clever covariate which is a function of the inverse probabilities of treatment assignment (based on $P(A = 1|\mathbf{L})$). This update reduces bias (if present) and improves efficiency (if no bias present). TMLE can readily incorporate machine learning to estimate $E(Y|A, \mathbf{L})$ and $P(A = 1|\mathbf{L})$ while retaining valid inference. We have implemented TMLE with super learning, a combination of machine learning and statistical forecasting.

Super learning means considering a set of prediction algorithms, for example regression models, shrinkage estimators or boosting. Instead of choosing the algorithm with the smallest cross validation error, super learning chooses a weighted

[3] Good introductions for epidemiologists can be found in Luque Fernandez et al. (2018) and Schuler and Rose (2017)

[4] The empirical distribution, i.e. the data, is taken for $L = l$

combination of different algorithms, that is the weighted combination which minimizes the cross validation error. It can be shown that this weighted combination will perform (asymptotically) at least as good as the best algorithm, if not better (Van der Laan et al., 2008) and this is known as the oracle property of super learning. Briefly, super learning works as follows:

1. First split the data into blocks of equal size (i.e. ten blocks of 100 observations for a sample size of 1,000 units and the choice of 10-fold cross-validation) and fit each of the selected algorithms on the training set (i.e. on 9 out of the 10 blocks).

2. Then, predict the estimated probabilities of the outcome (Y) using the validation set (i.e. the remaining one block) for each algorithm.

3. Repeat steps 1 and 2 for each of the ten blocks. This yields predictions for all 1,000 observations for each learning algorithm.

4. Now, estimate the cross validated risk for each learning algorithm, that is a function of the true values of Y and the respective predictions, for example the (vector of the) squared differences.

5. Then, use non-negative least squares estimation to find the weighted linear combination of cross validated risks (related to each learner) which predicts Y best. Note that the weights sum up to one.

6. Then, use the weights to create a weighted prediction from the different learning algorithms. This yields the super learner estimate of $E(Y|A,\mathbf{L})$ and $P(A=1|\mathbf{L})$ respectively.

In our analysis we have implemented TMLE with the following specifications [for the main result in Table 1]:

- Our outcome Y is the number of hospital admissions, $A=\bar{A}^2$ denotes whether load shedding occurred on the same or any of the two preceding days, and the considered confounders are (based on the model building considerations above, i.e. variable reduction with model averaging for computational feasibility) $L^1, L^2, L^5, L^3_{i-1}, L^5_{i-1}, L^6_{i-1}$, L^2_{i-2}, L^7, L^8, $L^9, L^{10}, L^{11}, L^{12}, Y_{i-1}, Y_{i-3}, Y_{i-7}, Y_{i-9}$.

- Our target quantity is ψ, as discussed above.

- We use the R-package `tmle` to implement super learning (Gruber and van der Laan, 2012).

- We use super learning to estimate $E(Y|A,\mathbf{L})$ and $P(A=1|\mathbf{L})$ with 10-fold cross validation, a squared loss function (for cross validation), and a truncation level of 0.01, meaning that estimated probabilities $P(A=1|\mathbf{L})$ (needed for the targeted update step) would have been truncated if they were lower than 0.01 (though this did not occur). We use the following learners to estimate $E(Y|A,\mathbf{L})$: a full linear regression model, a linear regression model with stepwise AIC based model selection, a linear regression model with AIC based forward selection, GLM's based on an EM-algorithm-Bayesian model fitting (Gelman and Su, 2016), the arithmetic mean, a linear regression model with stepwise AIC based model selection including interaction terms, LASSO estimation (Tibsharani, 1996), LASSO averaging (Schomaker, 2012) for the full model and models with interactions and squared terms (Schomaker, 2017), boosting (Ridgeway, 2017), multivariate adaptive regression splines (Milborrow, 2017), Mallow's model averaging (Hansen, 2007), Jackknife Model Averaging (Hansen and Racine, 2012), linear regression after screening variables with LASSO, multivariate adaptive regression splines after screening variables with LASSO, and boosting after screening with Cramer's V (Heumann et al., 2016, Chapter 4). The same learners have been used to estimate $P(A=1|\mathbf{L})$ except Mallow's Model Averaging, Jackknifel Model Averaging, and the LASSO. In addition k-nearest neighbour classification has been used.

