

A Difference-in-Differences Approach to Assess the Effect of a Heat Action Plan on Heat-Related Mortality, and Differences in Effectiveness According to Sex, Age, and Socioeconomic Status (Montreal, Quebec)

Tarik Benmarhnia,¹ Zinzi Bailey,¹ David Kaiser,² Nathalie Auger,³ Nicholas King,^{4,5} and Jay S. Kaufman^{1,5}

¹Institute for Health and Social Policy, McGill University, Montreal, Quebec, Canada; ²Direction de santé publique de l'Agence de la santé et des services sociaux de Montréal, Montréal, Québec, Canada; ³Institut National de Santé Publique du Québec, Montréal, Québec, Canada; ⁴Biomedical Ethics Unit, and ⁵Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada

BACKGROUND: The impact of heat waves on mortality and health inequalities is well documented. Very few studies have assessed the effectiveness of heat action plans (HAPs) on health, and none has used quasi-experimental methods to estimate causal effects of such programs.

OBJECTIVES: We developed a quasi-experimental method to estimate the causal effects associated with HAPs that allows the identification of heterogeneity across subpopulations, and to apply this method specifically to the case of the Montreal (Quebec, Canada) HAP.

METHODS: A difference-in-differences approach was undertaken using Montreal death registry data for the summers of 2000–2007 to assess the effectiveness of the Montreal HAP, implemented in 2004, on mortality. To study equity in the effect of HAP implementation, we assessed whether the program effects were heterogeneous across sex (male vs. female), age (≥ 65 years vs. < 65 years), and neighborhood education levels (first vs. third tertile). We conducted sensitivity analyses to assess the validity of the estimated causal effect of the HAP program.

RESULTS: We found evidence that the HAP contributed to reducing mortality on hot days, and that the mortality reduction attributable to the program was greater for elderly people and people living in low-education neighborhoods.

CONCLUSION: These findings show promise for programs aimed at reducing the impact of extreme temperatures and health inequities. We propose a new quasi-experimental approach that can be easily applied to evaluate the impact of any program or intervention triggered when daily thresholds are reached.

CITATION: Benmarhnia T, Bailey Z, Kaiser D, Auger N, King N, Kaufman J. 2016. A difference-in-differences approach to assess the effect of a heat action plan on heat-related mortality, and differences in effectiveness according to sex, age, and socioeconomic status (Montreal, Quebec). *Environ Health Perspect* 124:1694–1699; <http://dx.doi.org/10.1289/EHP203>

Introduction

Heat waves are a public health concern that is likely to grow in importance in the context of climate change (Huang et al. 2012; Li et al. 2013). The impacts of heat waves on mortality have been extensively documented (Åström et al. 2011; Basagaña et al. 2011; Basu 2009; Borrell et al. 2006; Bridger et al. 1976; Ellis 1972; Jones et al. 1982; Rooney et al. 1998). The 1995 Chicago, Illinois, heat wave resulted in 700 excess deaths over only a few days (Semenza et al. 1996). In Europe, around 15,000 excess deaths were reported during a heat wave in the first week of August 2003 (Robine et al. 2008). More recently, approximately 55,000 excess deaths (including potential contributing effects of wildfire smoke) were associated with the 2010 heat wave in Russia (Barriopedro et al. 2011). In response to such events, public health authorities have developed programs aimed at reducing heat wave-related health effects, including heat action plans (HAPs).

HAPs include early alerts and advisories combined with emergency public health measures to reduce heat-related morbidity and mortality (Kovats and Hajat 2008; Kovats and Kristie 2006; Lowe et al. 2011; McGregor

et al. 2015). HAPs are activated when meteorological conditions meet local criteria for classification of hot days, each jurisdiction having its own precise criteria (Tong et al. 2010). Variations exist in terminology to describe such public health programs (Toloo et al. 2013). In this paper, we use the term “HAPs” to refer to such programs.

The content of HAPs is generally based on existing empirical evidence of risk and protective factors for heat-related health impacts. In the absence of evaluations of actual HAP implementation, however, it remains unknown whether the various activities carried out in a HAP have an effect on heat-related mortality and morbidity. Such evidence would help inform future interventions aimed at reducing heat-related health impacts. Evaluating the effectiveness of public health programs is essential for several reasons: *a*) to assess whether policies effectively prevent and control mortality and morbidity; *b*) to evaluate whether programs are ethical, a good use of public resources, and efficient in reducing social inequalities; and *c*) to determine whether programs should be renewed or improved at the end of an initial trial period.

