Assessing the causal effects of a stochastic intervention in time series data: are heat alerts effective in preventing deaths and hospitalizations?

XIAO WU[*] (ID)

Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 W 168th St, New York, NY 10032, USA and Department of Biostatistics, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115, USA

xw2892@cumc.columbia.edu

KATE R. WEINBERGER

School of Population and Public Health, University of British Columbia, 2206 E Mall, Vancouver, BC V6T 1Z3, Canada

GREGORY A. WELLENIUS

Department of Environmental Health, Boston University School of Public Health, 715 Albany St, Boston, MA 02118, USA

FRANCESCA DOMINICI

Department of Biostatistics, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115, USA

DANIELLE BRAUN

Department of Biostatistics, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115, USA and Department of Data Science, Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02215, USA

SUMMARY

The methodological development of this article is motivated by the need to address the following scientific question: does the issuance of heat alerts prevent adverse health effects? Our goal is to address this question within a causal inference framework in the context of time series data. A key challenge is that causal inference methods require the overlap assumption to hold: each unit (i.e., a day) must have a positive probability of receiving the treatment (i.e., issuing a heat alert on that day). In our motivating example, the overlap assumption is often violated: the probability of issuing a heat alert on a cooler day is near zero. To overcome this challenge, we propose a *stochastic intervention* for time series data which is implemented via an incremental time-varying propensity score

*To whom correspondence should be addressed.

(ItvPS). The ItvPS intervention is executed by multiplying the probability of issuing a heat alert on day t—conditional on past information up to day t—by an odds ratio δ_t. First, we introduce a new class of causal estimands, which relies on the ItvPS intervention. We provide theoretical results to show that these causal estimands can be identified and estimated under a weaker version of the overlap assumption. Second, we propose nonparametric estimators based on the ItvPS and derive an upper bound for the variances of these estimators. Third, we extend this framework to multisite time series using a spatial meta-analysis approach. Fourth, we show that the proposed estimators perform well in terms of bias and root mean squared error via simulations. Finally, we apply our proposed approach to estimate the causal effects of increasing the probability of issuing heat alerts on each warm-season day in reducing deaths and hospitalizations among Medicare enrollees in 2837 US counties.

1. INTRODUCTION

Extreme heat events are a significant threat to public health (US Environmental Protection Agency, 2006). In the United States, heat waves have been associated with increased morbidity and mortality (Bobb *and others*, 2014; Weinberger *and others*, 2020). To reduce heat-related adverse health outcomes, the US National Weather Service (NWS) issues excessive heat warnings (to indicate more severe heat events) and heat advisories (to indicate less severe heat events) in advance of forecasted heat events to communicate these risks to the public and local government officials (Hawkins *and others*, 2017). However, how effective these excessive heat warnings and advisories (collectively, "heat alerts") are in reducing adverse health outcomes such as deaths and hospitalizations is largely unknown (Weinberger *and others*, 2021). To fill these knowledge gaps, we acquired daily time series data during the warm months (April–October) of 2006–2016 for $N = 2837$ US counties. For each county, we obtained (i) daily maximum heat index (an index that combines air temperature and relative humidity to posit a human-perceived equivalent temperature); (ii) daily issuance of heat alerts (binary); and (iii) daily number of deaths and hospitalizations among Medicare enrollees. We define the unit of analysis as *the day* to align with the time series literature in environmental epidemiology (Bell *and others*, 2004).

To analyze this data set within a causal inference framework, we need to overcome the following methodological challenges. First, the overlap assumption (Rosenbaum and Rubin, 1983), i.e., any day must have a positive probability of being treated (e.g., receiving a heat alert) or untreated (e.g., not receiving a heat alert), is often violated in time series data. This is because during the study period heat alerts were only issued on 2.52% of the warm-season days across US counties. Moreover, a heat alert is highly unlikely on a cool day, and more likely on a very hot day. Even if the overlap assumption did hold, unrealistically large sample sizes would be needed to avoid point estimates of the causal effects with very large estimated variances (Kang and Schafer, 2007). This is because the limited overlap in the covariate distribution between treated and untreated units leads to the existence of extreme propensity scores (close to 0 or 1), translating into large errors in the weights by inverse propensity weighting (IPW) methods and very low effective sample sizes for the resulting weighted populations. Second, time series observational studies are prone to time-varying confounding (Robins, 1986). To adjust for confounding bias, we must control for time-varying covariates (e.g., daily maximum heat index) associated with both the treatment (e.g., daily issuance of heat alerts) and outcomes of interest (e.g., daily deaths or hospitalizations). Third, we are dealing with multisite time series data where it is plausible that the true causal effects of interest (e.g., whether heat

alerts prevent deaths and hospitalizations in a given county) might be highly heterogeneous across counties.

Literature has focused on some aspects of the methodology gaps identified above. First, while the causal inference literature on time series studies is sparse, there are a few exceptions (Bojinov and Shephard, 2019; Sobel and Lindquist, 2014; Ning *and others*, 2019; Papadogeorgou *and others*, 2022; Shi *and others*, 2022). Bojinov and Shephard (2019) introduced an extended potential outcome framework for randomized experiments on time series data assuming that the overlap assumption always holds. They focused on the setting where the treatment at each time can be randomly assigned based on positive probabilities. Shi *and others* (2022) extended this framework to online randomized experiments with reinforcement learning (see Sutton and Barto (2018) for an overview). Sobel and Lindquist (2014) and Ning *and others* (2019) proposed a relevant causal inference framework for time series observational studies with the focus on modeling the correlation structures of multiple time series, whereas both require the overlap assumption to be held in observational data. Therefore, their methods do not directly apply to time series observational studies with overlap violations. Second, stochastic interventions (e.g., increasing or decreasing the probability of issuing heat alerts on each day), have been proposed to overcome violations of the overlap assumption in observational studies (Stock, 1989; Robins *and others*, 2004; Muñoz and Van Der Laan, 2012; Haneuse and Rotnitzky, 2013; Kennedy, 2019; Imai and Jiang, 2019; Naimi *and others*, 2021). Most recently, Kennedy (2019), Kim *and others* (2021), and Díaz *and others* (2021) proposed causal inference framework for incremental propensity scores and modified treatment policies based on stochastic interventions, accounting for time-varying confounding in longitudinal studies. However, there are important distinctions between longitudinal studies and multisite time series studies. More specifically, in longitudinal studies, the number of study units, which are often defined by individual patients or sites (N), is much larger than the number of repeated observations (T) for each study unit ($T << N$). Whereas, in multisite time series studies, the number of study units, which are instead defined by time points (e.g., days T), can be larger than the number of sites (e.g., counties N). Such distinctions in data structure lead to different considerations in defining causal estimands and developing statistical methods. Papadogeorgou *and others* (2022) are the first to bridge a stochastic intervention with time series observational data, but they focused on treatments that can be modeled by a spatio-temporal point process. In this context, they analyzed a single time series that contains information on the geographic coordinates of treatments (e.g., airstrikes). In the presence of heterogeneity of the true causal effects across counties, as is the case in our motivating example, the causal estimands defined on a homogeneous super-population—as is done in the context of longitudinal studies or for a spatio-temporal point process—might not capture the heterogeneous nature of the study population.

To the best of our knowledge, a causal inference approach in the context of a stochastic intervention for multisite time series is lacking. Accordingly, in this article, we introduce a stochastic intervention defined by the incremental time-varying propensity score (ItvPS) for time series data (i.e., the ItvPS intervention). The ItvPS intervention is executed by multiplying the probability of receiving the treatment at time t conditional on past information up to time t by an odds ratio δ_t (e.g., increase the probability of issuing heat alerts on day t). This novel framework allows us to define a broad class of stochastic causal estimands on time series data and ease the identification and estimation due to the violation of the overlap assumption.

In Section 2, we set up the notations in the context of our motivating example about national data on heat alerts. In Section 3, starting with a single time series, we define the stochastic causal estimands based on ItvPS intervention and their corresponding assumptions of identifiability. Next, in Section 4, we propose nonparametric estimators based on ItvPS, derive an upper bound for the variances of these estimators, and construct the corresponding asymptotic confidence intervals (CIs)

and time-uniform confidence sequences (CSs). In Section 5, we extend our approach into the context of multisite time series, introduce additional assumptions of identifiability for multisite time series, and propose a random-effects meta-analysis to pool causal estimands across time series from multiple sites. In Section 6, we illustrate the finite-sample performance of the proposed estimators via simulation studies. In Section 7, we apply our method to the national heat alert data set to estimate the effectiveness of heat alerts in reducing morbidity and mortality among Medicare beneficiaries. In Section 8, we conclude with a summary and discussion.

2. Set up

2.1. *Mathematical notations*

We introduce the following notation in the context of time series data. Let $Y_{i,t}, W_{i,t}, \mathbf{C}_{i,t}$ be the outcome, treatment and pretreatment covariates at time t, respectively, for $t \in \{1, 2, \ldots, T\}$ in site i, for $i \in \{1, 2, \ldots, N\}$. First, we introduce the causal estimands in the context of a single time series and omit the index i for conciseness. Later, in Section 5, we reintroduce the index i to indicate multisite time series.

For a single time series, each time t can be assigned to the treatment $W_t = 1$ or $W_t = 0$, and then the outcome Y_t is observed. We assume that W_t is binary; Y_t and $\mathbf{C}_t$ can be binary, categorical, or continuous. We define $\{\mathbf{C}_{1:T} = (\mathbf{C}_1, ..., \mathbf{C}_T), W_{1:T} = (W_1, ..., W_T),$ and $Y_{1:T} = (Y_1, ..., Y_T)\}$. We denote by $\mathcal{F}_t$ the filtration which captures past information prior to the treatment assignment at time t, that is, $\mathcal{F}_t = \{\mathbf{C}_{1:t}, W_{1:(t-1)}, Y_{1:(t-1)}\}$ (note, $\mathbf{C}_t$ included in the filtration $\mathcal{F}_t$ because it precedes the treatment that takes place at time t). Following the potential outcome framework for time series data (Bojinov and Shephard, 2019), we denote $Y_t(w_{1:t})$ the potential outcome at time t that would have been observed under the treatment path $w_{1:t} = (w_1, ..., w_t)$, and we introduce the potential outcome path $Y_{1:t}(w_{1:t}) = \{Y_1(w_{1:1}), Y_2(w_{1:2}), ..., , Y_t(w_{1:t})\}$.