The Negative Binomial Model. Recall the Quasi-Poisson model used to obtain the results reported in Table 1:

$$\begin{aligned} E(Y) &= \exp(\beta_0+\beta_1\bar{A}^2+\mathbf{L}^*\boldsymbol{\beta}_2+f_1(Y_{i-1})+f_2(Y_{i-3})+f_3(Y_{i-7})+f_4(Y_{i-9})+f_5(L^1)+f_6(L^2)+f_7(L^5) \\ &\quad +f_8(L^3_{i-1})+f_9(L^5_{i-1})+f_{10}(L^6_{i-1})+f_{11}(L^2_{i-2})) \end{aligned}$$

An alternative to using the above model, would still be using the above model, but instead of using the assumption that the observations Y_i follow a Poisson distribution (conditional on the covariates), one can assume that they follow a negative

binomial distribution:

$$P(Y = y) \quad = \quad \frac{\Gamma(\phi + y)}{\Gamma(y + 1)\Gamma(\phi)} \left(\frac{\phi}{\phi + \lambda}\right)^{\phi} \left(\frac{\lambda}{\phi + \lambda}\right)^{y}$$

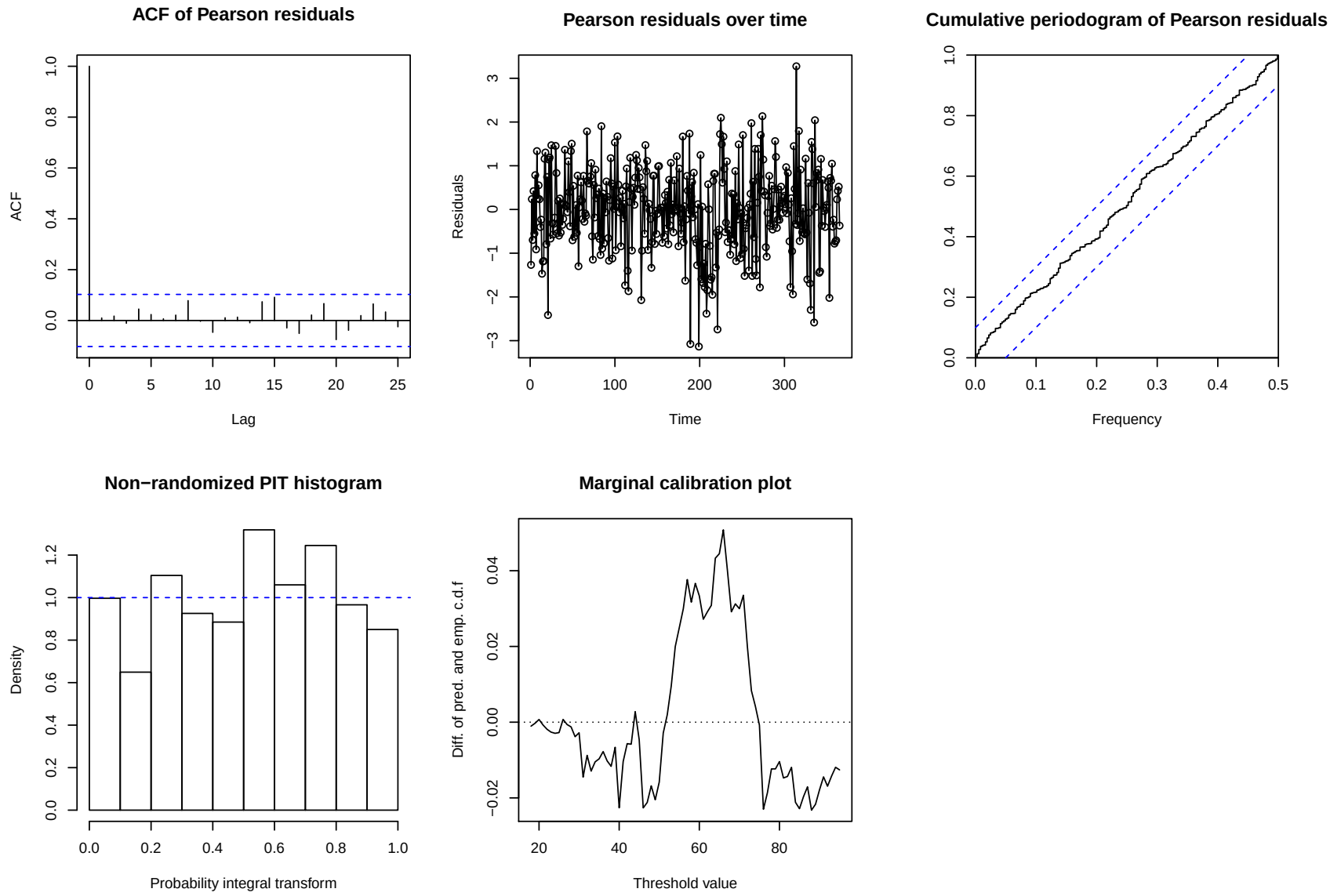
Note that $\Gamma(n)$ is the Gamma function, defined as $\Gamma(n) = (n-1)!$ for positive integers and $\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t)dt$ otherwise. For $\phi \to \infty$ this distribution approaches a Poisson distribution. While for a Poisson distribution we get a variance of λ, it is $\lambda + \lambda^2/\phi$ for the negative binomial approach, thus allowing to adjust for overdispersion. Our results are based on the negative binomial model as explained above, implemented with the `glm.nb` from the library `MASS`. No splines have been used to include the covariates.