Very few studies have assessed the effectiveness of HAPs on health. In two recent reviews (Boeckmann and Rohn 2014; Toloo et al. 2013), only seven studies worldwide assessed the effectiveness of HAPs in reducing mortality. Six studies found that fewer people died of heat-related complications after the implementation of a HAP than was expected had the program not been in place (Chau et al. 2009; Ebi et al. 2004; Fouillet et al. 2008; Palecki et al. 2001; Tan et al. 2007; Weisskopf et al. 2002), whereas one study did not report an association (Morabito et al. 2012). Some of these studies were descriptive, using a pre–post approach to estimate the reduction in mortality attributable to the HAP by comparing the number of deaths on hot days before HAP implementation with the number after implementation. A pre–post approach may document variation in mortality, but, because it does not account for potential confounding (e.g., factors other than HAPs that are associated with heat-related mortality and have changed over time), does not provide evidence of causal effects associated with the program, which is fundamental to formulating and implementing effective HAPs. Other studies (Ebi et al. 2004; Morabito et al. 2012) used control groups (e.g., other cities) and controlled for measured confounders, but still may have been biased by unmeasured confounders. To deal with unmeasured confounders, quasi-experimental methods

Address correspondence to T. Benmarhnia, Institute for Health and Social Policy, McGill University, Meredith, Charles, House, 1130 Pine Ave. West, Montreal, Quebec H3A 1A3 Canada. Telephone: 15146384393. E-mail: tarik.benmarhnia@mcgill.ca

This work is supported by the Montreal Health Equity Research Consortium which finances the postdoctoral awards for T.B. and Z.B.

The authors declare they have no actual or potential competing financial interests.

Received: 10 July 2015; Revised: 14 October 2015; Accepted: 4 May 2016; Published: 20 May 2016.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

have been developed for observational studies as alternatives to experimental methods and are used to provide estimates from observational studies. The term “quasi-experiment” refers to the analysis of an observed treatment in relation to an outcome measure in which assignment to the treatment is not made by physical randomization, as in a true experiment, but is instead assigned according to some mechanism that can be argued to be independent of the potential outcome of each experimental unit. Using the counterfactual framework, natural experiments can be used to estimate the causal effect of specific policies on population health.

The European heat wave of 2003 prompted the Montreal Public Health Department (PHD) to develop a HAP to reduce heat-related mortality and morbidity (Price et al. 2013). Implemented in 2004, the HAP comprises a spectrum of interventions triggered at different alert levels based on temperature forecasts of Environment Canada. Important interventions undertaken as part of the HAP by the Montreal PHD and partners are triggered at the alert level (called “Active watch” level in the Montreal HAP), when daily temperatures exceed 30°C (Price et al. 2013). These interventions include, for example, public advisories via different media about preventive measures, and intensified surveillance and implementation of preventive measures in health care facilities (Appendix 1). The alert level is followed by two other levels, associated with additional interventions such as extended operational hours of public pools and opening of air-conditioned shelters. The HAP is active on days defined as hot, and not active on other summer days. The program is automatically triggered by meteorological conditions, and targets the entire population of the Island of Montreal. Although implemented in 2004, mortality attributable to heat continues to be observed in Montreal (Benmarhnia et al. 2014b), with 105 heat-related deaths occurring during a 5-day heat wave in 2010 (Price et al. 2013) suggesting the need to evaluate this program. Evidence about the effectiveness of the HAP is essential to improve this program.

The impact of heat waves on mortality is uneven, and economically or socially disadvantaged populations are at higher risk for heat-related death. Potential factors associated with vulnerability to heat-related mortality, including age, sex, education, and neighborhood socioeconomic status (SES), have been documented (Benmarhnia et al. 2015a; Gronlund 2014). It is thus worthwhile to assess whether HAPs are effective in reducing or eliminating unequal health outcomes associated with heat. However, an assessment of the heterogeneity of potential benefits across subpopulations or territories

(e.g., neighborhoods) is lacking in the policy evaluation literature (Bauman et al. 2014; Benmarhnia et al. 2014a).

The equity impact of a policy, defined as the differential reduction of mortality between vulnerable and nonvulnerable populations, can be assessed by using a quasi-experimental method to identify heterogeneity in potential benefits among subpopulations or territories. We define as “equitable” any policy that reduces inequality in which vulnerable populations start from a comparatively worse-off position. For example, a policy that reduces inequalities in mortality between the elderly and other age groups (e.g., daily mortality differences between the two age groups) is by our definition equitable, if the elderly experienced higher mortality rates before its implementation. We stress, however, that one can employ alternative definitions of equity when using our method.

Using a difference-in-differences approach (a quasi-experimental method that can be used to estimate the causal effects of public health policies) (Angrist and Pischke 2008; Basu et al. 2016), we aimed to estimate the causal effect of the HAP program on mortality in Montreal. We assessed whether the HAP program had a heterogeneous effect across sex, age, and neighborhood socioeconomic status (SES).