We define the time-varying propensity score at time t as $p_t(w_t, \mathcal{F}_t) = Pr(W_t = w_t | \mathcal{F}_t)$, for $w_t = \{0, 1\}$. In our motivating example, the time-varying propensity score for treatment $W_t = 1$ denotes the probability of issuing a heat alert on day t conditional on past information up to day t prior to the heat alert issuance. We extend the incremental propensity score interventions proposed by Kennedy (2019) to the time series studies. The incremental propensity score intervention is an intervention that shifts the value of the propensity score. This is in contrast with the traditional deterministic intervention, which instead is implemented by shifting the value of treatment (see Section 2.2 for an example). First, we introduce the notation of the ItvPS as:

$$p_t^{\text{ItvPS}}(w_t, \mathcal{F}_t) := \frac{w_t \delta_t p_t(1, \mathcal{F}_t) + (1 - w_t) p_t(0, \mathcal{F}_t)}{\delta_t p_t(1, \mathcal{F}_t) + p_t(0, \mathcal{F}_t)}, \text{ for } w_t = \{0, 1\}. \tag{2.1}$$

Note that $p_t^{ItvPS}(w_t, \mathcal{F}_t)$ is obtained by solving the following equation

$$\delta_t = \frac{p_t^{\text{ItvPS}}(1, \mathcal{F}_t)/p_t^{\text{ItvPS}}(0, \mathcal{F}_t)}{p_t(1, \mathcal{F}_t)/p_t(0, \mathcal{F}_t)}, \tag{2.2}$$

where the odds ratio δ_t is set *a priori* by the analyst. Next, we define a stochastic intervention by the notation of ItvPS, and call it an ItvPS intervention. By executing an ItvPS intervention, we replace the probability of receiving the treatment $p_t(1, \mathcal{F}_t)$ at time t by $p_t^{\text{ItvPS}}(1, \mathcal{F}_t)$. Such a stochastic intervention reflects an odds ratio change of δ_t in the time-varying propensity scores. In the policy-relevant question, this can be explained as multiplying the probability of receiving the treatment on day t—conditional on past information up to day t, $\mathcal{F}_t$—by an odds ratio δ_t (see 2.2).

In the context of a time series with length T, we denote $\delta_{1:T} = \{\delta_1, \delta_2, ..., \delta_T\}$ as the whole intervention path and $w_{1:T}^{\text{ItvPS}}(\delta_{1:T})$ as the postintervened treatment path under the intervention path $\delta_{1:T}$.

2.2. *Motivating example*

To illustrate these mathematical notations in our motivating example, the binary treatment W_t indicates the issuance of a heat alert on day t in a given county. The covariates $\mathbf{C}_t$ include (forecasted) population-weighted daily maximum heat index that occurred just prior to treatment on day t, day of the week, and federal holidays which may be related to both heat alert issuance and adverse health outcomes (Weinberger *and others*, 2021). The outcome Y_t denotes the daily number of all-cause deaths or cause-specific hospitalizations for disease causes that were found to be associated with extreme heat in the Medicare population (Bobb *and others*, 2014). Note, we do not have information on cause-specific deaths, therefore we focus on all-cause deaths. The filtration $\mathcal{F}_t$ represents the past information prior to the treatment assignment on day t. The inclusion of past treatments $W_{1:t-1}$ and/or outcomes $Y_{1:t-1}$ is also allowed through the filtration, aligning with other literature (Bojinov and Shephard, 2019; Kennedy, 2019).

While the exact criteria used to issue heat alerts varies across the jurisdictions of local NWS offices, a key commonality across jurisdictions is that the issuance of heat alerts is based on forecasts of future weather conditions based on past information. Therefore, the probability of issuing a heat alert on day t, $p_t(w_t, \mathcal{F}_t)$ is modeled based on the past information $\mathcal{F}_t$ as discussed in the previous paragraph.

We consider a hypothetical ItvPS intervention in Los Angeles, CA, where one could increase the probability of issuing heat alerts by an odds ratio $\delta_t = 10$ for every day t during the warm months (April–October) of 2006–2016. Figure 1 illustrates the comparison between the heat alerts that were issued, and the heat alerts that could have been issued under this ItvPS intervention. For example, if the probability of issuing a heat alert on day t is 90%, under the incremental propensity score intervention with an odds ratio $\delta_t = 10$, the probability of issuing a heat alert would increase from 90% to $\frac{10 \times 90\%}{10 \times 90\% + 1 - 90\%} \approx 98.9\%$. Note that increasing the probability of issuing a heat alert by an odds ratio of 10 for each day is not equivalent to increasing the total number of heat alerts by 10 times. We find that if we hypothetically increase the probability of issuing a heat alert by an odds ratio of 10 on each warm-season days, the NWS office would have issued 128 heat alerts (an average of approximately 12 alerts per year) in the warm months of 2006–2016, compared to the total of 56 heat alerts (an average of approximately 5 alerts per year) actually issued.

3. CAUSAL ESTIMAND AND IDENTIFICATION

3.1. *General estimand*

We focus on the following scientific question "If we had changed the probability of issuing heat alerts by some prespecified amount on each day in the study period, how many total deaths would have been averted and how many cause-specific hospitalizations for heat-related diseases could have been avoided?" This question can be answered by quantifying the causal effect of an ItvPS intervention. We first define the general causal estimand of an ItvPS intervention in the context of time series data. The parameter $\tau_t(\delta_{1:t}) := E[Y_t\{w_{1:t}^{\text{ItvPS}}(\delta_{1:t})\}]$ denotes the mean potential outcome at time t with respect to the postintervened treatment path $w_{1:t}^{\text{ItvPS}}(\delta_{1:t})$. In econometrics, this parameter belongs to the class of dynamic causal effects since the parameter changes over time (Rambachan and Shephard, 2021). In our motivating example, this would be the counterfactual daily deaths or hospitalizations on day t given the ItvPS intervention.
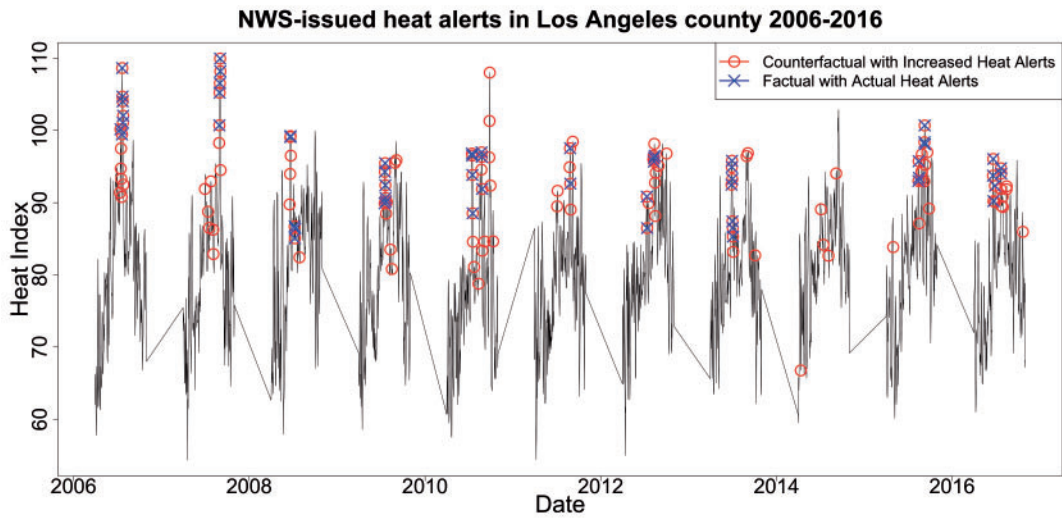
Fig. 1. NWS-issued heat alerts for Los Angeles, CA, during the warm months (April–October) of 2006–2016. The cross x's represent the heat alerts that were issued *factually*, and the circles represent the anticipated heat alerts under a counterfactual scenario where the probability of issuing heat alerts was increased for every warm-season day by an odds ratio of 10. Therefore, we find under this counterfactual scenario, Los Angeles, CA, would have issued 128 heat alerts from 2006 to 2016, compared to 56 heat alerts that were actually issued.

In cross-sectional or longitudinal studies, causal estimands are generally defined as the average of potential outcomes across N patients or sites. In the context of a single time series, we define the causal estimand as the temporal average of the potential outcomes up to time T. Under the postintervened treatment path $w_{1:t}^{\text{ItvPS}}(\delta_{1:t})$, where the probability of receiving the treatment on day t has been multiplied by an odds ratio δ_t for $t = 1, \ldots, T$, we define following temporal-average causal estimand $\bar{\tau}(\delta_{1:t,T}) = \frac{1}{T} \sum_{t=1}^{T} \tau_t(\delta_{1:t})$. In our motivating example, for Los Angeles, CA (see Figure 1), the parameter $\bar{\tau}(\delta_{1:t,T})$ denotes the average number of deaths or cause-specific hospitalizations per day under the hypothetical scenario where the probability of issuing a heat alert is multiplied by an odds ratio $\delta_t = 10$ for each day $t = 1, \ldots, T$ during the warm season.

3.2. *Causal estimand on observed treatment path*

Without further modeling assumptions, estimating the causal estimand $\bar{\tau}(\delta_{1:t,T})$ defined in Section 3.1 might be challenging for large T because this parameter depends on a treatment path with length T. Building upon Bojinov and Shephard (2019), we define a duration-t_0 causal estimand conditional on the observed treatment path $w_{1:(t-t_0-1)}^{\text{obs}}$ denoted as $\tau_t(\delta_{(t-t_0):t})$ (i.e., the duration of interventions is up to t_0 days), and its temporal average $\bar{\tau}(\delta_{(t-t_0):t,T})$, as follows

$$\tau_t(\delta_{(t-t_0):t}) = E[Y_t\{w_{1:(t-t_0-1)}^{\text{obs}}, w_{(t-t_0):t}^{\text{ItvPS}}(\delta_{(t-t_0):t})\}], \quad t_0 = 0, \ldots, t-1;$$

$$\bar{\tau}(\delta_{(t-t_0):t,T}) = \frac{1}{T-t_0} \sum_{t=t_0+1}^{T} \{\tau_t(\delta_{(t-t_0):t})\}.$$

Defining estimands dependent on prespecified fixed memory is common in time series model-ing, e.g., in an autoregressive integrated moving average model, the fixed memory up to lagged t_0

data points was prespecified. They are also compatible with the causal inference literature since they can be treated as conditional causal estimands conditioning on partial historical information $\{W_{1:(t-t_0-1)} = w^{\text{obs}}_{1:(t-t_0-1)}\}$ (Imbens and Rubin, 2015). Related theoretical developments were studied by van der Laan and Malenica (2018), studying a class of conditional (or named "context-specific") causal parameters for a single time series, conditioning on a fixed dimensional summary measure of the historical information. Furthermore, they established the statistical properties of their corresponding estimators under a targeted maximum likelihood estimation framework. Along this line, we define the causal estimands dependent on the observed treatment path $w^{\text{obs}}_{1:(t-t_0-1)}$ as a class of conditional causal estimands conditioning on historical information. We later show the estimands can be estimated using a weighted estimation strategy (see Section 4).