The INGARCH model. The integer-valued GARCH model, also known as the autoregressive Poisson model, is similar to the models used above – with the difference that the mean conditional on the past ($E(Y_i|\mathcal{F}_i)$) is modeled, i.e. we explicitly acknowledge the time-series structure by modeling a stochastic process conditional on the past. With $\mathcal{F}_i$ we mean the history of the joint process $\{Y_i, A_i, \mathbf{L}_i\}$. In our analysis, we use the following INGARCH model for which we assume a negative-binomial distribution of Y_i:

$$E(Y) \quad = \quad \exp(\beta_0 + \alpha_1 \lambda_{i-1} + \alpha_2 \lambda_{i-2} + \alpha_3 \lambda_{i-3} + \beta_1 \bar{A}^2 + \tilde{\mathbf{L}}\boldsymbol{\beta}_2)$$

where $\tilde{\mathbf{L}}$ contains all variables from equation (2) and $\lambda_i = E(Y_i|\mathcal{F}_i)$. The decision to include the conditional means up to a lag of 3 has been based on model selection with AIC. Thus, the INGARCH model is identical to our negative-binomial model above, with the exception of the inclusion of three conditional means. Background on inference of these models, as well as good references, can be found in Liboschik et al. (2017). This reference also explains useful diagnostic for the INGARCH model, all of which are shown in eFigure 6. The autocorrelation function of the residuals does not exhibit any serial correlation that hasn't been taken into account by the model (as the autocorrelation is small and below the blue limits). There is also no sign that the residuals are inappropriate. However, the probability integral transform (PIT) histogram is not ideal, it should be more uniformly distributed. Moreover, the marginal calibration plot, which plots the difference between the average predictive cumulative distribution function (cdf) and the empirical cdf, shows major deviations from zero. This means that the number of hospital admissions, given the past, is not modelled well by the INGARCH model. It is also worth mentioning that with the INGARCH approach we haven't been able to include non-linear relationships with splines.

eFigure 6. Diagnostics for the fitted INGARCH model.



References

Barton, K. (2017). *MuMIn: Multi-Model Inference*. R package version 1.40.1/r411.

Buckland, S. T., K. P. Burnham, and N. H. Augustin (1997). Model selection: an integral part of inference. *Biometrics 53*, 603–618.

Burnham, K. and D. Anderson (2002). *Model selection and multimodel inference. A practical information-theoretic approach*. Springer, New York.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society A 158*, 419–466.

Gelman, A. and Y.-S. Su (2016). *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. R package version 1.9-3.

Gruber, S. and M. J. van der Laan (2012). tmle: An R package for targeted maximum likelihood estimation. *Journal of Statistical Software 51*(13), 1–35.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica 75*, 1175–1189.

Hansen, B. E. and J. Racine (2012). Jackknife model averaging. *Journal of Econometrics 167*, 38–46.

Heumann, C., M. Schomaker, and Shalabh (2016). *Introduction to Statistics and Data Analysis - With Exercises, Solutions and Applications in R*. Heidelberg: Springer.

Hjort, L. and G. Claeskens (2003). Frequentist model average estimators. *Journal of the American Statistical Association 98*, 879–945.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: a tutorial. *Statistical Science 14*, 382–417.

Liboschik, T., K. Fokianos, and R. Fried (2017). tscount: An R package for analysis of count time series following generalized linear models. *Journal of Statistical Software 82*(5), 1–51.

Luque Fernandez, M. A., M. Schomaker, B. Rachet, and M. E. Schnitzer (2018). Targeted maximum likelihood estimation for a binary treatment: A tutorial. *Statistics in Medicine in press*.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models*. Chapman and Hall/CRC.

Milborrow, S. (2017). *earth: Multivariate Adaptive Regression Splines*. R package version 4.6.0.

Ridgeway, G. (2017). *gbm: Generalized Boosted Regression Models*. R package version 2.1.3.

Schomaker, M. (2012). Shrinkage averaging estimation. *Statistical Papers 53*(4), 1015–1034.

Schomaker, M. (2017). *MAMI: Model Averaging (and Model Selection) after Multiple Imputation*. R package version 0.9.10.

Schuler, M. S. and S. Rose (2017). Targeted maximum likelihood estimation for causal inference in observational studies. *Amreican Journal of Epidemiology 185*(1), 65–73.

Tibsharani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B 58*, 267–288.

Van der Laan, M., E. Polley, and A. Hubbard (2008). Super learner. *Statistical Applications in Genetics and Molecular Biology 6*, Article 25.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R* (2 ed.). Chapman and Hall/CRC.