Methods

Data Sources

The study population included all residents of the Island of Montreal who died during the months of June, July, and August between 2000 and 2007, inclusively, and were listed in the provincial death registry. We restricted the population to non-accidental causes of death, excluding deaths for *International Classification of Diseases, 9th Revision* (ICD-9) codes 800–999 (injury and poisoning) and *10th Revision* (ICD-10) codes S00–T98 (injury, poisoning, and other consequences of external causes). Individual information on age, sex, date of death, and census tract of residence was available for each death. Deaths were aggregated daily. We used the percentage of residents > 20 years of age without a high school diploma in the census tract of residence (at time of death) to assign a “neighborhood SES” to each participant in the study; data were extracted from the 2006 census (Benmarhnia et al. 2015b). We generated three strata of neighborhood SES by dividing the data into tertiles, and compared the first and third tertiles in the analysis.

We obtained maximum outdoor temperatures for the months of June through August between 2000 and 2007 from the Environment Canada meteorological observation station at the Montreal Pierre Elliott Trudeau International Airport (Environment

Canada 2015). The meteorological observation station used in this study is also used for activation of the Montreal HAP.

Study Design and Statistical Analysis

We used a difference-in-differences (DID) design, in which daily mortality for the population of Montreal was compared before and after HAP implementation in 2004. Summer days (June–August) were classified as “eligible” if they were hot enough to have triggered an “active watch” alert level according to the HAP (i.e., if they would have been eligible for a HAP intervention), or as “non-eligible” otherwise, regardless of whether they occurred during 2000–2003 (before the HAP was implemented) or during 2004–2007 (after the HAP was implemented). The counterfactual quantity of interest is the difference in the daily number of deaths between eligible (hot) and non-eligible (non-hot) days that would have occurred during 2004–2007 if the HAP had not been implemented. Therefore, the causal effect of the HAP on daily mortality was estimated as the difference between two values: *a*) the difference in number of deaths on eligible (hot) days before and after implementation of the HAP, and *b*) the difference in number of deaths on ineligible (non-hot) days before and after implementation of the HAP. We thus obtained a DID estimate to represent the causal effect of the HAP program on daily mortality in Montreal.

The DID approach will generate a valid estimate of the causal effect if the implementation of the HAP was the only factor that might cause a change in the association between heat and mortality before and after HAP implementation. We chose a short interval of time (4 years before and 4 years after initiation of HAP), to limit potential confounding due to population acclimatization and urban infrastructure changes (Petkova et al. 2014) that might reduce the effect of hot days on mortality during the post-HAP time period. We also compared the distributions (using a Kolmogorov–Smirnov test for equality of distribution functions) of maximum daily temperatures on eligible and non-eligible days during each time period to determine whether, for example, higher maximum daily temperatures on eligible days during the post-HAP time period might bias our estimate of the effect of the HAP by increasing mortality on eligible days.

We estimated the HAP effect on daily mortality using a time series analysis in DID Poisson models adjusted to account for temporal patterns (month, week, and day). To calculate the daily number of deaths “prevented” during hot days after HAP implementation, we predicted the Poisson

count from the adjusted models using the DID estimate. The DID estimate represents the product interaction term between the estimated effect of eligible (hot) versus non-eligible (non-hot) days and the estimated effect of the pre-HAP (2000–2003) versus post-HAP (2004–2007) time period. A positive value of the DID estimate generated with this method represents an estimated reduction of daily mortality attributable to the HAP, because we compared pre-HAP deaths to post-HAP deaths. We estimated 95% confidence intervals (CIs) for the adjusted DID estimate by bootstrapping (1,000 samples) (Carpenter and Bithell 2000).

As sensitivity analyses, we tested the validity of the causal model by conducting the same analyses above, but defining HAP at arbitrary implementation periods: *a*) in 2000 and *b*) in 2002. We compared differences in mortality on eligible and non-eligible days between 4-year time periods that were unrelated to the HAP intervention: *a*) 1996–1999 versus 2000–2003; and *b*) 1998–2001 versus 2002–2005. We then tested different definitions of hot days (thus modifying the assignment of “eligible” and “non-eligible” days) in order to test the validity of the causal model. We also used a definition of hot days as maximum temperatures above 28°C and above 32°C, expecting no effect of the HAP on mortality reduction. We considered the cumulative heat effect (lag 0–5: up to 5 consecutive hot days) (Åström et al. 2011), expecting a larger effect of the HAP on mortality reduction by considering consecutive hot days. We selected this lag period based on previous studies in Montreal (Goldberg et al. 2011). We also calculated the DID estimate considering a harvesting effect. This is the hypothesis that some frail individuals who died during a hot day would have died

in the subsequent days or weeks even without the hot day. These deaths are strictly speaking not caused by the hot day, but only hastened by it. To estimate this effect, we calculated a displacement ratio after 15 days (Saha et al. 2014). We then multiplied the displacement ratio we obtained with the DID estimate to get a DID estimate that takes into account the harvesting effect. Finally, we calculated the DID estimate by restricting “non-eligible” days to maximum temperatures between 25°C and 29°C.