3.3. *Assumptions and identification*

Following the potential outcomes framework for time series data (Bojinov and Shephard, 2019), we introduce the following assumptions of identifiability:

ASSUMPTION 1 (SUTVA) The outcome path satisfies nonanticipating, consistency, and noninterference assumptions, that is, $Y^{\text{obs}}_t = Y_t(w^{\text{obs}}_{1:T}) = Y_t(w^{\text{obs}}_{1:t}) \ \forall \ t = 1, ..., T$.

We use the stable unit treatment value assumption (SUTVA) to articulate the three conditions that are required for the potential outcomes framework for time series data (VanderWeele and Hernan, 2013; Bojinov and Shephard, 2019). (i) The potential outcomes at time t can depend on the treatment path up to time t but are not allowed to depend on future treatments (nonanticipating). (ii) There is only one version of the treatment for each time t, and each treatment path up to time t realizes a unique observed outcome at time t (consistency). (iii) Since we only have one site, there is no spillover effect from other sites (noninterference).

ASSUMPTION 2 (Unconfoundedness) The assignment mechanism is unconfounded if for all $W_{1:T} \in \mathcal{W} = \{0, 1\}^T$, $W_t \perp Y_s(w_{1:s}) \mid \mathcal{F}_t \ \forall \ t = 1, ..., T$ and $t \leq s \leq T$.

Assumption 2 aligns with the "sequential randomization" assumption introduced in longitudinal studies by Robins *and others* (1994), which states that the treatment assignment only depends on past information and thus is conditionally independent of future potential outcomes. This assumption also excludes the possibility that future potential outcomes could impact the current treatment assignment retrospectively (a phenomenon that has been discussed in Granger (1980)).

ASSUMPTION 3 (Weak overlap) The assignment mechanism weakly overlaps if, for all $t \in \{1, 2, ..., T\}$, there exists a constant $\gamma > 0$, such that $p^{\text{ItvPS}}_t(w_t, \mathcal{F}_t) \leq \gamma p_t(w_t, \mathcal{F}_t), \forall w_t \in \{0, 1\}$.

The overlap assumption in traditional causal inference literature requires that each time t has a positive probability of being treated or untreated. In contrast, a stochastic intervention framework may avoid the overlap assumption by not intervening at time points with zero probability of being either treated or untreated (Kennedy, 2019). Instead, the stochastic intervention framework only requires a weak overlap assumption to hold (Papadogeorgou *and others*, 2022). Specifically, the weak overlap assumption always holds under the proposed ItvPS intervention (a special type of stochastic intervention), since, according to (2.1), $p^{\text{ItvPS}}_t(w_t, \mathcal{F}_t) \leq \max\{1, \delta_t\} p_t(w_t, \mathcal{F}_t)$, and $p^{\text{ItvPS}}_t(1, \mathcal{F}_t) \equiv 0$ when $p_t(1, \mathcal{F}_t) = 0$ and $p^{\text{ItvPS}}_t(0, \mathcal{F}_t) \equiv 0$ when $p_t(0, \mathcal{F}_t) = 0$ regardless of the value of δ_t (Naimi *and others*, 2021).

Under the assumptions defined above, our proposed causal estimand is identified: at any time point t, $\tau_t(\delta_{1:t}) = \sum_{w_{1:t} \in W_{1:t}} \int_{\partial \mathcal{R}_{1:t}} \mu(w_t, \mathcal{F}_t) \times \prod_{s=1}^{t} \Big[\underbrace{\frac{w_s \delta_s p_s(1, \mathcal{F}_s) + (1 - w_s) p_s(0, \mathcal{F}_s)}{\delta_s p_s(1, \mathcal{F}_s) + p_s(0, \mathcal{F}_s)}}_{:=p_s^{\text{ItvPS}}(w_s, \mathcal{F}_s)} \Big] \times d\Pr(\partial r_s | W_{s-1}$

$= w_{s-1}, \mathcal{F}_{s-1})$, where $\partial \mathcal{R}_{1:t} = \partial \mathcal{R}_1 \times \ldots \times \partial \mathcal{R}_t$, $\partial \mathcal{R}_s = \mathcal{F}_s / \{\mathcal{F}_{s-1}, W_{s-1}\}, s = 1, \ldots, t$, and $\mu(w_t, \mathcal{F}_t) = E(Y_t \mid W_t = w_t, \mathcal{F}_t), t = 1, \ldots, T$. The proof is provided in Section S.1 of the supplementary material available at *Biostatistics* online. However, although $\tau_t(\delta_{1:t})$ is causally identifiable, the corresponding quantity may not be stably estimated using a single time series. We instead focus on the temporal average causal estimand $\bar{\tau}(\delta_{1:t,T}) = \frac{1}{T} \sum_{t=1}^{T} \tau_t(\delta_{1:t})$, which can also be causally identified given $\tau_t(\delta_{1:t})$ is identifiable $\forall t = 1, 2, \ldots, T$. Likewise, the duration-t_0 causal estimand $\tau_t(\delta_{(t-t_0):t})$ and its temporal average $\bar{\tau}(\delta_{(t-t_0):t,T})$ can be identified under the same assumptions.

4. Estimation and inference

4.1. *Weighted estimation*

We focus on the estimation and inference of duration-t_0 causal estimands $\tau_t(\delta_{(t-t_0):t})$ given the theoretical and practical conveniences discussed in Section 3.2. If the value of time-varying propensity score $p_s(w_s, \mathcal{F}_s)$ is known, as in Section 6 of Kim *and others* (2021), the unbiased estimator of $\tau_t(\delta_{(t-t_0):t})$ is defined as,

$$\hat{\tau}_t(\delta_{(t-t_0):t}) = \prod_{s=t-t_0}^{t} \left[\underbrace{\frac{\{W_s \delta_s + (1 - W_s)\}}{\delta_s p_s(1, \mathcal{F}_s) + p_s(0, \mathcal{F}_s)}}_{:=p_s^{\text{ItvPS}}(W_s, \mathcal{F}_s) / p_s(W_s, \mathcal{F}_s)} \right] Y_t.$$

When the time-varying propensity score $p_s(w_s, \mathcal{F}_s)$ is unknown, the estimation strategy requires two steps: (i) at each time s, we estimate the time-varying propensity score $\hat{p}_s(w_s, \mathcal{F}_s)$; (ii) if the time-varying propensity scores can be modeled using correctly specified parametric models, we construct the following unbiased estimator for a duration-t_0 estimand on the observed treatment path,

$$\hat{\tau}_t(\delta_{(t-t_0):t}) = \prod_{s=t-t_0}^{t} \left[\underbrace{\frac{\{W_s \delta_s + (1 - W_s)\}}{\delta_s \hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)}}_{:=\hat{p}_s^{\text{ItvPS}}(w_s, \mathcal{F}_s) / \hat{p}_s(w_s, \mathcal{F}_s)} \right] Y_t,$$

$$\hat{\bar{\tau}}(\delta_{(t-t_0):t,T}) = \frac{1}{T - t_0} \sum_{t=t_0+1}^{T} \hat{\tau}_t(\delta_{(t-t_0):t}) = \frac{1}{T - t_0} \sum_{t=t_0+1}^{T} \Big(\prod_{s=t-t_0}^{t} \left[\frac{\{W_s \delta_s + (1 - W_s)\}}{\delta_s \hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)} \right] Y_t \Big).$$

The proposed estimator is defined as a weighted average of the observed outcomes Y_t weighted by time-varying weights, in which the weights correspond to the product of fractions where the numerators are $\hat{p}_s^{\text{ItvPS}}(w_s, \mathcal{F}_s)$ and denominators are $\hat{p}_s(w_s, \mathcal{F}_s)$. All estimators described above are generally referred to as IPW estimators, since they take estimates for the time-varying propensity score $\hat{p}_s(w_s, \mathcal{F}_s)$, and plug it into the denominator of the weights.

4.2. *Statistical inference*

We investigate the variance of the proposed IPW estimators. For a fixed t, conditioning on $\mathcal{F}_{t-t_0}$, the variance of the estimator $\hat{\tau}_t(\delta_{(t-t_0):t})$ can be defined as

$$\mathrm{Var}\{\hat{\tau}_t(\delta_{(t-t_0):t})|\mathcal{F}_{t-t_0}\} = \underbrace{E\Big(\prod_{s=t-t_0}^{t} \Big[\frac{\{W_s\delta_s + (1 - W_s)\}}{\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)} \Big]^2 Y_t^2 \Big|\mathcal{F}_{t-t_0} \Big)}_{\mathcal{V}_t}$$

$$- \Big\{E\Big(\prod_{s=t-t_0}^{t} \Big[\frac{\{W_s\delta_s + (1 - W_s)\}}{\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)} \Big] Y_t \Big|\mathcal{F}_{t-t_0} \Big)\Big\}^2,$$

where $\mathcal{V}_t$ is defined as

$$\mathcal{V}_t = \sum_{w_{(t-t_0):t} \in W_{(t-t_0):t}} \int_{\partial\mathcal{R}_{(t-t_0):t}} E(Y_t^2 \mid W_{(t-t_0):t} = w_{(t-t_0):t}, \mathcal{F}_t)$$

$$\times \prod_{s=t-t_0}^{t} \Big[\frac{\{w_s\delta_s p_s(1, \mathcal{F}_s) + (1 - w_s)p_s(0, \mathcal{F}_s)\}}{\delta_s p_s(1, \mathcal{F}_s) + p_s(0, \mathcal{F}_s)} \Big]^2 \times dPr(\partial r_s|W_{s-1} = w_{s-1}, \mathcal{F}_{s-1}).$$

The IPW estimator of $\mathcal{V}_t$ can be written as

$$\hat{\mathcal{V}}_t = \prod_{s=t-t_0}^{t} \Big[\frac{\{W_s\delta_s^2 + (1 - W_s)\}}{[\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)]^2} \Big] Y_t^2. \tag{4.3}$$

A straightforward IPW estimator of $\mathrm{Var}\{\hat{\tau}_t(\delta_{(t-t_0):t})|\mathcal{F}_{t-t_0}\}$ can be written as

$$\widehat{\mathrm{Var}}\{\hat{\tau}_t(\delta_{(t-t_0):t})|\mathcal{F}_{t-t_0}\} = \prod_{s=t-t_0}^{t} \Big[\frac{\{W_s\delta_s^2 + (1 - W_s)\}}{[\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)]^2} \Big] Y_t^2 - \Big(\prod_{s=t-t_0}^{t} \Big[\frac{\{W_s\delta_s + (1 - W_s)\}}{\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)} \Big] Y_t \Big)^2.$$

