Bayesian Structural Time Series for Biomedical Sensor Data: A Flexible Modeling Framework for Evaluating Interventions

Jason Liu[1*], Daniel J. Spakowicz[2,3*], Garrett I. Ash[4,5], Rebecca Hoyd[2], Andrew Zhang[1], Shaoke Lou[1], Donghoon Lee[6], Jing Zhang[7], Carolyn Presley[2], Ann Greene[8], Matthew Stults-Kolehmainen[9], Laura Nally[8], Julien S. Baker[10], Lisa M. Fucito[11], Stuart A. Weinzimer[8], Andrew V Papachristos[12], Mark Gerstein[11+]

[*] Co-first authors
[+] Corresponding author

[1]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA; Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA
[2]Division of Medical Oncology, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA
[3]Department of Biomedical Informatics, The Ohio State University College of Medicine, Columbus, OH, USA
[4]Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA
[5]Center for Medical Informatics, Yale University
[6]Department of Genetics and Genomic Sciences, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY
[7]Department of Computer Science, University of California, Irvine, CA
[8]Department of Pediatrics, Yale School of Medicine, New Haven, CT, USA
[9]Bariatric Surgery Program, Yale-New Haven Hospital, New Haven, CT, USA
[10]Faculty of Social Sciences, Hong Kong Baptist University, Kowloon Tong, Hong Kong
[11]Department of Psychiatry, Yale School of Medicine, New Haven, CT USA
[12]Department of Sociology, Northwestern University, Chicago, IL, USA
[13]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA; Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA; Department of Computer Science, Yale University, New Haven, CT, USA. Electronic address: pi@gersteinlab.org

ABSTRACT

The development of mobile-health technology has the potential to revolutionize personalized medicine. Biomedical sensors (e.g. wearables) can assist with determining treatment plans for individuals, provide quantitative information to healthcare providers, and give objective measurements of health, leading to the goal of precise phenotypic correlates for genotypes. Even though treatments and interventions are becoming more specific and datasets more abundant, measuring the causal impact of health interventions requires careful considerations of complex covariate structures as well as knowledge of the temporal and spatial properties of the data. Thus, biomedical sensor data need to make use of specialized statistical models. Here, we show how the Bayesian structural time series framework, widely used in economics, can be applied to these data. We further show how this framework corrects for covariates to provide accurate assessments of interventions. Furthermore, it allows for a time-dependent confidence interval of impact, which is useful for considering individualized assessments of intervention efficacy. We provide a customized biomedical adaptor tool around a specific Google implementation of the Bayesian structural time series framework that uniformly processes, prepares, and registers diverse biomedical data. We apply the resulting software implementation to a structured set of examples in biomedicine to showcase the ability of the framework to evaluate interventions with varying levels of data richness and covariate complexity. In particular, we show how the framework is able to evaluate an exercise intervention's effect on stabilizing blood glucose in a diabetes dataset. We also provide a future-anticipating illustration from a behavioral dataset showcasing how the framework integrates complex spatial covariates. Overall, we show the robustness of the Bayesian structural time series framework when applied to biomedical sensor data, highlighting its increasing value for current and future datasets.

INTRODUCTION

Background

The wearable sensor market was valued at $10.8 billion USD in 2019 and is expected to triple in value over the next five years (1). Investment is bolstered by the great potential for advancing personalized medicine in the near future (2-4). Whereas medicine has previously focused on determining the right interventions, it is now more focused on for whom and when (5). Identifying the right time for and timing of treatments remains relatively understudied, but this trend is expected to change soon as large streams of sensor data are released (6). As sensor technology develops, data-rich features such as physical, chemical, behavioral, and biological variables will be measurable. In addition to time series data, spatial information is becoming more popular as well (7), all of which can be used for a more detailed understanding of sensors. A survey of distinct types of data are presented in Figure 1.

Though increasing amounts of sensor data is being released, there still is a paucity of analytical methods in the biomedical field that can accommodate the complex covariate structures as well as the temporal and local trend considerations necessary to analyze these longitudinal data. Existing methods applied to these situations are limited. They address longitudinally observed patterns that could inform the timing of an intervention but do not evaluate the impact and efficacy of such an intervention.

While the nature of data from sensors and wearables will vary depending on the context, most of them share certain properties. A variety of such data can be seen in Figure 1, all of which come from various contexts. It is important to discern these properties and to establish a flexible model for emerging sensor and wearable data, as it will have broad implications in fields such as personalized medicine (2). Therefore, in this study we aim to contextualize a model and establish a flexible statistical framework to model various sensor and mobile health data. The model we adapted here is a combination of the Bayesian structural time series model and the Causal Impact model from Google (8). The principles of this modeling framework stem from Bayesian inference and the analysis of time series data, which have been well established for decades (9). While the idea

of using Bayesian inference models has been extensively used in fields such as finance (10), ad campaigns (11), and marketing (12), they have rarely been used in the context of wearable and sensor data analysis to specifically evaluate impact intervention over various time periods. A survey of the application of Bayesian modeling of biosensor time series data yields few results. Therefore, we hope to apply this widely accepted statistical framework to the specialized context of biomedicine. By doing so we showcase the effectiveness when applied to a variety of sensor data we collected and accurately assess the impact that various interventions have on individuals.

Modelling Framework

To establish an intuitive modeling framework for data taken from sensors and other mobile health sources, it is important to lay out the structure of how the framework is formed as well as any assumptions that may be essential for the model. First, we must recognize that the dataset is a time series data with a known response variable that evolves over time. For example, this could be a person's weight. This response variable could be a direct measurement taken from a sensor, or it could be a derived value – some scalar calculated from various underlying variables measured by a device – such as calories burned (13) or a Nike Fuel Score (14). The response variable's evolution in time is important, as it should be dynamic and change based on covariates and interventions. In the example of weight, the measured weight fluctuates over time based on things such as seasonality (15), temperature (16), or diet (17).

Because we are interested in the impact that an intervention has on such a variable, an important assumption is that we know when the intervention occurs as well as the duration. There exist methods that detect intervention times (18,19) though to jointly assess the impact would result in a significant statistical disadvantage due to the multiple hypothesis corrections since the predictions from these models only account for one time point rather than a set of time points. Therefore, by having prior knowledge on the time of occurrence and duration of the intervention, we are able to more accurately assess the impact on an individualized level.