Equity of Program Impact

To study equity in implementation, we assessed whether the HAP had a heterogeneous effect by sex, age, and neighborhood SES. We first dichotomized each of the variables: *a*) sex: men vs. women; *b*) age (≥ 65 years vs. < 65 years); *c*) SES: first versus third tertile of neighborhood education. We then calculated daily mortality differences between each of the two categories (e.g., daily mortality among men minus daily mortality among women). The HAP effect on daily mortality differences was estimated following the same method above. Thus, to assess heterogeneity in the program effect, we calculated differences-in-differences-in-differences (DIDID) estimates using a DID estimate (i.e., interaction term) as above, but considering as the health outcome a daily difference between two groups.

Results

Descriptive Statistics

During the study period (2000–2007), 28,137 non-accidental deaths occurred during the months of June–August, including 14,779 among women and 13,358 among men. The average number of daily deaths

was 38.70, with a minimum of 16 deaths per day, and a maximum of 73 for the whole period. During this interval, 75 heat-wave days occurred, including 39 before the HAP and 36 after.

Table 1 presents descriptive statistics before and after HAP implementation for non-eligible and eligible groups. The number of eligible hot days was similar between the two periods (39 and 36 days, respectively). Among eligible and non-eligible days, we did not find a difference in maximum temperature distribution between before- and after-program periods (Table 1). We found no difference in mortality distribution among non-eligible days before and after HAP implementation.

Estimated Effect of HAP on Mortality

The adjusted DID model estimated that over the entire population, the HAP program reduced mortality during hot days by 2.52 deaths per day (95% CI: $-0.34, 5.38$). This result suggests that HAP may have helped reduce mortality during hot days.

Estimated Effect of HAP on Equity

Adjusted estimates for the heterogeneous effect by sex, age, and SES are presented in Table 2. We estimated that HAP reduced differences in mortality between the elderly (≥ 65 years) and non-elderly (0–64 years) during hot days by 2.44 deaths per day (95% CI: 0.27, 4.59). We also estimated that HAP program reduced differences in mortality between individuals living in neighborhoods with low SES and those living in neighborhood with high SES (the middle tertile of SES was not included in the analysis) during hot days by 2.48 deaths per day (95% CI: 0.69, 4.27). However we did not find much evidence of heterogeneity in the program

Table 1. Descriptive statistics before and after heat action plan program implementation for non-eligible and eligible days.

Variables ^a	Before program (2000–2003)		After program (2004–2007)	
	Non eligible group <i>n</i> = 329 days	Eligible group <i>n</i> = 39 days	Non eligible group <i>n</i> = 332 days	Eligible group <i>n</i> = 36 days
Median maximum temperature [°C (IQR)]	25 (22, 27)	32 (31, 33)	25 (22, 27)	31 (30, 32)
Number of hot days (%)	—	39 (11)	—	36 (10)
Daily number of deaths (IQR)	39 (36, 44)	44 (39, 52)	36 (32, 41)	39 (33, 43)
Daily number of deaths among men (IQR)	19 (16, 22)	20 (17, 24)	17 (14, 20)	18 (15, 22)
Daily number of deaths among women (IQR)	20 (18, 24)	24 (20, 29)	19 (16, 22)	21 (17, 25)
Daily number of deaths among elderly (≥ 65 years) (IQR)	32 (27, 35)	37 (32, 42)	29 (24, 32)	33 (28, 35)
Daily number of deaths among non-elderly (< 65 years) (IQR)	7 (5, 9)	7 (4, 9)	7 (5, 8)	6 (5, 8)
Daily number of deaths among low-SES group (IQR)	16 (14, 19)	20 (16, 22)	15 (12, 17)	16 (13, 18)
Daily number of deaths among high-SES group (IQR)	12 (10, 14)	13 (9, 16)	11 (9, 13)	12 (10, 14.5)
Daily number of daily differences in mortality according to sex ^b (IQR)	2 (–2, 6)	4 (–1, 8)	2 (–2, 6)	4 (–2, 7)
Daily number of daily differences in mortality according to age ^c (IQR)	24 (19, 28)	30 (26, 35)	21 (18, 26)	25.5 (20, 30)
Daily number of daily differences in mortality according to SES ^d (IQR)	4 (1, 8)	6 (1, 12)	4 (1, 7)	3 (–1, 6)

^aTo compare maximum temperature distributions among eligible and non-eligible days between the two periods, we conducted a Kolmogorov–Smirnov test for equality of distribution functions. This test confirmed that there was no strong evidence of a systematic difference between the distribution of daily temperatures between the two periods (*p*-value > 0.10). For each continuous variable, mean values are presented followed by their IQR.

^bDaily number of deaths among men – daily number of deaths among women.

^cDaily number of deaths among elderly (≥ 65 years) – daily number of deaths among non elderly (< 65 years).

^dDaily number of deaths among low-SES group (lowest tertile) – daily number of deaths among high-SES group (highest tertile).

effect according to sex, with a DID estimate of 1.38 (95% CI: -1.60, 4.36).