Note, this variance estimator on a single time series is always equal to 0, i.e.,

$$\prod_{s=t-t_0}^{t} \Big[\frac{\{W_s\delta_s^2 + (1 - W_s)\}}{[\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)]^2} \Big] Y_t^2 - \Big(\prod_{s=t-t_0}^{t} \Big[\frac{\{W_s\delta_s + (1 - W_s)\}}{\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)} \Big] Y_t \Big)^2$$

$$= \Big[\frac{\prod_{s=t-t_0 \& W_s=1}^{t} \delta_s^2}{\prod_{s=t-t_0}^{t}\{\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)\}^2} \Big] Y_t^2 - \Big(\Big[\frac{\prod_{s=t-t_0 \& W_s=1}^{t} \delta_s}{\prod_{s=t-t_0}^{t}\{\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)\}} \Big]^2 Y_t^2 \Big) = 0.$$

Therefore, instead, following the time series literature (Bojinov and Shephard, 2019; Papadogeorgou *and others*, 2022), we use $\mathcal{V}_t$ as an upper bound of the conditional variance of $\hat{\tau}_t(\delta_{(t-t_0):t})$ conditioning on $\mathcal{F}_{t-t_0}$. This is because we have

$$\mathrm{Var}\{\hat{\tau}_t(\delta_{(t-t_0):t})|\mathcal{F}_{t-t_0}\} = \mathcal{V}_t - \Big\{E\Big(\prod_{s=t-t_0}^{t} \Big[\frac{\{W_s\delta_s + (1 - W_s)\}}{\delta_s\hat{p}_s(1, \mathcal{F}_s) + \hat{p}_s(0, \mathcal{F}_s)} \Big] Y_t \Big|\mathcal{F}_{t-t_0} \Big)\Big\}^2 \le \mathcal{V}_t,$$

and the IPW estimator $\hat{\mathcal{V}}_t$ can be directly calculated using observed data as shown in Equation 4.3.

To calculate the variance of the temporal average estimator $\hat{\bar{\tau}}(\delta_{(t-t_0):t,T}) = \frac{1}{T-t_0}\sum_{t=t_0+1}^{T}\hat{\tau}_t(\delta_{(t-t_0):t})$, we firstly define a sequence for the estimation error $u_{t,t_0} \equiv \hat{\tau}_t(\delta_{(t-t_0):t}) - \tau_t(\delta_{(t-t_0):t}), \forall t = 1, 2, \ldots, T$ (t_0 is fixed). And then, we have $\mathrm{Var}\{\hat{\bar{\tau}}(\delta_{(t-t_0):t,T})\} = \mathrm{Var}\{\frac{1}{T-t_0}\sum_{t=t_0+1}^{T}u_{t,t_0}\}$. Analogous to the Proof of Theorem 1 in Bojinov and Shephard (2019) and Lemma 1 in Papadogeorgou *and others* (2022), u_{t,t_0} is a martingale difference sequence with respect to $\mathcal{F}_{t-t_0}$ given u_{t,t_0} is bounded and $E(u_{t,t_0}|\mathcal{F}_{t-t_0}) = 0$. Consequently, the sequence u_{t,t_0} is uncorrelated through time t, and thus

$$\mathrm{Var}\{\hat{\bar{\tau}}(\delta_{(t-t_0):t,T})\} = \frac{1}{(T-t_0)^2}\sum_{t=t_0+1}^{T}\mathrm{Var}\{u_{t,t_0}\} = \frac{1}{(T-t_0)^2}\sum_{t=t_0+1}^{T}E[\mathrm{Var}\{u_{t,t_0}|\mathcal{F}_{t-t_0}\}]$$

$$= \frac{1}{(T-t_0)^2}\sum_{t=t_0+1}^{T}E[\mathrm{Var}\{\hat{\tau}_t(\delta_{(t-t_0):t})|\mathcal{F}_{t-t_0}\}] \leq \frac{1}{(T-t_0)^2}\sum_{t=t_0+1}^{T}\mathcal{V}_t,$$

and the IPW estimator of an upper bound of $\mathrm{Var}\{\hat{\bar{\tau}}(\delta_{(t-t_0):t,T})\}$ can be expressed as $\frac{1}{(T-t_0)^2}\sum_{t=t_0+1}^{T}\hat{\mathcal{V}}_t$.

Using the martingale sequence property for the estimation error u_{t,t_0}, we can establish a central limit theorem of the proposed estimators $\hat{\bar{\tau}}(\delta_{(t-t_0):t,T})$ that allows us to build CIs.

PROPOSITION 1 (Asymptotic Normality) Suppose that Assumptions 1-3 hold, if $\frac{1}{T-t_0}\sum_{t=t_0+1}^{T}E(u_{t,t_0}^2|\mathcal{F}_{t-t_0}) \xrightarrow{p} V^*$ for a positive constant V^*, then as $T \rightarrow \infty$, for a given t_0 we have, $\sqrt{T-t_0}\{\hat{\bar{\tau}}(\delta_{(t-t_0):t,T}) - \bar{\tau}(\delta_{(t-t_0):t,T})\} = \sqrt{T-t_0}\{\frac{1}{T-t_0}\sum_{t=t_0+1}^{T}u_{t,t_0}\} \xrightarrow{d} \mathcal{N}(0, V^*)$, where $V^* = \lim_{T\to\infty}\frac{1}{(T-t_0)}\sum_{t=t_0+1}^{T}\mathrm{Var}\{u_{t,t_0}|\mathcal{F}_{t-t_0}\}$.

The proof is provided in Section S.2 of the supplementary material available at *Biostatistics* online. Here, we note that an upper bound of V^* is $\lim_{T\to\infty}\frac{1}{(T-t_0)}\sum_{t=t_0+1}^{T}\mathcal{V}_t$. Then, we construct the pointwise Wald $100(1-\alpha)\%$ CI of $\hat{\bar{\tau}}(\delta_{(t-t_0):t,T})$ as $\{\hat{\bar{\tau}}(\delta_{(t-t_0):t,T}) \pm z_{1-\alpha/2} \times \sqrt{\frac{1}{(T-t_0)^2}\sum_{t=t_0+1}^{T}\hat{\mathcal{V}}_t}\}$, where $z_{1-\alpha/2}$ denotes the upper $\alpha/2$ critical value of a standard normal distribution.

We also construct a time-uniform CS for the parameter $\bar{\tau}(\delta_{(t-t_0):t,T})$. The $100(1-\alpha)\%$ CS is a sequence of CIs (L_T, U_T) that are constructed from the first T samples and have a uniform (simultaneous) coverage guarantee (Darling and Robbins, 1967), i.e., $P\{\forall T \geq t_0 + 1 : \bar{\tau}(\delta_{(t-t_0):t,T}) \in (L_T, U_T)\} \geq 1 - \alpha$. We yield the following $100(1-\alpha)\%$ CS for $\bar{\tau}(\delta_{(t-t_0):t,T})$ based on Theorem 2 of Waudby-Smith *and others* (2021): Setting $\hat{\bar{V}}_T = \frac{1}{T-t_0}\sum_{t=t_0+1}^{T}\hat{\mathcal{V}}_t$, for any prespecified constant $\rho > 0$, the $100(1-\alpha)\%$ Lypaunov-type asymptotic CS for $\bar{\tau}(\delta_{(t-t_0):t,T})$ is $\left\{\hat{\bar{\tau}}(\delta_{(t-t_0):t,T}) \pm \right.$ $\left. \sqrt{\frac{2\{(T-t_0)\hat{\bar{V}}_T\rho^2(T)+1\}}{(T-t_0)^2\rho^2(T)}\log\left(\frac{\sqrt{(T-t_0)\hat{\bar{V}}_T\rho^2(T)+1}}{\alpha}\right)}\right\}$, with an approximate solution of time-dependent $\rho(T) := \sqrt{\frac{-\alpha^2-2\log\alpha+\log(-2\log\alpha+1-\alpha^2)}{T-t_0}}$ to optimize the boundary of CS.

5. META-ANALYSIS ON MULTISITE TIME SERIES

In many applications, time series data are available from multiple sites. In this section, we generalize our ItvPS intervention framework to combine information from multisite time series. To formalize

the causal identification in the context of random-effects across multisite time series, in Section 5.1, we introduce and discuss new assumptions of identification. In subsequent Section 5.2, we describe the estimation and inference of the pooled estimator based on meta-analysis models.

5.1. *Assumptions and identification*

There are at least two popular statistical models for meta-analysis, the fixed-effect (FE) model and the random-effects model (Borenstein *and others*, 2010). The FE model assumes that the underlying true effect is homogeneous for different sites, i.e., there is a common effect across all sites. The random-effects model allows effect heterogeneity across sites, whereas assumes the effect in each site is treated as a random sample from a super-population. In general, we found the random-effects assumption is more plausible in many applications, including our motivating example, allowing us to assume that the true causal effects of heat alerts on health outcomes vary across counties. Therefore, we propose a meta-analysis method that relies on a random-effects model to obtain a pooled estimator to summarize the overall causal effect of time series data across multiple counties (DerSimonian and Laird, 2015).

Before introducing the new pooled estimator, we introduce additional assumptions to allow for causal inference that synthesizes evidence from multisite time series. We use the index i to indicate the sites (i.e., counties in our motivating example), for $i \in \{1, 2, ..., N\}$. We denote by $\mathcal{F}_t = \{\mathcal{F}_{1,t}, ..., \mathcal{F}_{N,t}\}$ the filtration which captures the past information prior to time t in N sites.

ASSUMPTION 4 (Multisite Time Series SUTVA) The multisite outcome paths satisfy nonanticipating, consistency, and noninterference assumptions, that is, $Y_{i,t}^{\text{obs}} = Y_{i,t}(w_{1:N,1:T}^{\text{obs}}) = Y_{i,t}(w_{1:N,1:t}^{\text{obs}}) = Y_{i,t}(w_{i,1:t}^{\text{obs}}) \ \forall \ i = 1, ..., N; t = 1, ..., T.$

Assumption 4 generalizes Assumption 1 to the context of multisites. The nonanticipating and consistency assumptions are the same as in Assumption 1. The noninterference assumption, which holds trivially for a single time series, requires that the potential outcomes for one site are only affected by their own treatment path, and not by spillover effects across sites. In our motivating example, since each county often covers relatively large areas, especially in rural areas, and our study populations are older adults who are less to commute long-distance daily, we expect spillover effects across county boundaries to be small compared to the causal effects of heat alerts within each given county. However, we cannot rule out the possibility of the daily movement of people from one county to neighboring counties or that people may be exposed to heat alerts from neighboring counties, which may lead to some spillover effects.

ASSUMPTION 5 (Multisite Time Series Unconfoundedness) The assignment mechanism is unconfounded if for all $W_{i,1:T} \in \mathcal{W} = \{0,1\}^T$, and $\mathcal{F}_t$, $W_{i,t} \perp Y_{i,s}(w_{i,1:s}) \mid \mathcal{F}_t \ \forall \ i = 1, ..., N; t = 1, ..., T$ and $t \leq s \leq T$.