Given a specific intervention and its duration, we can split our time series into two segments – the pre- and post- intervention periods. These periods are used to describe all variables, known and unknown, as well as any parameters associated with them. For the example of weight and diet, the pre-intervention would be the period before the diet starts and the post-intervention is the period after the diet starts. This is distinct from "post-cessation of intervention" (e.g., return to unhealthy eating following several weeks of dieting) which is not addressed by our present model.

Our goal is to model the pre-intervention period using covariates to most accurately assess any changes in behavior in the response variable before any intervention has occurred. We also assume that any covariates used in our model should not be affected by the intervention, to provide a similar control in the post-intervention period. Though the post-intervention covariates themselves may change in value, the assumption is that they are derived from the same distribution as in the pre-intervention. Finally, using the model derived from the pre-intervention period and the covariates in the post-intervention, we can calculate a counterfactual in the post-intervention. The counterfactual predicts the response variable without intervention. It serves as a baseline to compare the actual observation in the post-intervention and ultimately is what we use to calculate the impact of an intervention. Compared to linear models, this framework allows for an evolving measure of impact, due to the dynamic confidence interval for the difference between counterfactual and observation that is inherent to using Bayesian structural time series. This temporal consideration, in addition to the more common advantage of using hyper-parameters and priors, is an important consideration for the Bayesian framework, and sets it apart from other models.

This Bayesian structural time series framework can make use of complex covariate structures, which is useful and necessary to get an unbiased measure of impact. Two main types of covariates can be used: those that have known effects on the response and those that can account for hidden effects. The first type refers to covariates that could be correlated to the response, or cause changes to the response unrelated to the intervention. In relation to our example about weight, covariates of this type may include temperature, weather, or season. Since it is very unlikely to know all covariates of the first type that may perturb the response variable, we also introduce the notion of a second type of covariate, termed a "paired covariate". The paired covariate is

an independent stream of sensor data of the same type as our original data in question. They could be collected in different locations, but the ideal case is they receive little to no treatment or intervention. If we imagine a scenario where the roommate of our subject is subject to many of the same external factors, but the roommate does not participate in the dieting, then this roommate's weight could be considered the paired covariate. Since both the original data and paired covariate share underlying biases, by using the paired covariates we can determine the true impact of the intervention.

As a whole, the Bayesian structural time series framework addresses many of the limitations of other methods and serves to advance the field by providing a structured method for analyzing data from sensors and mobile health sources. We give more details in the Statistical Formalism section and provide a schematic of the framework in Figure 2.

STATISTICAL FORMALISMS

Bayesian Structural Time Series

In order to understand the causal impact of interventions on longitudinal datasets that may be affected by a wide range of factors it is important to have a statistical framework that considers carefully the prior and posterior measurements as well as covariates that may affect the response variable. We detail aspects of this framework below.

We assume for a given time point, t, there exists an observation y_t, linked to a variety of other parameters, $a_t: \{\mu_t, \delta_t, \boldsymbol{x}_t, \boldsymbol{\eta_t} \sim N(0, \boldsymbol{\sigma_t^2})\}$. As time progresses from t to $t+1$, the other parameters also progress due to their time dependency.

There also exists a set of hyper-parameters $\pi: \{\varepsilon_t, \eta_t \sim N(0, \sigma_t^2), \ldots\}$ that define the initial conditions of the parameters above. The hyper-parameter ε_t is a prior that reflects previous knowledge about the system and is used for the initial set of parameters in a_t. More specifically, we use the following equations to represent our time series data.

$$y_t = \mu_t + \tau_t + \beta^T \boldsymbol{x}_t + \eta_t^y$$

$$\mu_{t+1} = \mu_t + \delta_t + \eta_t^\mu$$

$$\delta_{t+1} = \delta_t + \eta_t^\delta$$

Here we see that the equation involving the observation or response variable, y_t, depends on μ_t, and additionally a variable $\boldsymbol{x}_t$ to represent the covariates. Furthermore, we additionally include another layer of dependency, δ_t, which μ_t depends on. Because our observation, y_t, can be biased, it is important to take into account as many covariates in the form of $\boldsymbol{x}_t$. Each equation has an error term represented by η_t. Since each of the processes are different, the error term is also different. For example, the error associated with μ_t is η_t^μ. These η_t are drawn from some distribution satisfying $N(0, \sigma_t^2)$. Similar to other regression-based models, the covariates here are scaled by a coefficient vector β^T. The model also employs a spike and slab method to

penalize and select for important covariates when the covariate matrix is large. Additionally, one important feature of each covariate is the posterior inclusion probability (PIP), which can be calculated from the sum of all posterior probabilities of all regressions that include that particular covariate (20). The PIP gives a ranking measure to show how favorable the inclusion of a particular covariate is. The variable μ_t, the underlying "state" of the response, contributes to our response variable, and its representation is given as a time dependent equation. If we ignore the δ_t variable, our equation becomes

$$\mu_{t+1} = \mu_t + \eta_t^{\mu}$$

and this is a representation for the random walk. That is to say, in this simplified form of μ_t, our observation is a random walk, where at each time point, there is some progression to the next time point in random fashion, based on the parameter $\eta_{t,\mu}$. However, one optimization that can be done for this parameter is the inclusion of δ_t, which results in the pair of equations

$$\mu_{t+1} = \mu_t + \delta_t + \eta_t^{\mu}$$

$$\delta_{t+1} = \delta_t + \eta_t^{\delta}$$

Here δ_t follows a random walk, and the μ_t equation is dependent on δ_t. That is to say, δ_t serves as a trajectory or slope parameter that helps to guide the behavior of μ_t. We can still imagine μ_t as a random walk, but now with some additional trend vector. If μ_t is steadily increasing, it is likely that at μ_{t+1} there is also an increase, due to the inclusion of δ_t. In other words, the δ_t allows for more stability between each time point and serves a purpose similar to a slope parameter. We provide more examples in Supplemental Figures S1-5.