Sensitivity Analyses

Table 3 presents the adjusted DID estimates in sensitivity analyses. When we compared 4-year periods that were unrelated to HAP implementation, we did not find evidence of a difference in the effect of hot days on mortality between the earlier and later time periods based on the heterogeneity test. We also failed to find evidence of an effect on mortality when hot days were defined at a threshold of 28°C, thus misclassifying some days during which the Montreal HAP was not active as “eligible,” and when the threshold was 32°C, thus misclassifying some days when the HAP was active as “non-eligible. The latter result may be explained in part by the small number of hot days when using this definition (26 eligible days for the whole period). We then defined hot days according to cumulative heat (temperature on the day of death and up to the 5 days before death), and found that HAP program was effective in reducing daily mortality during hot days by 4.87 deaths per day (95% CI: 0.67, 8.20), a stronger effect, in comparison with 2.52 deaths per day (95% CI: -0.34, 5.38). Including a harvesting effect yielded a DID estimate of 1.87 (95% CI: 0.29, 3.47). Finally, restricting non-eligible days to maximum daily temperatures above 25°C yielded a DID estimate of 2.23 (95% CI: -0.80, 5.27), which is close to the point estimate of the main analysis but with a wider confidence interval.

Discussion

We found some evidence that the heat action plan implemented in Montreal in 2004 contributed to reduce mortality overall on hot days between 2004 and 2007. Furthermore, we observed that this program may have a positive effect in reducing some inequities in heat-related health impacts, because the estimated beneficial effects associated with the HAP were greater for elderly people and people living in low-SES neighborhoods. These findings show promise for further implementation of policies aimed at reducing the impact of extreme temperatures. Indeed, providing evidence about public health policies' effectiveness may be useful to justify such investments for stakeholders/political leaders, and inform other jurisdictions that are considering implementing such policies. In addition, we apply this approach to policy impact evaluation.

To our knowledge, this is the first study to use a quasi-experimental method to estimate the causal effects associated with a HAP. In addition to providing relevant findings regarding this policy's effectiveness in protecting populations, we feel that the

quasi-experimental methodological approach is a major contribution to environmental health research. The DID approach that we used allows researchers to estimate a policy's causal effect by using non-eligible days within the jurisdiction of interest, obviating the need for data from other cities or countries as control groups, the approach used in time series analyses with quasi-experimental methods (Cook et al. 2008). Our approach is distinct from previously published methods (Chau et al. 2009; Ebi et al. 2004; Fouillet et al. 2008; Morabito et al. 2012; Palecki et al. 2001; Tan et al. 2007; Weisskopf et al. 2002) because it controls for unmeasured confounders if the assumptions of the model are satisfied. For the DID method, the most important assumptions are the adequacy of the control group for the counterfactual contrast of interest and the absence of other interventions contemporaneous with the implementation of the new policy. In our setting where the control group was based on non-hot days in the same city, the validity of the causal estimates requires that *a*) there be no important time trends in the outcome of interest in the non-eligible group before and after policy implementation; and *b*) the distribution of the variable that defines the eligibility (temperature in this case) should not be different between the before and after periods.

Quasi-experimental methods have been widely used in recent years as alternatives to experimental methods and to observational designs that are affected by unmeasured confounders. They are designed to be more robust than pre-post comparisons that lack a control group (Angrist and Pischke 2008;

Basu et al. 2016). In addition, the DID approach used here is not necessarily more difficult to implement than methods used previously (pre-post approach with a control group), provided that DID assumptions are met. In addition, our approach can be applied to other policies in which implementation is based on a daily threshold. For example, policies aimed at reducing air pollution levels on “smog episode days” (Zivin and Neidell 2009) (e.g., adjusting speed limits, alternative-day driving) can be evaluated following the same approach.

We also proposed a DID approach to evaluate heterogeneity in the HAP effect, and its potential impact on health equity. Our results suggest that targeting specific populations vulnerable to extreme heat (or other hazards) may reduce health inequalities between vulnerable and comparison groups. This might be explained by various actions undertaken as part of the HAP (Appendix 1) targeting directly some populations identified as vulnerable, such as frequent visits to home care patients (including elderly individuals) or daily phone calls to home care patients. DID have been used in studies with other exposures and health outcomes (Currie et al. 2014; Harper et al. 2014) to provide information about heterogeneity in policy effects. With this method, we propose a complementary approach that can be applied to policy evaluation using a time-series design either with or without a control group, as long as the data meet the assumptions described above.

This study is particularly relevant considering recent recommendations noted (Woodward et al. 2014) on the need

Table 2. Estimated effect of the heat action plan program on equity.

Potential modifiers of the program benefits	Heterogeneity in the program effect ^a estimate	95% CI	p-Value ^b
Sex (men vs. women)	1.38	(-1.60, 4.36)	0.36
Age (≥ 65 vs. < 65 years)	2.44	(0.27, 4.59)	0.03
Neighborhood SES (lowest SES tercile vs. highest SES tercile)	2.48	(0.69, 4.27)	< 0.01

^aFrom DID (differences-in-differences-in-differences) estimates (Poisson model adjusted for temporal trends); 95% CIs were obtained by bootstrapping (1,000 samples).