Assumption 5 is similar to Assumption 2, which states the treatment path for each site depends on the past information only. This does not rule out the possibility that the probability of receiving the treatment on day t in one site could depend on covariates observed in another site prior to time t. There is no need to assume that the potential outcome in site i, $Y_{i,1:t}(w_{i,1:t})$, is independent from the potential outcome in a different site j, $Y_{j,1:t}(w_{j,1:t})$, for $i \neq j$. In our motivation example, we adjusted for time-varying confounders identified by Weinberger *and others* (2021). Collaborating with environmental health experts to collect additional pre-exposure covariates that are associated with both exposures and outcomes and adjust for those covariates may further reduce the confounding bias.

ASSUMPTION 6 (Spatial Random-effects across Multisites) Each site was given the same interven-
tion path $\delta_{1:T}$, and the duration-t_0 causal estimands $\bar{\tau}_i(\delta_{(t-t_0):t})$ for site i follows $\bar{\tau}_i(\delta_{(t-t_0):t,T}) = \bar{\tau}^N(\delta_{(t-t_0):t,T}) + u_i + \epsilon_i$, where $\bar{\tau}^N(\delta_{(t-t_0):t,T})$ is the weighted-average pooled causal estimand across
N sites, $\forall t = 1, ..., T$; $t_0 = 1, ..., t$, u_i is a random effect to allow for heterogeneity and spatial corre-
lation in the causal effects across site $i = 1, ..., N$, and ϵ_i is sampling variability for site i with known
within-site variance, V_i.

Assumption 6 has been used in meta-analysis allowing both effect heterogeneity and spatial cor-
relation (DerSimonian and Laird, 2015; Maire *and others*, 2019). This assumption consists of two
components: (i) we assume the causal estimands at each site are defined by the same ItvPS interven-
tion $\delta_{1:T}$ and thus combining results in a meta-analysis is practically meaningful (e.g., one usually
chooses to combine clinical studies of the same drug, rather than completely different drugs, in one
meta-analysis); (ii) among the well-defined causal estimands $\bar{\tau}_i(\delta_{(t-t_0):t})$, we further assume effects
across multisites follow a spatial random-effects model. Such a spatial random-effects assumption
is previously stated in random-effect meta-analysis literature (Maire *and others*, 2019) and was used
in multisite environmental epidemiology studies (Bell *and others*, 2004).

We define a weighted-average causal estimand across N sites, as the estimand of interest:

$$\bar{\tau}^N(\delta_{(t-t_0):t,T}) = \sum_{i=1}^{N} c_i \bar{\tau}_i(\delta_{(t-t_0):t,T}), \text{ where } \sum_{i=1}^{N} c_i = 1 \text{ and } \{c_i, i = 1, 2, ..., N\} \text{ are fixed.}$$

Such a weighted causal estimand provides an overall summary of the causal effect drawn from a
super-population, whereas the observations from each site are treated as a random sample from this
super-population (DerSimonian and Laird, 2015). The remaining question is how to choose suitable
weights c_i for each site i. To answer this question, we propose a random-effects meta-analysis model.

5.2. *Estimation and inference*

The random-effects allow different sites to have heterogeneous causal effects that are sampled from a
distribution characterizing the overall causal effect. The primary purpose is to make inferences about
the pooled causal effect and provide a quantitative measure of how the causal effects differ across the
sites. In the meta-analysis, to obtain a weighted-average causal estimand that better characterizes the
pooled causal effect synthesized from multiple sites, we assign more weight to sites that yield a more
precise estimate of the overall causal effect. Random-effect models use an inverse variance scheme
to assign weights to each site based on inverse proportions of the total variance from each site.
Specifically, there are two sources of variance under a random-effects model (Borenstein *and others*,
2010). First, the observed causal effect $\hat{\bar{\tau}}_i(\delta_{(t-t_0):t,T})$ for any time series in one site differs from that time
series' true causal effect because of within-site variance, V_i. Second, the true causal effect for each
time series differs from the overall causal effect because of between-site variance, Δ^2. DerSimonian
and Laird (1986) are among the first to propose the following method to estimate the between-site
variance, Δ^2:

(1) Obtain a common effect estimator under a FE meta-analysis model, $\hat{\bar{\tau}}_{\text{FE}} = \frac{\sum_{i=1}^{N} \hat{\bar{\tau}}_i(\delta_{(t-t_0):t,T})/V_i}{\sum_{i=1}^{N} 1/V_i}$.

(2) Based on $\hat{\bar{\tau}}_{\text{FE}}$, calculate the Cochran's Q-statistic, $Q = \sum_{i=1}^{N} \frac{[\hat{\bar{\tau}}_i(\delta_{(t-t_0):t,T}) - \hat{\bar{\tau}}_{\text{FE}}]^2}{V_i}$.

(3) Obtain the estimator for the between-site variance Δ^2 as $\hat{\Delta}_{DL}^2 = \frac{Q - (N-1)}{\sum_{i=1}^{N} 1/V_i - \frac{\sum_{i=1}^{N} 1/V_i^2}{\sum_{i=1}^{N} 1/V_i}}$.

The weight assigned to each site under the inverse variance scheme is $c_i = \frac{1}{V_i + \Delta^2}$, $i = 1, 2, ..., N$, where the within-site variance, V_i, is unique to each site, whereas the between-site variance Δ^2 is a quantity that is common across sites. Various methods to estimate Δ^2 are summarized in Borenstein *and others* (2010). To allow spatial correlation in the causal effects across sites, we additionally use a spatial random-effects meta-analysis model proposed by Maire *and others* (2019). For this model, we first identify the spatial coordinates (longitude and latitude) for the centroids of each county. We then specify a Gaussian spatial correlation structure based on the relative Euclidean distances computed by the longitude and latitude, d. Finally, we fit the following spatial random-effects meta-analysis model using **metafor** R package (Viechtbauer, 2010): $\bar{\tau}_i(\delta_{(t-t_0):t,T}) = \bar{\tau}^N(\delta_{(t-t_0):t,T}) + u_i + \epsilon_i$, where $\bar{\tau}_i(\delta_{(t-t_0):t})$ is the duration-t_0 causal effects for site i, $\bar{\tau}^N(\delta_{(t-t_0):t,T})$ is the weighted-average pooled causal effect across N sites. The random effects $u_1, ..., u_N$ were assumed to follow a multivariate normal distribution with a mean of zero and a variance–covariance matrix $Cov(u_i, u_j) = \Delta^2 \times \exp(-d_{i,j}^2/\rho^2)$, where Δ^2 denotes the between-site variance, $d_{i,j}$ denotes the Euclidean distance between two spatial points i and j. ρ is the spatial correlation parameter for the Gaussian correlation structure, which can be estimated by restricted maximum likelihood.

6. SIMULATION STUDY

We study finite-sample properties of the proposed estimators via simulation studies on a single time series, in which we vary: (i) the length of the time series T; (ii) the duration of interventions t_0; and (iii) the assignment mechanisms of the treatment $p_t(W_t = 1 \mid \mathcal{F}_t)$. We compare the IPW estimator to the nonparametric influence-function-based estimator initially proposed by Kennedy (2019), modified to the time series setting. To reflect the nature of treatment assignment in our motivating example, we generate the treatment W_t under a nearly nonoverlapping setting. In particular, we consider a time series with length T. For each $t = 1, ..., T$, $\mathbf{C}_t = (C_{1,t}, C_{2,t}, C_{3,t}, C_{4,t}, C_{5,t}) \sim N(\mathbf{0}, \mathbf{I}_5)$; $p_t(W_t = 1 \mid \mathcal{F}_t) = \text{expit}\{10 \times (\sum_{j=1}^5 C_{j,t}/5 - W_{t-1} + 0.5)\}$; $Y_t \mid W_t, W_{t-1}, \mathbf{C}_t \sim N(3 \times W_t + W_{t-1} + \sum_{j=1}^5 C_{j,t}/5, 1)$. At time t, the random assignment mechanism of the treatment W_t depends on both $\mathbf{C}_t$, and the treatment at time $t-1$, W_{t-1}. Also the outcome Y_t depends on the treatments both at time t and $t-1$ (duration-1 effect). The data-generating mechanism described above mimics a nearly nonoverlap setting (see Figure S.1 of the supplementary material available at *Biostatistics* online, showing the distributions of the time-varying propensity scores have little overlap across treated vs. untreated units). The main quantity of interest is the duration-t_0 causal estimand on the observed treatment path. We assess the performance of the selected estimators by calculating the integrated bias and root mean squared error (RMSE) defined as; $\widehat{\text{Integrated Bias}} = \frac{1}{J} \sum_{j=1}^J \left| \frac{1}{K} \sum_{k=1}^K \left\{ \hat{\bar{\tau}}^k(\delta_{t-t_0:t,T,j}) - \bar{\tau}^k(\delta_{t-t_0:t,T,j}) \right\} \right|$, $\widehat{\text{RMSE}} = \frac{\sqrt{N}}{J} \sum_{j=1}^J \left[\frac{1}{K} \sum_{k=1}^K \left\{ \hat{\bar{\tau}}^k(\delta_{t-t_0:t,T,j}) - \bar{\tau}^k(\delta_{t-t_0:t,T,j}) \right\}^2 \right]^{1/2}$, where $\bar{\tau}^k(\delta_{t-t_0:t,T,j}) := \frac{1}{T-t_0} \sum_{t=t_0+1}^T \bar{\tau}^k(\delta_{t-t_0:t,j})$ is the true temporal average causal quantity, and $\hat{\bar{\tau}}^k(\delta_j)$ is its estimator based on the estimated time-varying propensity score, in which the superscript k indicates the simulation replicate. We assess the estimation performances at $J = 50$ values of δ_j equally spaced between 0.1 and 10 evaluated on $K = 500$ simulation replicates. We also assess the average coverage of our proposed point-wise Wald 95% CIs, and the coverage of time-uniform 95% CSs.