Evaluating Intervention Impact

While the above model describes longitudinal sensor data well, it is important to consider how the pre- and post- intervention periods differ. To determine if an intervention was effective in bringing about a change in our dependent measurement, y_t, it is important to have a framework in which comparisons can be quantified. To do so, we define the pre-intervention period as time points $t = 1, \dots, n$, while the post-intervention period is defined as $t = n + 1, \dots, N$. Furthermore, we define observations as y and predictions of the model as y'. Therefore, the set of observations in the pre-intervention period, $y_{1,\dots,n}$, serve as the training data, based on all

covariates. At each time point $t = 1, \dots, n$, we update our parameter set of the model to better fit $y_{1,\dots,n}$ using a Markov Chain Monte Carlo (MCMC) method of sampling from the posterior distribution. After estimation of parameters in the pre-intervention period, we predict the counterfactual in the post-intervention, $y'_{n+1,\dots,N}$. This is done by using the parameter set defined in the pre-intervention period and the covariates from the post-intervention period. Because $y'_{n+1,\dots,N}$ has an associated prediction interval, we can calculate a significance p-value associated with the difference between $y'_{n+1,\dots,N}$ and $y_{n+1,\dots,N}$. While we can get a p-value at every time point in the post intervention, it is more useful to consider the impact that an intervention had on the whole post-intervention period. The p-value associated with the full post-intervention time period ($t = n + 1, \dots, N$) is known as the *cumulative impact*. It should be noted that the prediction interval associated with $y'_{n+1,\dots,N}$ generally will increase as time progresses. This is one advantage of the Bayesian model and allows for an evolving cumulative impact. This is due to the fact that at every time point after the intervention, the variance associated with the distribution that the prediction is drawn from is compounded at every time point. It is useful to factor in this temporal aspect since the confidence that an intervention resulted in a causal impact may be dependent on time.

Biomedical Adapter Tool and Software Implementation

We provide a customized biomedical adaptor tool around a specific Google implementation of the Bayesian structural time series framework that uniformly processes, prepares, and registers diverse biomedical data. Specifically, time series data from biosensor data that are measured at different time points and intervals are unified so that given a time interval set $t = \{1, \dots, N\}$, there exists a set of variables (observation and covariates), y_t and $\boldsymbol{x}_t$ for each time point in t. The adaptor tool that applies Causal Impact and bsts to evaluate a user defined intervention and gives a report of results (8,21).

RESULTS

To showcase the wide applicability of the Bayesian structural time series framework for biomedical applications, we provide examples of analysis from various types of sensor data. First, we apply the Bayesian structural time series framework on a toy example where we collected data from an Android phone sensor in order to give intuition for and demonstrate several key features of our model. Second, we apply the model to a real-world example – environmental sensor data collected from an air quality device – and show the usefulness of our model in identifying the effect of simple interventions in real world data. Third, we provide a core biomedical example that showcases a strong application of the framework and the potential it has in personalized medicine. In particular, we collected extensive biosensor data from a diabetes and exercise study which aims to understand how structured exercise can help to stabilize glucose levels throughout the day. The framework was then used to assess the impact of an exercise regimen on clinical health markers of glucose level. We find a strong stabilizing effect on glucose after this exercise intervention. While these data are informative, they currently lack several attributes that we expect future biomedical data to have. In particular, we expect future biomedical datasets to have a large number of context-rich covariates (e.g. activity in different locations or environments) that can dramatically increase intervention assessment. Thus, we give a final example dataset rich in covariates to showcase how these (paired) covariates can be used creatively to more accurately assess the effect of interventions. Specifically, we collected human behavioral sensor data to showcase the patterns of crime across different neighborhoods in a city, and how the introduction of a mobile application aimed at increasing social cohesion affected these patterns. Below, we give more details regarding each of the specific sub-studies and their corresponding findings.

Toy Example: Google Android sensor

As a toy example to illustrate how the Bayesian structural time series framework works, we used Google Science Journal app on Android and collected various mobile sensor data for a simple experiment involving a person spinning in different scenarios. The goal of this analysis was to demonstrate the flexibility of our model

in determining significant causal impact of human physical actions before and after an intervention as well as to provide intuition for the utility of paired covariates at reducing the effects of noise. A schematic of the results is given in Figure 3A, B, and C. First, a single user's longitudinal data was used to estimate the effect of an intervention (Figure 3A). An example of this could be sleep data and the effect of changing one's pillow, and where no other data are available to the user. Here we show an individual spinning with a sensor measuring acceleration when held at arm's length; the intervention is to bring the sensor closer to the body while spinning at the same rate, thereby increasing the acceleration. The model successfully detected the state change associated with the intervention, as shown by the cumulative change relative to the predicted values.

Next, we demonstrate how the previous scenario is susceptible to noise, i.e. something that affects the signal but is not related to the intervention. In this case the noise is represented by a hop, which occurred after the intervention began and was therefore not predicted by the model (Figure 3B). In this case the model does not detect an effect of the intervention – the hop affected the signal to such a degree that the confidence in the prediction decreased, i.e. the credible intervals widened greatly. To continue the sleep study analogy mentioned above, the noise could be loud neighbors moving in next door (adding noise in a literal sense), which leads to poor sleep in a way that is unrelated to the pillow intervention.

Finally, we show how including a paired covariate -- another data stream that affects the state being measured but is not related to the intervention – can effectively correct for noise (Figure 3C). In this case two sensors are held while spinning, but only one is subject to the intervention (brought close to the chest). Both, however, experience the noise (hop), which the model is then able to control and correctly identifies the intervention like the case without noise (Figure 3A). In the sleep analogy, an effective paired covariate could be another device measuring the sleep of one's partner sharing the same bed, but who did not change their pillow. This is distinct from a non-paired covariate which could be similarly effective, e.g. a device measuring the decibel level in the room; it is relevant to the state (sleep quality), is not affected by the intervention (pillow change) and identifies the noise that affects the state (loud neighbors). However, a paired covariate has the characteristic that it is able to control for unknown confounders, akin to a control arm of a randomized controlled trial.

Simple Real-world Example: Environmental Sensor Data

We then transitioned to a simple real-world example where we performed a similar analysis using data collected from an AWAIR monitor, with measurements of CO_2, dust, humidity, and temperature (Figure 1E). The data were collected over a one-month period, with measurements taken every 15 minutes. The data were aggregated at each time point across all 30 days to give a smoother signal and limit the analysis to one intervention with a clearly delineated a pre- and post- intervention period, namely exposure of the room to people. It should be noted that though intervention often implies some treatment put in place, it can be widely adapted in the realm of Bayesian structural time series to allow for any disruption or change in status such as the effect that people have on environmental CO_2 levels. After correcting for covariates such as dust, humidity, and temperature (22), there was a significant increase in CO_2 in the hours when workers were in in the room with the sensor (p-value < 0.001). Figure 3D shows the CO_2measurements across the aggregated time points in a 24-hour time frame. We can see that as CO_2 levels increased drastically after the 9am time point, the cumulative causal impact in the post-intervention period increased. This cumulative impact increases until around 5pm in the evening, where the cumulative impact tapered off, signaling a state was reached in the post-intervention that was similar to that of the pre-intervention period. Although there is one p-value associated with the whole post-intervention period, this framework can show confidence intervals for each time point in the post-intervention period as a function of both the observation and covariates. This helps us to define period of time where the intervention is most effective.