^bp-Values are obtained from a Wald test on the interaction term (i.e., DID estimate considering as health outcome the daily difference between two groups).

Table 3. Sensitivity analyses for the estimated effects of the heat action plan program.

Sensitivity analyses	DID estimate	95% CI	p-Value ^a
Arbitrary programs			
Program implemented in 2000 ^b	0.94	(-2.08, 3.96)	0.54
Program implemented in 2002 ^c	0.42	(-3.62, 2.77)	0.80
Other hot days definitions			
When maximum temperature is above 28°C	0.58	(-1.77, 2.93)	0.63
When maximum temperature is above 32°C	2.79	(-2.65, 8.23)	0.32
Cumulative heat ^d	4.87	(0.67, 8.20)	0.03
Accounting for displacement ratio ^e	1.87	(0.29, 3.47)	0.02
Restriction to non-eligible days above 25°C	2.23	(-0.80, 5.27)	0.15

^ap-Values are obtained from a Wald test on the interaction term (i.e., DID estimate).

^bUsing mortality and temperature data for periods 1996–1999 vs. 2000–2003.

^cUsing mortality and temperature data for periods 1998–2001 vs. 2002–2005.

^dConsidering a cumulative heat effect up to 5 consecutive hot days (lag 0–5).

^eThe displacement ratio (Saha et al. 2014) was 0.65.

for evidence of effectiveness of HAPs in promoting health and reducing health inequalities. Many countries, regions, and cities have recently instituted HAPs to reduce heat-related mortality and morbidity (Lowe et al. 2011; Toloo et al. 2013); however, few have evaluated the effects associated with their implementation. We therefore encourage further studies to estimate the effect of HAPs on mortality in other contexts. Further, these methods can also be used to assess effects on other health outcomes, such as hospitalization for fluid and electrolyte disorders or heat stroke (Bobb et al. 2014).

There are some limitations to our study. First, we assessed the effect of Montreal's HAP without considering the spatial variability of benefits. Other studies have shown intra-city variability in health impacts of extreme heat (Hondula and Barnett 2014), and it is plausible that HAP benefits had a spatial pattern in our case as well. Further studies could conduct spatiotemporal analysis to address this gap. Because HAP measures have been triggered based on forecast daily temperatures, it is possible that some observed daily temperatures did not correlate with forecast data (Åström et al. 2014).

Appendix 1: Examples of Actions Pertaining to the Montreal Heat Action Plan

- Transmission of information coming from the Montreal PHD concerning levels of the plan that must be implemented in the health care network.
- Intensification of surveillance of signs and symptoms of heat-related illness, reminder of preventive measures to patients, distribution of water bottles.
- Air conditioning of common areas and opening of these areas during the day, the evening, and the night for patients in institutions.
- Frequent visits to home care patients.
- In institutions, frequent visits to housed patients.
- Distribution of water, refreshments, lighter meals.
- Monitoring of temperature in work areas, especially in warmer environments (e.g., kitchen, laundry room).
- Frequent work breaks for workers in hot, non-air-conditioned environments.
- Transfer of patients to common areas with air conditioning.
- Transfer of vulnerable home care patients to air-conditioned shelters.
- Daily contact by telephone or home visits to home care patients. Registry of calls and compilation of questionnaires for home evaluation.

In addition, by measuring temperature at a single site (i.e., the airport), we were not able to get within-city temperature variability that would inform us about exposure misclassification that might be correlated with neighborhood SES spatial variability (through micro-heat islands, for example). Another issue is redundancy between categories in the equity assessment (for example, prevented deaths among older individuals might be also counted in prevented deaths among individuals living in low-SES neighborhoods). We also did not distinguish actions undertaken > 30°C. Moreover, the Montreal HAP has been revised and updated yearly since 2007 (Price et al. 2013); this was not specifically considered in the present study. Further research could consider how incremental changes in policies affected mortality reduction, if at all. In addition, we did not qualitatively or quantitatively explore mechanisms through which the program contributed to reducing mortality during hot days. Implementation evaluation studies are a necessary complement to studies such as ours in producing a complete portrait of the effectiveness of a complex program such as a HAP.

Evidence of the effectiveness of HAPs is timely because the implementation of such programs has increased significantly across the world in recent years. Here, we provide evidence that a local program and a quasi-experimental methodology can help facilitate similar evaluations in other contexts.