We vary the following combinations of (T, t_0), $T = (200, 1000, 5000)$, and $t_0 = (1, 4, 9)$. We apply both parametric and nonparametric models to estimate the time-varying propensity scores, following the recommendation by Bonvini *and others* (2021), (i) logistic regression; (ii) Super Learner (Van der Laan *and others*, 2007), which combines generalized additive models, multivariate adaptive regression splines, support vector machines, and random forests, along with parametric generalized linear models (with and without interactions, and with terms selected stepwise via Akaike

Table 1. *Characteristics for NWS-issued heat alerts, all-cause deaths among Medicare enrollees, and cause-specific hospitalizations for five heat-related diseases among Medicare fee-for-service (FFS) enrollees across April–October of 2006–2016*

Variables	2837 counties	550 populous counties[†]
% Days with heat alerts	2.52	2.22
No. of deaths	10 467 201	7 653 987
No. of heat stroke	97 399	72 649
No. of urinary tract infections	1 424 046	1 061 060
No. of septicemia	2 614 871	1 954 011
No. of renal failure	1 207 903	894 164
No. of fluid and electrolyte disorders	928 270	673 007

Counties with population size $> 100\,000$.

information criterion (AIC)), consistent with Kennedy (2019) (implemented by the **SuperLearner** R package). We assume that the form of the time-varying propensity score model is correctly specified as $W_t \mid \mathbf{C}_t, W_{t-1}$ with a **logit** link. We omit the RMSE and coverage rate results for the IPW estimators that use Super Learner for propensity score models since these estimators fail to achieve asymptotic linearity and thus are generally incompatible with theoretical claims requiring asymptotic normality, unlike those using a correctly specified parametric logistic regression for propensity score models. We conduct additional simulations in Section S.3 of the supplementary material available at *Biostatistics* online when the form of the time-varying propensity score model is misspecified.

Table 2 shows the integrated bias and RMSE of the IPW and influence-function-based estimator, along with the average coverage rate of their corresponding Wald 95% CIs. Under the same duration t_0, we observed that the integrated bias and RMSE of the estimator generally decrease when the length of the time series, T, increases, regardless of how the time-varying propensity score is estimated, either by logistic regression or Super Learner. We also observed a decreased performance of the estimator as the duration, t_0, increases. This finding is not surprising given that the estimated weights in the proposed weighting estimator depend on the product of t_0+1 terms of estimated ItvPS. We expect that the proposed estimator will likely be unstable, and thus the integrated bias and RMSE will be larger when t_0 increases. In practice, the exact duration t_0 is unknown, and researchers need to specify t_0 based on their prior domain knowledge. The simulation results suggest that the choice of t_0 should be parsimonious, i.e., one should choose a duration t_0 that can capture the data complexity yet is as small as possible. We found, in general, the IPW estimator and influence-function-based estimator perform comparably in the simulation settings considered. Although, under these simulation settings, the IPW estimators that pair with Super Learner perform well in terms of integrated bias, possibly due to reduced model misspecification bias, such finite-sample results are not generally held, especially considering the potential large asymptotic bias arising from the slower converge rate of Super Learner. Additional simulations in Section S.3 of the supplementary material available at *Biostatistics* online present similar findings when the underlying data-generating mechanism is unknown and potentially misspecified.

We also found that the coverage rates of the Wald 95% CIs for the IPW estimator were near or above the nominal level (95%) when T is relatively large, as shown in Table 2. The conservative performance is likely due to the fact that we used an upper bound for the variance estimates when constructing the CIs. The coverage rates of the Wald 95% CIs for the influence-function-based estimator were more frequently below the nominal level. The coverage rates decrease when t_0 increases, which is consistent with the coverage results shown in Papadogeorgou *and others* (2022). Figure 2

Table 2. *Simulation results for the scenario assuming the treatment assignment mechanism is specified with a **logit** link. The integrated bias and RMSE (multiplied by 10 for easier interpretation) of proposed estimators, the average coverage of proposed confidence intervals (CIs), and the uniform coverage of proposed confidence sequences (CSs) for IPW estimators*

| Propensity score model | | | IPW | | | | Influence-function-based | | |
	T	t_0	Bias	RMSE	Coverage (%)	Uniform coverage (%)	Bias	RMSE	Coverage (%)
		1	1.34	1.57	99.97	98.20	1.31	1.57	99.84
	200	4	2.95	3.15	97.80	98.20	2.80	3.03	96.56
		9	5.57	5.73	89.58	16.40	5.18	5.38	84.22
		1	0.25	0.45	100.00	100.00	0.30	0.48	98.06
Logistic regression	1000	4	0.59	0.74	99.98	100.00	0.78	0.90	79.30
		9	1.06	1.18	96.45	100.00	1.60	1.68	65.46
		1	0.12	0.20	100.00	100.00	0.16	0.23	99.42
	5000	4	0.30	0.35	100.00	100.00	0.40	0.43	91.84
		9	0.60	0.64	97.68	100.00	0.78	0.81	70.41
		1	1.08				1.26	1.50	99.16
	200	4	2.24				2.39	2.62	95.24
		9	4.11				4.23	4.43	82.65
		1	0.04				0.08	0.35	99.97
Super Learner	1000	4	0.08				0.25	0.46	99.73
		9	0.21				0.40	0.59	98.70
		1	0.04				0.04	0.16	100.00
	5000	4	0.14				0.15	0.24	99.89
		9	0.33				0.34	0.42	98.14

visualizes the curve of duration-1 causal effects when the probability of the treatment assignment is multiplied by odds ratios $\delta_t \in [0.1, 10]$. The solid red line represents the estimated causal effects along with point-wise Wald 95% CIs (dashed line). We found that the bias between the estimated curve and the true curve reduces when T increases. The point-wise Wald 95% CIs capture the true curve in all three scenarios; $T = 200, 1000, 5000$. The time-uniform CSs generally perform unstably when the sample sizes are small ($T = 200$) yet perform conservatively when the sample sizes are relatively large ($T = 1000, 5000$).

We design our simulation studies to reflect scenarios of nearly nonoverlap between treated and untreated units observed in our data application. Our simulation results show that the proposed estimators perform well in terms of bias and RMSE even when the overlap assumption is nearly violated.

7. APPLICATION

We analyze the multisite time series data described in Section 2.2. First, we apply the proposed methods to estimate the causal estimands $\bar{\tau}_i(\delta_{(t-t_0):t,T})$. We assume δ_t is the same for every day t during the warm months (April–October) of 2006–2016. In the time-varying propensity score model, we included the following observed covariates: daily maximum heat index, lag-1 daily maximum heat index, lag-2 daily maximum heat index, moving average heat index during the current warm season, lag-1 day heat alert, lag-2 day heat alert, the running total number (the summation of the sequence of numbers updated daily) of heat alerts that have been issued during the current warm season, the moving average number of deaths/hospitalizations during the current warm season, day of the week, and federal holidays. Note, as described in Section 2.2, the inclusion of past treatments
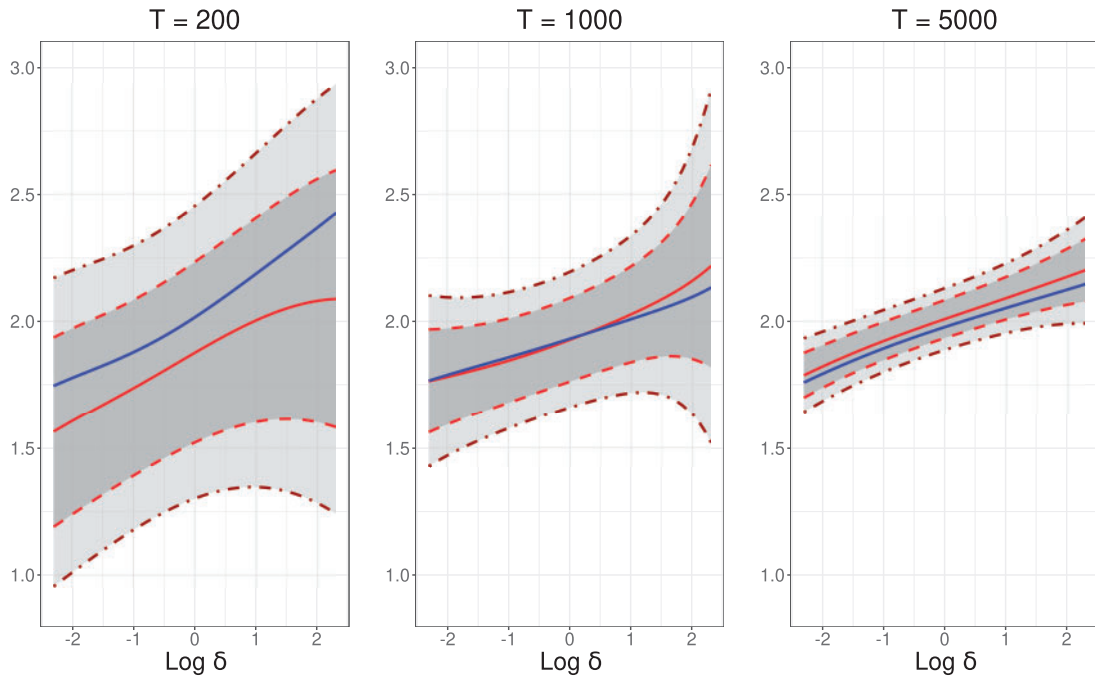
Fig. 2. The duration-1 causal effect curves obtained by the IPW estimation when the probability of treatment assignments was multiplied by an odds ratio $\delta_t \in [0.1, 10]$. The solid red line represents the estimated causal effect curve along with the point-wise Wald 95% CIs (red dashed line) and the time-uniform 95% CSs (dot-dash line). The solid blue line represents the true causal effect curve. The left, middle, and right panels reflect the simulation scenarios with $T = 200, 1000, 5000$. The propensity scores were estimated by a logistic regression model.

(lagged heat alerts) and past outcomes (historical deaths/hospitalizations) is allowed in the time-varying propensity score model. Considering the relationships between time-varying treatment and covariates are potentially complex, the model misspecification bias is of greater concern in this data application; we estimate the time-varying propensity score by the Super Learner using the same combination of algorithms described in Section 6.

For each county i, we estimate the daily numbers of all-cause deaths and cause-specific hospitalizations for five heat-related diseases (heat stroke, urinary tract infections, septicemia, renal failure, fluid, and electrolyte disorders) under several ItvPS intervention scenarios ranging from $\delta_t = 1$ to $\delta_t = 10$. We consider the causal estimands with duration $t_0 = 2$, consistent with Weinberger *and others* (2018). We define the county-specific causal effect curve for county i as $\bar{\tau}_i(\delta_{(t-2):t,T}) - \bar{\tau}_i(1, 1, 1)$, where $\bar{\tau}_i(\delta_{(t-2):t,T})$ denotes the estimated daily average number of deaths or hospitalizations under various ItvPS interventions ($\delta_t \in [1, 10]$) and $\bar{\tau}_i(1, 1, 1)$ denotes the quantity corresponding to the daily average number of deaths or hospitalizations observed *factually* (as if there was no change in the treatment assignments). After obtaining all $N = 2837$ county-specific causal effect curves, we utilize the spatial random-effects meta-analysis approach proposed in Section 5 to pool the estimated county-specific causal effect curves across multiple counties and obtain the estimated pooled causal effect curve, $\sum_{i=1}^{N} c_i[\bar{\tau}_i(\delta_{(t-2):t,T}) - \bar{\tau}_i(1, 1, 1)]$, where the weights $c_i, i = 1, 2, \ldots, N$ are obtained by the spatial random-effects meta-analysis model.