Core Biomedical Example: Clinical Sensor Data

We next provide a core biomedical example of how the framework can be applied to clinical sensor data. In particular, we collected biosensor data from a person with type 1 diabetes over a 12-week period who completed an exercise regimen (intervention). We chose to model this set of biomedical data because it is arguably one of the most established applications of personalized medicine today (23,24). People with type 1 diabetes are advised to intensively manage their blood glucose to maintain it within the target range. Since this

process is quite involved and requires continuous adjustments based on numerous biological factors, the result is a complex and high-stakes problem in personalized medicine.

Figure 4A shows the data from a continuous glucose sensor and insulin pump as well as a comprehensive set of Apple Watch data from the study. The 12-week evaluation spanned an initial 2-week sedentary period followed by a 10-week exercise regimen. The Apple Watch and the insulin pump used by the patient provided several potential covariates, which could help us understand the glucose sensor data. We took glucose readings from the participant's glucose sensor and aggregated them into 24-hour values by transforming the values into two clinically relevant indicators of glucose stability, percent-in-target and percent-above-target (25). In general, values above the target range are predictive of long-term organ damage, while values below the target range lead to acute hypoglycemia (26), a state of low blood glucose with symptoms that can range from mild fatigue and confusion to life-threatening coma, posing immediate threats to safety as well as long-term psychological consequences (e.g., fear of hypoglycemia (27)).

Maintenance of glucose levels in the target range is achieved by strategically managing a triad of factors: insulin administration, diet, and exercise. The timing and dosing of insulin (which decreases blood glucose) with carbohydrate ingestion (which increases blood glucose) must be carefully balanced. The role of exercise, however, is less clearly defined because it can either decrease or increase blood glucose during and up to 24hr after a session. Determinants of the direction and magnitude of the glycemic response to exercise are numerous and include **1)** exercise characteristics (high vs. moderate or low intensity), **2)** individual characteristics (endogenous insulin sensitivity, and the effect of physical fitness to increase insulin sensitivity), and **3)** contextual factors (pre-exercise blood glucose level, insulin- and carbohydrates-on-board, and concentration of counter regulatory hormones) (23).

Using our Bayesian structural time series framework, we find that the exercise regimen was effective in increasing the percent-in-target and decreasing the percent-above-target (p = 0.014 and 0.002 respectively). These results are shown in Figure 4B and 4C. Thus, our model supports the existence of a positive causal effect of this particular exercise regimen on maintaining a healthy glucose level for this individual. Furthermore, of all the covariates we used, insulin on board (IOB) was found to have the highest posterior inclusion probability,

which is a ranking measure used to see how favorable the inclusion of a variable is in the model. This is reasonable since insulin is essential to glucose control.

Example with Data Rich in Covariates: Human Behavioral Data

In our final example, we demonstrate how data rich in covariates can significantly improve assessments of intervention impact. While the example we provide is not clinical in the traditional sense, we provide this example as a way to showcase how the Bayesian structural time series framework performs when given extensive data and paired covariates – an important consideration for future biomedical data. One aspect of a rich covariate set could be location information, an important factor in many clinical or personal health applications. For example, many watches collect GPS tracks of a run as well as steps and heart rate (e.g., Apple Watch, Garmin Forerunner). The Bayesian structural time series framework can flexibly accommodate location data through the inclusion of spatial correlation matrices. In addition, the spatial data can be used to segment the data where the intervention does and does not have an effect. By this method one can create "synthetic" paired-covariates, in that data are from different spatial segments from the same source. However, spatial data are often protected for privacy concerns and are therefore less accessible to researchers. We therefore demonstrate the utility of the framework on more accessible behavioral data that share many characteristics: the effect of a social monitoring application called SeeClickFix on negative behavioral patterns (crime).

Crime patterns show similar characteristics to spatial mobile health data. They are affected by covariates such as temperature and precipitation (28) and can be linked to many other features such as census data of household incomes or education (29) (analogous to covariates of movement such as age, physical fitness, etc. (30)). SeeClickFix is a smartphone and web application developed to allow users to report issues in their communities including non-violent crimes. Posts can be voted on and supported by other users' comments and local government agencies acknowledge issues and post when they have been addressed. It has been hypothesized that SeeClickFix and similar tools may reduce crime through establishing social cohesion, promoting collective efficacy (31).

Following the intervention (creation of SeeClickFix) there was no detectable decrease in crime across the entire area (Figure 5A), nor in particular neighborhoods (Figure 5B). However, one would expect many other factors besides the introduction of SeeClickFix to affect crime in this time (e.g. increases in the police force, changes to local employment, city-wide initiatives). The Bayesian structural time series framework can leverage the spatial information in these behavioral data to search for a paired-covariate, in this case referring to locations not affected by the intervention (no SeeClickFix use) in order to control for other, unobserved effects on the outcome variable (increases in the city police force). We aggregated crimes and SeeClickFix posts by neighborhood (Figure 5C) and then modeled used a neighborhood that was not affected by the intervention as a control (i.e. a neighborhood without SeeClickFix posts), but that would experience any city-wide initiatives that might affect crime (Figure 5D). Neighborhoods with heavy SeeClickFix use showed an effect of the intervention on crime when controlling for unobserved factors with synthetic paired covariates (Figure 5E).

DISCUSSION

In this paper, we demonstrated how the Bayesian structural time series framework can be applied to biomedical sensor data and personalized medicine as well as created a wrapper software to facilitate the framework's use. We additionally showcase how a rich dataset with complex covariates can benefit from the framework. While some studies demonstrating the success that linear models have on understanding such data do exist (18,19,32-34), there is a lack of emphasis on considerations for the temporal aspect of interventions and how effective interventions are at different time points in the post-intervention period. Linear modeling frameworks lack the flexibility to evaluate the intervention's effect strength at all points in the post-intervention period, while existing Bayesian frameworks forecast based upon predictors (e.g., forecast continuous blood glucose based upon covariates like the ones we obtained from insulin and smartwatch devices) but do not evaluate deviation from this forecast following an intervention of known timing and duration to determine its impact (35-37). In this study, we illustrate the benefit of using a Bayesian structural time series model in modeling the behavior of various longitudinal data collected via apps and sensors. These data demonstrate properties that are commonly found across other wearable sensor data, which are increasingly gaining in popularity (3). In order to effectively integrate such data with personalized medicine and health in the future, we must understand how specific timings and types of interventions impact individual patients (2-4).