REFERENCES

- Angrist JD, Pischke JS. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ:Princeton University Press.
- Åström C, Ebi KL, Langner J, Forsberg B. 2014. Developing a heatwave early warning system for Sweden: evaluating sensitivity of different epidemiological modelling approaches to forecast temperatures. *Int J Environ Res Public Health* 12:254–267.
- Åström DO, Forsberg B, Rocklöv J. 2011. Heat wave impact on morbidity and mortality in the elderly population: a review of recent studies. *Maturitas* 69:99–105.
- Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R. 2011. The hot summer of 2010: redrawing the temperature record map of Europe. *Science* 332:220–224.
- Basagaña X, Sartini C, Barrera-Gómez J, Davvand P, Cunillera J, Ostro B, et al. 2011. Heat waves and cause-specific mortality at all ages. *Epidemiology* 22:765–772.
- Basu R. 2009. High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. *Environ Health* 8:40. doi: 10.1186/1476-069X-8-40.
- Basu S, Meghani A, Siddiqi A. 2016. Evaluating the health impact of large-scale public policy changes: classical and novel approaches [Abstract]. *Annu Rev Public Health* 38:1. Available: <http://www.annualreviews.org/doi/abs/10.1146/annurev-publhealth-031816-044208> [accessed 22 September 2016].
- Bauman AE, King L, Nutbeam D. 2014. Rethinking the evaluation and measurement of health in all policies. *Health Promot In* 29(suppl 1):i143–i151.
- Benmarhnia T, Deguen S, Kaufman JS, Smargiassi A. 2015a. Vulnerability to heat-related mortality: a systematic review, meta-analysis and metaregression analysis. *Epidemiology* 26:781–793.
- Benmarhnia T, Grenier P, Brand A, Fournier M, Deguen S, Smargiassi A. 2015b. Quantifying vulnerability to extreme heat in time series analyses: a novel approach applied to neighborhood social disparities under climate change. *Int J Environ Res Public Health* 12:11869–11879.
- Benmarhnia T, Rey L, Cartier Y, Clary CM, Deguen S, Brousselle A. 2014a. Addressing equity in interventions to reduce air pollution in urban areas: a systematic review. *Int J Public Health* 59:933–944.
- Benmarhnia T, Sottile MF, Plante C, Brand A, Casati B, Fournier M, et al. 2014b. Variability in temperature-related mortality projections under climate change. *Environ Health Perspect* 122:1293–1298. doi: 10.1289/ehp.1306954.
- Bobb JF, Obermeyer Z, Wang Y, Dominic F. 2014. Cause-specific risk of hospital admission related to extreme heat in older adults. *JAMA* 312(24):2659–2667.
- Boeckmann M, Rohn I. 2014. Is planned adaptation to heat reducing heat-related mortality and illness? A systematic review. *BMC Public Health* 14:1112. doi: 10.1186/1471-2458-14-1112.
- Borrell C, Mari-Dell'Olmo M, Rodríguez-Sanz M, García-Olalla P, Caylà JA, Benach J, et al. 2006. Socioeconomic position and excess mortality during the heat wave of 2003 in Barcelona. *Eur J Epidemiol* 21:633–640.
- Bridger C, Ellis F, Taylor HL. 1976. Mortality in St. Louis, Missouri, during heat waves in 1936, 1953, 1954, 1955, and 1966: coroner's cases. *Environ Res* 12:38–48.
- Carpenter J, Bithell J. 2000. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 19(9):1141–1164.
- Chau PH, Chan KC, Woo J. 2009. Hot weather warning might help to reduce elderly mortality in Hong Kong. *Int J Biometeorol* 53:461–468.
- Cook TD, Shadish WR, Wong VC. 2008. Three conditions under which experiments and observational studies produce comparable causal estimates: new findings from within-study comparisons. *J Policy Anal Manage* 27(4):724–750.
- Currie J, Stabile M, Jones L. 2014. Do stimulant medications improve educational and behavioral outcomes for children with ADHD? *J Health Econ* 37:58–69.
- Ebi KL, Teisberg TJ, Kalkstein LS, Robinson L, Weiher RF. 2004. Heat watch/warning systems save lives: estimated costs and benefits for Philadelphia 1995–98. *Bull Am Meteorol Soc* 85:1067–1073.
- Ellis FP. 1972. Mortality from heat illness and heat-aggravated illness in the United States. *Environ Res* 5:1–58.
- Environment Canada. 2015. *Historical Climate Data*. Available: <http://www.climate.weatheroffice.gc.ca> [accessed 30 March 2015].
- Fouillet A, Rey G, Wagner V, Laaidi K, Empereur-Bissonnet P, Le Tertre A, et al. 2008. Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003? A study of the 2006 heat wave. *Int J Epidemiol* 37:309–317.
- Goldberg MS, Gasparrini A, Armstrong B, Valois MF. 2011. The short-term influence of temperature on daily mortality in the temperate climate of Montreal, Canada. *Environ Res* 111(6):853–860.
- Gronlund CJ. 2014. Racial and socioeconomic disparities in heat-related health effects and their mechanisms: a review. *Curr Epidemiol Rep* 1:165–173.
- Harper S, Strumpf EC, Burris S, Smith GD, Lynch J. 2014. The effect of mandatory seat belt laws on seat belt use by socioeconomic position. *J Policy Anal Manage* 33(1):141–161.