Figure 3 shows the estimated pooled causal effects for the daily average number of all-cause deaths and cause-specific hospitalizations for five heat-related diseases per county among 2837 counties
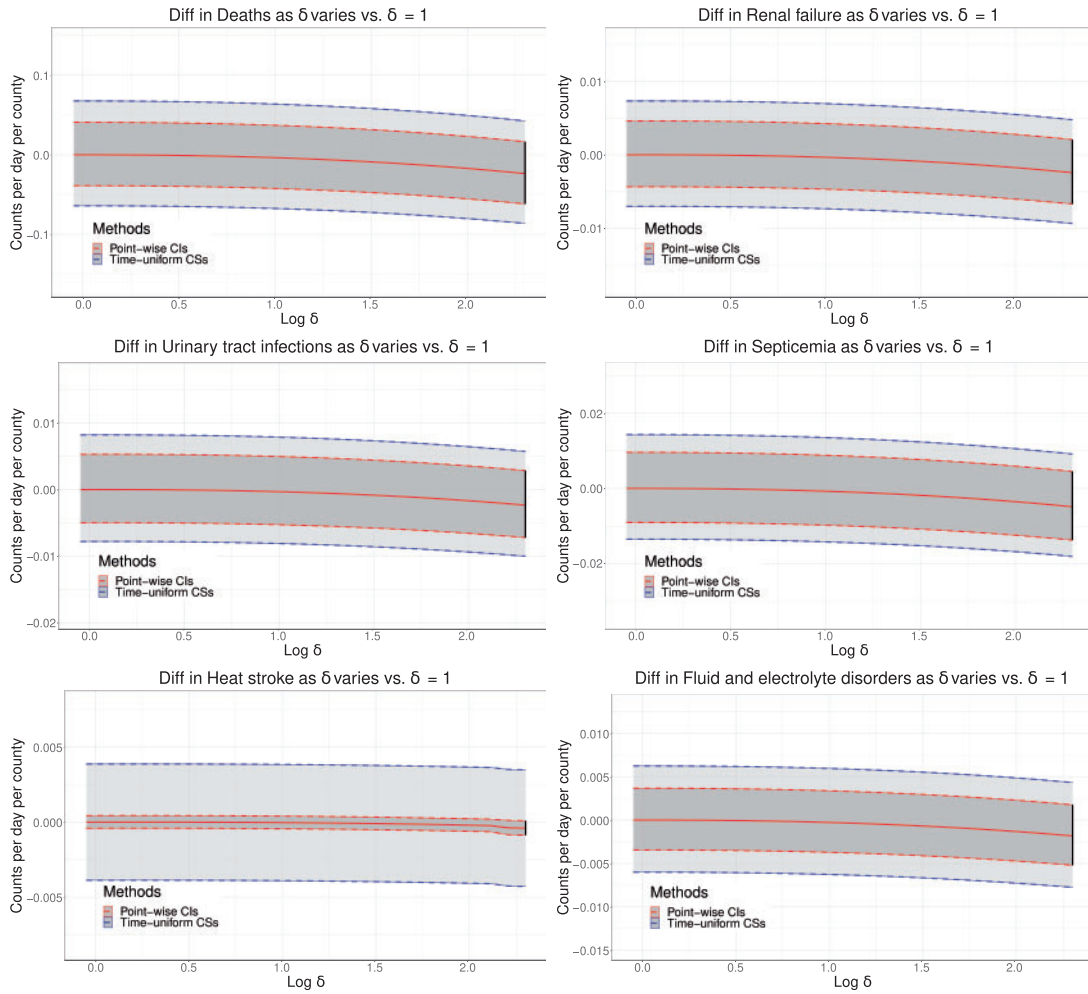
Fig. 3. The estimated pooled causal effect curves of average all-cause deaths and cause-specific hospitalizations for five heat-related diseases per day per county among 2837 counties, assuming the duration of interventions was 2 days. The curves represent the differences in deaths and hospitalizations comparing the counterfactual situations where the probability of issuing heat alerts was multiplied by an odds ratio $\delta_t \in [1, 10]$ to the factual situations where the probability of issuing heat alerts remains unchanged ($\delta_t = 1$). The dashed red lines represent the corresponding point-wise Wald 95% CIs of the differences. The dashed blue lines represent the time-uniform 95% CSs.

across the warm months of 2006–2016. The curves represent the differences in daily average numbers of deaths and hospitalizations averaged across 2837 counties comparing the counterfactual scenarios where the probability of issuing heat alerts was multiplied by an odds ratio $\delta_t \in [1, 10]$ to the factual scenario, where the probability of issuing heat alerts remains unchanged ($\delta_t = 1$). The dashed red lines represent the corresponding point-wise Wald 95% CIs of the differences. The dashed blue lines represent the time-uniform 95% CSs. The vertical lines represent the average number of deaths and hospitalizations that could be avoided per day per county and their corresponding CIs if we had increased the probability of issuing heat alerts by the maximum odds considered, $\delta_t = 10$. We found,

in general, consistent downward patterns among the pooled causal effect curves for each health outcome as the $\log(\delta_t)$ increases above 0, indicating slight reductions in average all-cause deaths and cause-specific hospitalizations for five heat-related diseases among Medicare enrollees as $\log(\delta_t)$ increases. The CIs contain 0 throughout the range of $\delta_t \in [1, 10]$.

Quantitatively, we found that if we had increased the probability of issuing heat alerts by an odds ratio $\delta_t = 10$, on average 0.15 (95% CI: -0.01 to 0.32) deaths could be averted per day per county (see vertical line in Figure 3a). Based on these estimates, solely among extremely hot days (i.e., the top 5% hottest days when the heat alerts are more likely to be issued), we estimated 4653 avoidable deaths (95% CI: -415 to 9576) across all 2837 counties in one warm season. Similarly, if we increase the probability of issuing heat alerts by an odds ratio $\delta_t = 10$, the numbers of hospitalizations averted are 42 (95% CI: -7 to 91) for heat stroke; 634 (95% CI: -81 to 1327) for urinary tract infections; 1157 (95% CI: -134 to 2409) for septicemia; 585 (95% CI: -18 to 1169) for renal failure; 405 (95% CI: -43 to 840) for fluid and electrolyte disorders among extremely hot days in one warm season. However, given the wide CIs for the results of all outcomes, we did not find any statistically significant causal effects of increasing the probability of issuing heat alerts on health outcomes comparing $\delta_t = 10$ (increased odds) vs. $\delta_t = 1$ (unchanged). We also found there is large between-county heterogeneity in the random-effects meta-analysis model, (P-value of heterogeneity test < 0.001), for all outcomes. Overall, our findings suggest weak evidence that increasing the probability of issuing heat alerts may bring health benefits to the US Medicare population.

In Section S.4 of the supplementary material available at *Biostatistics* online, we conducted additional county-specific analyses that include time series data from three selected counties: Santa Clara, CA; Maricopa, AZ; and New York, NY, separately. Notably, we found that in Maricopa, AZ, increasing the probability of issuing heat alerts could significantly reduce all-cause deaths, whereas less dramatic but still statistically significant results were found in Santa Clara, CA and New York, NY (see Figure S.2 of the supplementary material available at *Biostatistics* online). As a sensitivity analysis, we analyzed a subset of 550 counties with population sizes $> 100\,000$ ("populous counties"). These counties included more than 70% of all-cause deaths and cause-specific hospitalizations observed among the 2837 counties. As shown in Figure S.3 of the supplementary material available at *Biostatistics* online, the shapes of the causal effect curves based on the subset of 550 populous counties are very similar to the estimated curves based on data from all 2837 counties. We found nonsignificant results for all outcomes. The random-effects meta-analysis model still indicates considerable between-county heterogeneity (P-value of heterogeneity test < 0.001) for all outcomes. Additional analysis was conducted in which the duration of interventions was varied. We also estimated pooled causal effect curves assuming the duration of interventions was 0 and 5 days, respectively. As shown in Figures S.4 and S.5 of the supplementary material available at *Biostatistics* online, the causal effect of increasing the frequency of issuing heat alerts may have more pronounced effects on health outcomes when considering a longer duration of interventions. Importantly, we observed statistically significant results of the causal effects of increasing the probability of issuing heat alerts on health outcomes comparing $\delta_t = 10$ (increased odds) vs. $\delta_t = 1$ (unchanged) when considering cumulative effects up to duration $t_0 = 5$, i.e., on average 0.15 (95% CI 0.06–0.24) avoidable deaths per day per county. We also found the widths of CIs for the causal effect curves of duration-0, duration-2, and duration-5 estimands did not change notably as t_0 varies.

8. DISCUSSION

We have developed a novel causal inference framework for multisite time series data, which relies on stochastic interventions. Under this framework, we introduced a class of causal estimands and their unbiased estimators with theoretical justification. In the context of multisite time series data,

we link our ItvPS intervention framework to the spatial random-effects meta-analysis. The ultimate goal is to obtain a pooled causal estimator that summarizes the potentially heterogeneous causal effects from multisite time series studies.

Our approach was motivated by our data application, where the goal was to assess whether or not increasing the probability of issuing heat alerts reduces morbidity and mortality. More specifically, we estimated the causal effects of increasing the probability of issuing heat alerts on all-cause deaths and cause-specific hospitalizations for five heat-related diseases. We found some evidence of reductions in morbidity and mortality with large statistical uncertainty. Importantly, the random-effects meta-analysis model indicated large heterogeneity of the causal effects across different counties.

One key distinction of our approach compared to more traditional statistical methods in environmental epidemiology is that most time series studies in this context use matched case-crossover or difference-in-difference designs, where the focus is often the (average) causal contrast between treated days (i.e., days with heat alerts) vs. untreated days (i.e., days without heat alerts), and the number of heat alert days is then fixed for a given study period in one chosen site (Weinberger *and others*, 2018, 2021). In contrast, our analysis focuses on the causal effect of a stochastic intervention, estimating the causal effect as the probability of issuing heat alert changes, and as such, the counterfactual heat alert days are not fixed. Therefore, our results are not directly comparable to previously published studies (Weinberger *and others*, 2021).