It is also important to consider the contexts in which this modeling framework does not yield significant advantage over a linear framework. A stable and consistent intervention effect over the entire post-intervention period should be evaluated equally well by linear models. Given the computational complexity of the Bayesian method due to MCMC sampling for each time point, such considerations could be very important for large datasets. Furthermore, due to the nature of this type of model and the slight degree of randomness in parameter estimation, it is possible to have varying results in calculating the impact of the intervention. This is in contrast to the results found from linear models, which generally demonstrate a singular solution.

We show that the Bayesian modeling framework can take into account the rich covariates in our behavioral sensor dataset – specifically the paired covariate structure between different locations – and give

results that are unbiased. Furthermore, the framework also considers the temporal properties, ensuring that predictions of intervention impact are conditional on duration of the intervention. This is especially important for future data from sensors and mobile health sources as one of the major concerns of personalized medicine is ensuring the right time and duration of a treatment or intervention is considered (5). With time varying confidence intervals, we can begin to understand not only the effectiveness of an intervention but determine the most effective intervention plan necessary to reach a desired result.

When applying this method to various datasets, an additional consideration is how informative the covariates are. For example, in the case of the sensor data derived from the diabetes study, even though we were able to isolate the effect of the 10-week exercise regimen upon both metrics of glucose stability, more covariates could be used in the future to improve the prediction. Similar biological studies in the future could benefit from obtaining other clinically relevant covariates such as plasma cortisol, epinephrine, growth hormone, glucagon, and directly sensed (not estimated) insulin levels or more to use high precision sensors with reduced noise. Furthermore, paired covariates (e.g. control group) should be taken into consideration when designing studies, as they can greatly increase model accuracy. While it is true that even an improved set of covariates could lead to more accurate results, we showcase here that the framework we have now is able to demonstrate results that converge with current literature showing favorable impact of exercise upon blood glucose control metrics (38). We also note that one assumption of our model is that the intervention does not affect any of the covariates in the post-intervention period. For example, though the exercise regimen is aimed at stabilizing glucose levels, we note that exercise and physical activity can have a general effect on various biological factors within the human body, including some of the covariates used. We minimize the effect of this through our transformation of variables and aggregation on a 24-hour block.

The Bayesian structural time series framework is a flexible modeling approach that, with relatively minor specialization we provide through our tool, can be applied to diverse biosensor data. It has features that few methods in biomedicine share, such as dynamic forecasts to observe the effect of an intervention and its evolution over time. These features are critical to advancing personalized medicine and realizing the challenge of relating genotype and phenotype data in the context of human research.

METHODS AND MATERIALS

Data Collection

There exist many datasets similar to those of mobile sensors and we should not confine ourselves to just the traditional accelerometer and gyroscope data that one would traditionally think of when envisioning sensor data. In studying the causal impact of an intervention, we show that most data sharing longitudinal qualities allow for development of algorithms and exploring how such algorithms can be useful in analyzing various sensor data. The data we analyze in this paper consists of environmental sensor data, physical activity sensor data and human behavioral sensor data.

Due to the longitudinal nature of sensor and wearable data, these data not only demonstrate interesting patterns in a response variable, but also are closely tied to the temporal property of a phenomena. By introducing the aspect of time, it becomes important to find models that leverage temporal considerations and use them to make accurate assessments about the data. Also, some of the data showcase complex covariates (paired) and can be used to better correct for unknown and hidden biases. More details about the data collection and data types are given below.

<u>Google science journal data collection</u>

Data were collected using the linear accelerometer measurement function in the Google Science Journal application on a Samsung Galaxy S8 smartphone. Each experiment lasted 20 seconds. One or two instruments were held at arm's length while spinning at a constant rate for 10 seconds. Next, one instrument was brought in close to the body for 10 seconds while maintaining the spinning rate. In the case of the noise simulation, the experimenter hopped once after 15 seconds (halfway through the intervention period). Data were exported from Science Journal and analyzed using the CausalImpact package in R.

<u>AWAIR Collection</u>

Data were collected from a one-month period from the AWAIR device. The device was placed in an office lab setting where individuals frequented on a daily basis. CO_2 levels were measured in units of ppm. AWAIR also measures dust, temperature and humidity, which were used as covariates.

SeeClickFix Data collection

The application SeeClickFix is a smartphone and web application developed in New Haven, Connecticut, where users report issues in their communities including non-violent crimes. SeeClickFix posts can be supported and commented on by other users, and local government agencies acknowledge and address issues. The SeeClickFix data are publicly available, providing a rich longitudinal and spatial dataset for monitoring behavior and interactions with other users and city representatives. Posts were aggregated by month for the New Haven metropolitan area and by neighborhood (n = 19) from 2007-2015.

Aggregated crime data were shared through a memorandum of understanding with the New Haven Police Department for 2000-2013. Rates were calculated using the 2014 ACS 5-year population estimates (crimes / 10K population per unit area).

Diabetes Data collection

We used data from one participant in a single-group clinical trial that was evaluating an exercise intervention for previously sedentary adults with type 1 diabetes. Participants completed a 2-week baseline period then a 10-week exercise intervention, while wearing sensors that continuously monitored blood glucose, heart rate, heart rate variability and physical activity. They continued their normal prescribed insulin therapy, and shared device-recorded dosing logs with the research team. Besides these continuous measures, we assessed chronic diabetes control at the beginning of the baseline period and the end of the 10-week intervention using blood glycosylated hemoglobin concentrations. The 10-week intervention included motivational enhancement of exercise (i.e., patient-centered exercise coaching including instructional videos) and health feedback from biosensors (full details available on clinicaltrials.gov, NCT04204733). Both were delivered through a customized mobile digital application and supported by a coach internally certified in exercise for diabetes (GlucoseZone^{TM}, Fitscript^{LLC}, New Haven, CT). The study was approved and overseen by the Yale University Institutional Review Board, and all participants provided written informed consent.