- Hondula DM, Barnett AG. 2014. Heat-related morbidity in Brisbane, Australia: spatial variation and area-level predictors. *Environ Health Perspect* 122:831–836, doi:10.1289/ehp.1307496.
- Huang C, Barnett AG, Wang X, Tong S. 2012. The impact of temperature on years of life lost in Brisbane, Australia [Letter]. *Nat Clim Change* 2:265–270.
- Jones TS, Liang AP, Kilbourne EM, Griffin MR, Patriarca PA, Wassilak SGF, et al. 1982. Morbidity and mortality associated with the July 1980 heat wave in St Louis and Kansas City, Mo. *JAMA* 247:3327–3331.
- Kovats RS, Hajat S. 2008. Heat stress and public health: a critical review. *Annu Rev Public Health* 29:41–55.
- Kovats RS, Kristie LE. 2006. Heatwaves and public health in Europe. *Eur J Public Health* 16:592–599.
- Li T, Horton RM, Kinney PL. 2013. Projections of seasonal patterns in temperature-related deaths for Manhattan, New York. *Nat Clim Chang* 3:717–721.
- Lowe D, Ebi KL, Forsberg B. 2011. Heatwave early warning systems and adaptation advice to reduce human health consequences of heatwaves. *Int J Environ Res Public Health* 8:4623–4648.
- McGregor GR, Bessemoulin P, Ebi K, Menne B, eds. 2015. *Heatwaves and Health: Guidance on Warning System Development*. WMO-No. 1142. Geneva:World Meteorological Organization, World Health Organization. Available: http://www.who.int/entity/globalchange/publications/WMO_WHO_Heat_Health_Guidance_2015.pdf [accessed 1 July 2015].
- Morabito M, Profili F, Crisci A, Francesconi P, Gensini GF, Orlandini S. 2012. Heat-related mortality in the Florentine area (Italy) before and after the exceptional 2003 heat wave in Europe: an improved public health response? *Int J Biometeorol* 56:801–810.
- Palecki MA, Changnon SA, Kunkel KE. 2001. The nature and impacts of the July 1999 heat wave in the Midwestern United States: learning from the lessons of 1995. *Bull Am Meteorol Soc* 82:1353–1367.
- Petkova EP, Gasparri A, Kinney PL. 2014. Heat and mortality in New York City since the beginning of the 20th century. *Epidemiology* 25:554–560.
- Price K, Perron S, King N. 2013. Implementation of the Montreal heat response plan during the 2010 heat wave. *Can J Public Health* 104:e96–e100.
- Robine JM, Cheung SL, Le Roy S, Van Oyen H, Griffiths C, Michel JP, et al. 2008. Death toll exceeded 70,000 in Europe during the summer of 2003. *CR Biol* 331:171–178.
- Rooney C, McMichael AJ, Kovats RS, Coleman MP. 1998. Excess mortality in England and Wales, and in Greater London, during the 1995 heatwave. *J Epidemiol Community Health* 52:482–486.
- Saha MV, Davis RE, Hondula DM. 2014. Mortality displacement as a function of heat event strength in 7 US cities. *Am J Epidemiol* 179:467–474.
- Semenza JC, Rubin CH, Falter KH, Selanikio JD, Flanders WD, Howe HL, et al. 1996. Heat-related deaths during the July 1995 heat wave in Chicago. *N Engl J Med* 335:84–90.
- Tan J, Zheng Y, Song G, Kalkstein LS, Kalkstein AJ, Tang X. 2007. Heat wave impacts on mortality in Shanghai, 1998 and 2003. *Int J Biometeorol* 51:193–200.
- Toloo G, FitzGerald G, Aitken P, Verrall K, Tong S. 2013. Evaluating the effectiveness of heat warning systems: systematic review of epidemiological evidence. *Int J Public Health* 58:667–681.
- Tong S, Wang XY, Barnett AG. 2010. Assessment of heat-related health impacts in Brisbane, Australia: comparison of different heatwave definitions. *PLoS One* 5:e12155, doi: 10.1371/journal.pone.0012155.
- Weisskopf MG, Anderson HA, Foldy S, Hanrahan LP, Blair K, Török TJ, et al. 2002. Heat wave morbidity and mortality, Milwaukee, Wis, 1999 vs 1995: an improved response? *Am J Public Health* 92:830–833.
- Woodward A, Smith KR, Campbell-Lendrum D, Chadee DD, Honda Y, Liu Q, et al. 2014. Climate change and health: on the latest IPCC report. *Lancet* 383(9924):1185–1189.
- Zivin JG, Neidell M. 2009. Days of haze: Environmental information disclosure and intertemporal avoidance behavior. *J Environ Econ Manage* 58(2):119–128.