While our proposed novel causal inference approach provides a new powerful tool in policy evaluations, some methodology considerations are needed when applying this approach to future studies. First, we expect causal effects defined by stochastic interventions to answer a different policy question than that of a deterministic intervention (Kennedy, 2019). The proposed causal estimand is an intuitive quantity answering the following causal question "how many adverse health outcomes could have been averted if we changed the frequency of heat alerts?" which was the focus of our work. However, other practitioners' interests may be more aligned with estimating causal effects of deterministic interventions, for instance, "how many adverse health outcomes could have been avoided if the temperature was, on average, one degree lower during extremely hot days (Weinberger *and others*, 2020)?" In that case, causal inference methods based on deterministic interventions (Bojinov and Shephard, 2019; Bojinov *and others*, 2021; Rambachan and Shephard, 2021) may be used. Second, the proposed weighting estimator maintains ideal asymptotic properties only when the time-varying propensity scores can be modeled using correctly specific parametric models. Using adaptive nonparametric models for propensity score estimation may reduce the model misspecification bias in weights; however, such IPW estimators do not attain asymptotic normality and may suffer from large asymptotic bias. Along this line, an influence-function-based estimator may be compatible with the nonparametric propensity score estimation (Kennedy, 2022); yet its asymptotic properties under the time series setting require further investigation. As part of future work, we plan to conduct a theoretical analysis of doubly robust targeted estimation of an ItvPS intervention. Third, the random-effects meta-analysis creates a weighted-average causal estimand and its corresponding pooled estimator, in which the weights are calculated by combining between-site and within-site variances. For this reason, the weighted-average causal quantity is defined on a hypothetical weighted population. Dahabreh *and others* (2020) have criticized that standard meta-analyses may produce results that do not belong to a clear target population when each site represents a different population, and the treatment effect varies across these populations. As part of future work, we plan to develop approaches that allow inferences to be transported from multisite time series to a clearly specified target population (Dahabreh *and others*, 2020). Fourth, the noninterference assumption may not hold in many climate and health studies, such as this one, which rely on spatial–temporal data. For instance, the NWS-issued heat alerts in one county may impact

people in adjacent counties. As part of future work, we hope to extend the time series intervention path to a multivariate intervention path defined by random matrices (Papadogeorgou *and others*, 2022), to potentially overcome the violation of noninterference assumption, and identify direct and spillover effects under this stochastic intervention framework. Fifth, the stable estimation of time-varying propensity score based on time series observational data is challenging since the time-varying confounder sets are potentially high-dimensional. While in our data application we used the state-of-the-art Super Learner (an ensemble of flexible parametric/nonparametric models) to estimate time-varying propensity scores, we had to limit the size of confounder sets to avoid unstable estimates due to the curse of dimensionality. Also, the proposed weighting estimator that relies on products of the estimated ItvPS may be unstable when accounting for the longer duration of interventions. We plan to generalize covariate balance methods to improve high-dimensional propensity score estimation and stability of the weighting estimators in time series observational studies (Athey *and others*, 2018).

The stochastic intervention framework for time series data introduced in this article is the first approach that allows the identification and estimation of causal quantities defined by stochastic interventions on multisite time series. We believe that this framework addresses one of the emerging methodological needs in climate and health research, where researchers often collect time series data from multiple geographic locations seeking causal evidence among diverse populations (Liu *and others*, 2019; Lee *and others*, 2020). Furthermore, we expect that the proposed framework can be applied to science and policy-relevant research in political science, economics, and law, where a considerable amount of spatial–temporal data are generated and collected.

9. Software

Software in the form of R code to implement both simulation studies and data applications is available at https://github.com/wxwx1993/TS_Stochastic.

Supplementary material

Supplementary material is available online at http://biostatistics.oxfordjournals.org.

Acknowledgments

Funding

References

Athey, S., Imbens, G. W. and Wager, S. (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 597–623.

Bell, M. L., Samet, J. M. and Dominici, F. (2004). Time-series studies of particulate matter. *Annual Review of Public Health* **25**, 247–280.

Bobb, J. F., Obermeyer, Z., Wang, Y. and Dominici, F. (2014). Cause-specific risk of hospital admission related to extreme heat in older adults. *JAMA* **312**, 2659–2667.

Bojinov, I., Rambachan, A. and Shephard, N. (2021). Panel experiments and dynamic causal effects: a finite population perspective. *Quantitative Economics* **12**, 1171–1196.

Bojinov, I. and Shephard, N. (2019). Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association* **114**, 1665–1682.

Bonvini, M., McClean, A., Branson, Z. and Kennedy, E. H. (2021). Incremental causal effects: an introduction and review. arXiv, arXiv:2110.10532, preprint: not peer reviewed. Avaliable at: https://doi.org/10.48550/arXiv.2110.10532.

Borenstein, M., Hedges, L. V., Higgins, J. P. T. and Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods* **1**, 97–111.

Dahabreh, I. J., Petito, L. C., Robertson, S. E., Hernán, M. A., and Steingrimsson, J. A. (2020). Toward causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population. *Epidemiology* **31**, 334–344.

Darling, D. A. and Robbins, H. (1967). Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences of the United States of America* **58**, 66–68.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.

DerSimonian, R. and Laird, N. (2015). Meta-analysis in clinical trials revisited. *Contemporary Clinical Trials* **45**, 139–145.

Díaz, I., Williams, N., Hoffman, K. L. and Schenck, E. J. (2021). Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association*, 1–16.

Granger, C. W. J. (1980). Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control* **2**, 329–352.

Haneuse, S. and Rotnitzky, A. (2013). Estimation of the effect of interventions that modify the received treatment. *Statistics in Medicine* **32**, 5260–5277.

Hawkins, M. D., Brown, V. and Ferrell, J. (2017). Assessment of NOAA national weather service methods to warn for extreme heat events. *Weather, Climate, and Society* **9**, 5–13.

Imai, K. and Jiang, Z. (2019). Comment: The challenges of multiple causes. *Journal of the American Statistical Association* **114**, 1605–1610.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge, England, UK: Cambridge University Press.

Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22**, 523–539.

Kennedy, E. H. (2019). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association* **114**, 645–656.

Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review. arXiv, arXiv:2203.06469, preprint: not peer reviewed. Avaliable at: https://doi.org/10.48550/arXiv.2203.06469.

KIM, K., KENNEDY, E. H. AND NAIMI, A. I. (2021). Incremental intervention effects in studies with dropout and many timepoints. *Journal of Causal Inference* **9**, 302–344.

LEE, W., KIM, Y., SERA, F., GASPARRINI, A., PARK, R., CHOI, H. M., PRIFTI, K., BELL, M. L., ABRUTZKY, R., GUO, Y. *and others* (2020). Projections of excess mortality related to diurnal temperature range under climate change scenarios: a multi-country modelling study. *The Lancet Planetary Health* **4**, e512–e521.

LIU, C., CHEN, R., SERA, F., VICEDO-CABRERA, A. M., GUO, Y., TONG, S., COELHO, M. S.Z.S., SALDIVA, P. H.N., LAVIGNE, E., MATUS, P. *and others* (2019). Ambient particulate air pollution and daily mortality in 652 cities. *New England Journal of Medicine* **381**, 705–715.

MAIRE, A., THIERRY, E., VIECHTBAUER, W. AND DAUFRESNE, M. (2019). Poleward shift in large-river fish communities detected with a novel meta-analysis framework. *Freshwater Biology* **64**, 1143–1156.

MUÑOZ, I. D. AND VAN DER LAAN, M. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics* **68**, 541–549.

NAIMI, A. I., RUDOLPH, J. E., KENNEDY, E. H., CARTUS, A., KIRKPATRICK, S. I., HAAS, D. M., SIMHAN, H. AND BODNAR, L. M. (2021). Incremental propensity score effects for time-fixed exposures. *Epidemiology* **32**, 202–208.

NING, B., GHOSAL, S. AND THOMAS, J. (2019). Bayesian method for causal inference in spatially-correlated multivariate time series. *Bayesian Analysis* **14**, 1–28.

PAPADOGEORGOU, G., IMAI, K., LYALL, J. AND LI, F. (2022). Causal inference with spatio-temporal data: estimating the effects of airstrikes on insurgent violence in Iraq. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **84**, 1969–1999.

RAMBACHAN, A. AND SHEPHARD, N. (2021). *When Do Common Time Series Estimands have Nonparametric Causal Meaning?* [Online; cited October 2021] avaliable at: https://scholar.harvard.edu/shephard/publications/nonparametric-dynamic-causal-model-macroeconometrics. preprint: not peer reviewed.

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling* **7**, 1393–1512.

ROBINS, J. M., HERNÁN, M. A. AND SIEBERT, U. (2004). Effects of multiple interventions. In: Majid, E. *and others* (editors), *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, Geneva, Switzerland: World Health Organization, Volume 1. pp. 2191–2230.

ROBINS, J. M., ROTNITZKY, A. AND ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

ROSENBAUM, P. R. AND RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

SHI, C., WANG, X., LUO, S., ZHU, H., YE, J. AND SONG, R. (2022). Dynamic causal effects evaluation in a/b testing with a reinforcement learning framework. *Journal of the American Statistical Association*, 1–13.

SOBEL, M. E. AND LINDQUIST, M. A. (2014). Causal inference for fMRI time series data with systematic errors of measurement in a balanced on/off study of social evaluative threat. *Journal of the American Statistical Association* **109**, 967–976.

STOCK, J. H. (1989). Nonparametric policy analysis. *Journal of the American Statistical Association* **84**, 567–575.

SUTTON, R. S. AND BARTO, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, Massachusetts, USA: MIT Press.

US ENVIRONMENTAL PROTECTION AGENCY. (2006). *Excessive Heat Events Guidebook*. Washington, DC, USA: National Service Center for Environmental Publications (NSCEP).

van der Laan, M. J. and Malenica, I. (2018). Robust estimation of data-dependent causal effects based on observing a single time-series. arXiv, arXiv:1809.00734, preprint: not peer reviewed. Avaliable at: https://doi.org/10.48550/arXiv.1809.00734.

Van der Laan, M. J., Polley, E. C. and Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology* **6**, Article, Article 25, 1–21.

VanderWeele, T. J. and Hernan, M. A. (2013). Causal inference under multiple versions of treatment. *Journal of Causal Inference* **1**, 1–20.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* **36**, 1–48.

Waudby-Smith, I., Arbour, D., Sinha, R., Kennedy, E. H. and Ramdas, A. (2021). Time-uniform central limit theory with applications to anytime-valid causal inference. arXiv, arXiv:2103.06476, preprint: not peer reviewed. Avaliable at: https://doi.org/10.48550/arXiv.2103.06476.

Weinberger, K. R., Harris, D., Spangler, K. R., Zanobetti, A. and Wellenius, G. A. (2020). Estimating the number of excess deaths attributable to heat in 297 united states counties. *Environmental Epidemiology* **4**, e096.

Weinberger, K. R., Wu, X., Sun, S., Spangler, K. R., Nori-Sarma, A., Schwartz, J., Requia, W., Sabath, B. M., Braun, D., Zanobetti, A., Dominici, F. and Wellenius, G. A. (2021). Heat warnings, mortality, and hospital admissions among older adults in the United States. *Environment International* **157**, 106834.

Weinberger, K. R, Zanobetti, A., Schwartz, J. and Wellenius, G. A. (2018). Effectiveness of national weather service heat alerts in preventing mortality in 20 US cities. *Environment International* **116**, 30–38.