Biodata Collection. 1) 24hr blood glucose was measured every 5 minutes by the Dexcom G6 continuous glucose monitor (San Diego, CA), a subcutaneous wire sensor sampling interstitial fluid glucose content which is converted to estimated blood glucose (validated against venous blood glucose with mean absolute relative difference 9%) (39). 2). Insulin was delivered according to each participant's usual prescribed therapy. The participant in this manuscript received lispro insulin via the Tandem t-slim Control IQ pump (San Diego, CA). The pump subcutaneously infuses insulin every 5 minutes according to the patient's individualized settings and current blood glucose levels using proprietary algorithms (40). The patient can also manually dose insulin or adjust some standard settings for meals or other disturbances (e.g., planned exercise). Infusion doses are recorded, uploaded to a central server for exporting and analysis, and converted to estimated insulin on board by the manufacturer's proprietary pharmacokinetics algorithm. 3) Heart rate (beats per minute, validated against electrocardiography with mean absolute percentage error 1.1%-6.7%) (41) and heart rate variability (standard deviation of interbeat intervals, validated against electrocardiography with intraclass correlation coefficient 0.98) (42) were measured by the Apple Watch 3 (Cupertino, CA) using photoplethysmography. 4) Physical activity was measured by the Apple Watch 3 using accelerometry and converted to kcals per day (validated against calorimetry with mean absolute percentage error ~40%) (43). 5) Diabetes control was measured by glycosylated hemoglobin (DCA Vantage Analyzer (Bayer, Tarrytown, NY) at baseline and PTS Diagnostics A1cNow+ (Indianapolis, IN) at 10 weeks).

Clinical Outcomes. The participant was a 63-year-old white non-Hispanic female with type 1 diabetes for 50 years, receiving 81 units/day of insulin (0.9 units/kg body weight/day) and performing no regular exercise at baseline. During the 10-week intervention she received 93 units/day of insulin (1.1 units/kg body weight/day) and exercised on average 2.5 days per week, 26 minutes per session, at easy to moderate intensity (Borg rating of perceived exertion 2.5 / 10). The exercise routines were dynamic, interval-based, and equally emphasized all major muscle groups. Her chronic diabetes control indicated by glycosylated hemoglobin improved from 8.6% at baseline to 6.9% at 10 weeks, achieving the recommended target of less than 7.0% (44). Processing and analysis of all data was done in R and Python and our packaged tool can be found at https://github.com/gersteinlab/scf.

ACKNOWLEDGEMENTS

This work was funded by the National Institutes of Health (NIH).

GA was supported by a fellowship from the Office of Academic Affiliations at the United States Veterans Health Administration.

REFERENCES

(1) Mordor Intelligence. Wearable sensors market: growth, trends and forecast (2020 - 2025). [Cited July 19, 2020]. Available from: https://www.mordorintelligence.com/industry-reports/global-wearable-sensors-market.

(2) Hekler EB, Klasnja P, Chevance G, Golaszewski NM, Lewis D, Sim I. Why we need a small data paradigm. BMC Med 2019 Jul 17;17(1):133.

(3) Sim I. Mobile devices and health. N Engl J Med 2019:956-968.

(4) Whitney RL, Ward DH, Marois MT, Schmid CH, Sim I, Kravitz RL. Patient perceptions of their own data in mHealth technology-enabled N-of-1 trials for chronic pain: qualitative study. JMIR Mhealth Uhealth 2018 Oct 11;6(10):e10291.

(5) Corwin E, Redeker NS, Richmond TS, Docherty SL, Pickler RH. Ways of knowing in precision health. Nurs Outlook 2019 Jul - Aug;67(4):293-301.

(6) FitBit. Fitbit's 100+ billion hours of resting heart rate user data reveals resting heart rate decreases after age 40. 2018 Feb 14 [Cited 2020 Jul 19]. Available from: https://investor.fitbit.com/press/press-releases/press-release-details/2018/Fitbits-100-Billion-Hours-of-Resting-Heart-RateUser-DataReveals-Resting-Heart-Rate-Decreases-After-Age-40/default.aspx

(7) Seshadri DR, Li RT, Voos JE, Rowbottom JR, Alfres CM, Zorman C. A., et al. Wearable sensors for monitoring the internal and external workload of the athlete. NPJ Digit Med 2019:1-18.

(8) Brodersen K, Gallusser F, Koehler J, Remy N, Scott S. Inferring causal impact using Bayesian structural time-series models. Ann Appl Stat 2015;9:247-274.

(9) Fienberg SE. When did Bayesian inference become “Bayesian”? Bayesian Anal 2006;1:1-40.

(10) Jacquier E, Polson N. Bayseian methods in finance. In: Geweke J, Koop G, Van Dijk H, editors. The Oxford handbook of Bayesian econometrics. Oxford: Oxford University Press; 2011. doi: 10.1093/oxfordhb/9780199559084.013.0010

(11) Jin Y, Wang Y, Yunting S, Chan D, Koehler J. Bayesian methods for media mix modeling with carryover and shape e□ects. [2017]. [Cited 2020 Jul 19]. Available from: https://research.google/pubs/pub46001.pdf

(12) Rossi P, Allenby G. Bayesian statistics and marketing. Mark Sci 2003;22:304-328.

(13) Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781.

(14) Nike. What is Nikefuel? 2014 Aug 31 [cited 2020 Jul 19]. Available from: https://news.nike.com/news/what-is-nikefuel

(15) Schoeller DA. The effect of holiday weight gain on body weight. Physiol Behav 2014 Jul;134:66-69.

(16) Naya DE, Naya H, White CR. On the interplay among ambient temperature, basal metabolic rate, and body mass. Am Nat 2018 Oct;192(4):518-524.

(17) Geddes R, Harvey JD, Wills PR. The molecular size and shape of liver glycogen. Biochem J 1977 May 1;163(2):201-209.

(18) Hebert ET, Stevens EM, Frank SG, Kendzor DE, Wetter DW, Zvolensky MJ, et al. An ecological momentary intervention for smoking cessation: the associations of just-in-time, tailored messages with lapse risk factors. Addict Behav 2018 Mar;78:30-35.

(19) Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 3;53(6):573-582.

(20) Tsangarides CG. A Bayesian approach to model uncertainty. IMF Working Paper 2004;4:68.

(21) Scott SL, Varian H. Predicting the present with Bayesian structural time series. Int J Mathematical Modelling and Numerical Optimisation 2014;5:4-23.

(22) Lipsett MJ, Shusterman DJ, Beard RR. Inorganic compounds of carbon, nitrogen, and oxygen. In: Clayton GD, Clayton FD, editors. Patty's industrial hygiene and toxicology. New York: John Wiley & Sons; 1994. pp. 4523–4554.

(23) Chetty T, Shetty V, Fournier PA, Adolfsson P, Jones TW, Davis EA. Exercise management for young people with type 1 diabetes: A structured approach to the exercise consultation. Front Endocrinol (Lausanne) 2019 Jun 14;10:326.

(24) Riddell MC, Gallen IW, Smart CE, Taplin CE, Adolfsson P, Lumb AN, et al. Exercise management in type 1 diabetes: a consensus statement. Lancet Diabetes Endocrinol 2017 May;5(5):377-390.

(25) Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the International Consensus on Time in Range. Diabetes Care 2019 Aug;42(8):1593-1603.

(26) Diabetes Control and Complications Trial Research Group, Nathan DM, Genuth S, Lachin J, Cleary P, Crofford O, et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993 Sep 30;329(14):977-986.

(27) Cox DJ, Irvine A, Gonder-Frederick L, Nowacek G, Butterfield J. Fear of hypoglycemia: quantification, validation, and utilization. Diabetes Care 1987 Sep-Oct;10(5):617-621.

(28) Mares D. Climate change and crime: monthly temperature and precipitation anomalies and crime rates in St. Louis, MO 1990–2009. Crime Law Soc Change 2013;59:185-208.

(29) Reitzel JD. Socioeconomic status and offending. In: Reitzel JD, editor. The encyclopedia of women and crime. New York: John Wiley & Sons; 2019. pp. 1-2.

(30) Aleksovska K, Puggina A, Giraldi L, Buck C, Burns C, Cardon G, et al. Biological determinants of physical activity across the life course: a "Determinants of Diet and Physical Activity" (DEDIPAC) umbrella systematic literature review. Sports Med Open 2019 Jan 8;5(1):2.

(31) Mergel I. Distributed democracy: SeeClickFix.com for crowdsourced issue reporting. 2012 Jan 27 [Cited 2020 Jul 19]. Available from: https://ssrn.com/abstract=1992968

(32) Suchting R, Hebert ET, Ma P, Kendzor DE, Businelle MS. Using elastic net penalized cox proportional hazards regression to identify predictors of imminent smoking lapse. Nicotine Tob Res 2019 Jan 4;21(2):173-179.

(33) Koslovsky MD, Swartz MD, Chan W, Leon-Novelo L, Wilkinson AV, Kendzor DE, et al. Bayesian variable selection for multistate Markov models with interval-censored data in an ecological momentary assessment study of smoking cessation. Biometrics 2018 Jun;74(2):636-644.

(34) Koslovsky MD, Hebert ET, Swartz MD, Chan W, Leon-Novelo L, Wilkinson AV, et al. The time-varying relations between risk factors and smoking before and after a quit attempt. Nicotine Tob Res 2018 Sep 4;20(10):1231-1236.

(35) Zecchin C, Facchinetti A, Sparacino G, Dalla Man C, Manohar C, Levine JA, et al. Physical activity measured by physical activity monitoring system correlates with glucose trends reconstructed from continuous glucose monitoring. Diabetes Technol Ther 2013 Oct;15(10):836-844.

(36) Ewings SM, Sahu SK, Valletta JJ, Byrne CD, Chipperfield AJ. A Bayesian network for modelling blood glucose concentration and exercise in type 1 diabetes. Stat Methods Med Res 2015 Jun;24(3):342-372.

(37) Ruan Y, Wilinska ME, Thabit H, Hovorka R. Modeling day-to-day variability of glucose-insulin regulation over 12-week home use of closed-loop insulin delivery. IEEE Trans Biomed Eng 2017 Jun;64(6):1412-1419.

(38) Wu N, Bredin SSD, Guan Y, Dickinson K, Kim DD, Chua Z, et al. Cardiovascular health benefits of exercise training in persons living with type 1 diabetes: a systematic review and meta-analysis. J Clin Med 2019 Feb 17;8(2):253.

(39) Aleppo G, Webb K. Continuous glucose monitoring integration in clinical practice: a stepped guide to data review and interpretation. J Diabetes Sci Technol 2019 Jul;13(4):664-673.

(40) Brown SA, Kovatchev BP, Raghinaru D, Lum JW, Buckingham BA, Kudva YC, et al. Six-month randomized, multicenter trial of closed-loop control in type 1 diabetes. N Engl J Med 2019 Oct 31;381(18):1707-1717.

(41) Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34.

(42) Hernando D, Roca S, Sancho J, Alesanco Á, Bailón R. Validation of the Apple Watch for heart rate variability measurements during relax and mental stress in healthy subjects. Sensors (Basel) 2018 Aug 10;18(8):2619. doi: 10.3390/s18082619.

(43) Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):3.

(44) American Diabetes Association. 6. Glycemic targets: Standards of Medical Care in Diabetes-2020. Diabetes Care 2020 Jan;43(Suppl 1):S66-S76.

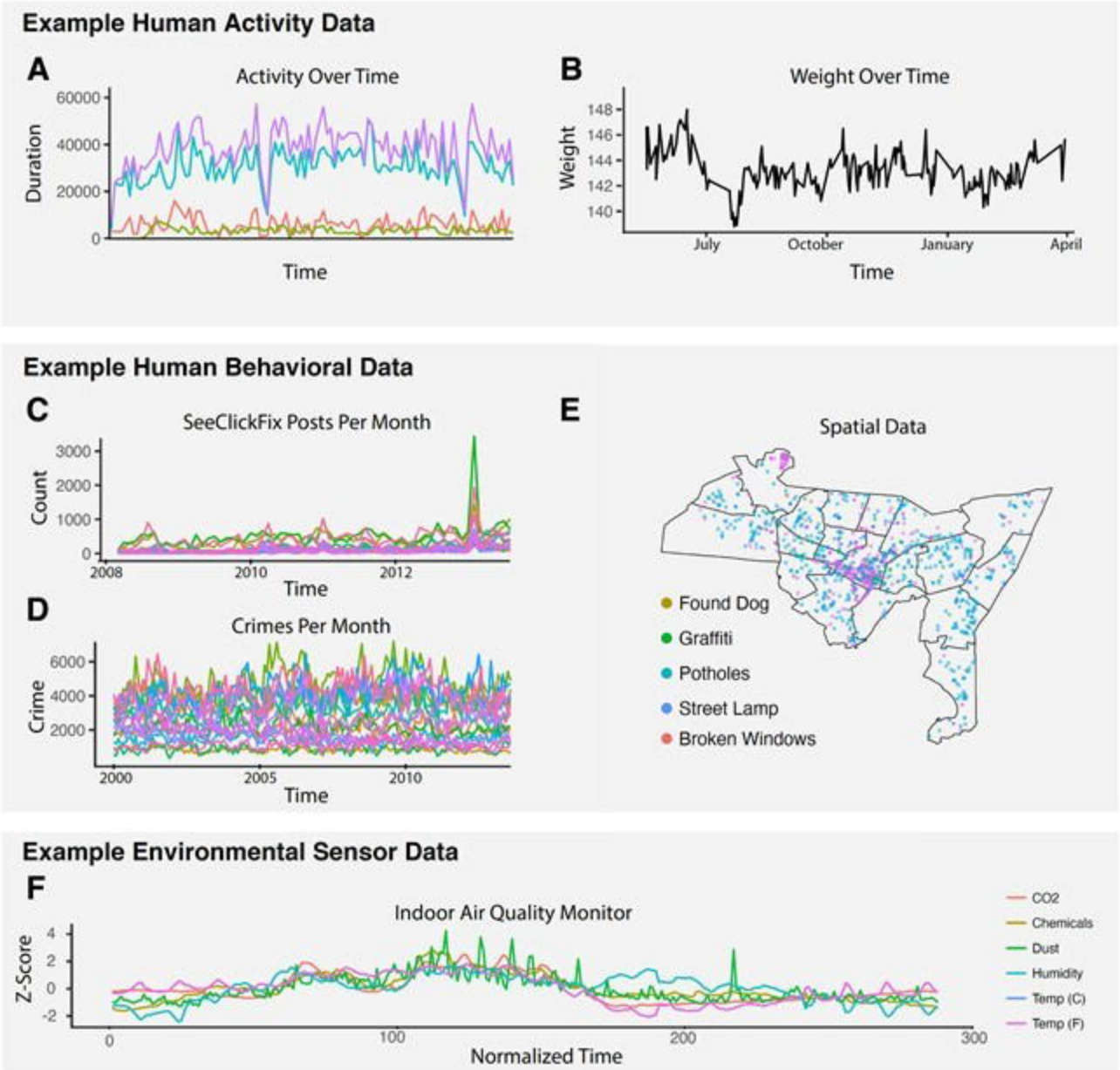

Figure 1. A) MOVES data set showing different exercise activities over time, with each color representing a different activity. B) Weight over time from the Withings data set. C) An example of human behavioral data, the use of SeeClickFix, over time. D) Various neighborhoods in New Haven and their crime data over time. E) Spatial map of New Haven illustrating the use of SeeClickFix and the associated geographical data. F) Various measurements taken from an indoor air quality sensor averaged over a day across a month.

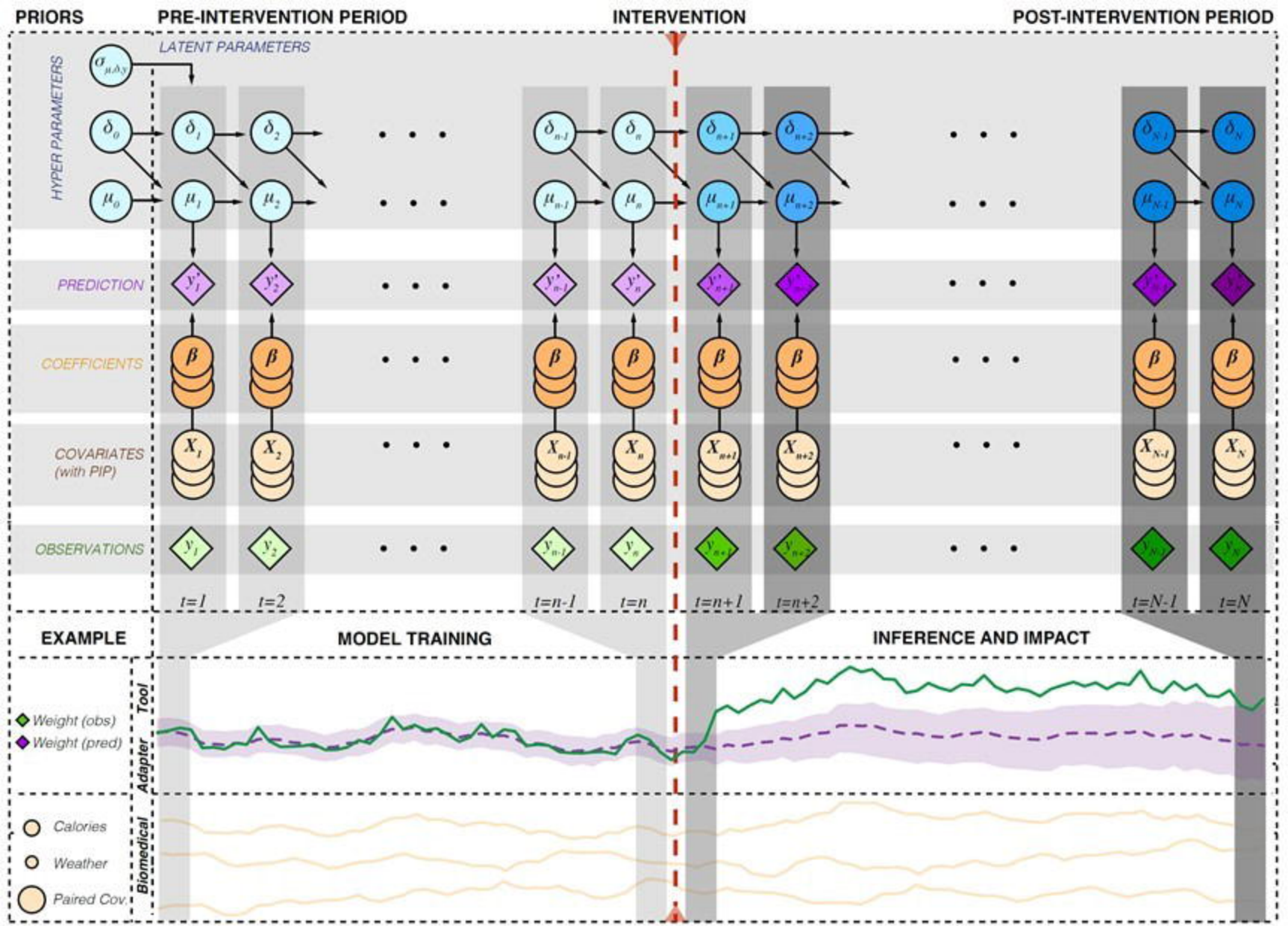

Figure 2. Schematic and illustration of the Bayesian structural time series. Latent parameters and hyperparameters are shown in blue, observations are shown in green, covariates in yellow (with a property of posterior inclusion probability), coefficients in orange, and predictions in purple. Predictions have an associated prediction interval shown in light purple. An illustrative example shows weight over time, with various covariates, being modeled. Covariate posterior inclusion probability is given by the size of the circle. Assuming an intervention of increased diet, the model detects a strong impact on the post-intervention weight.

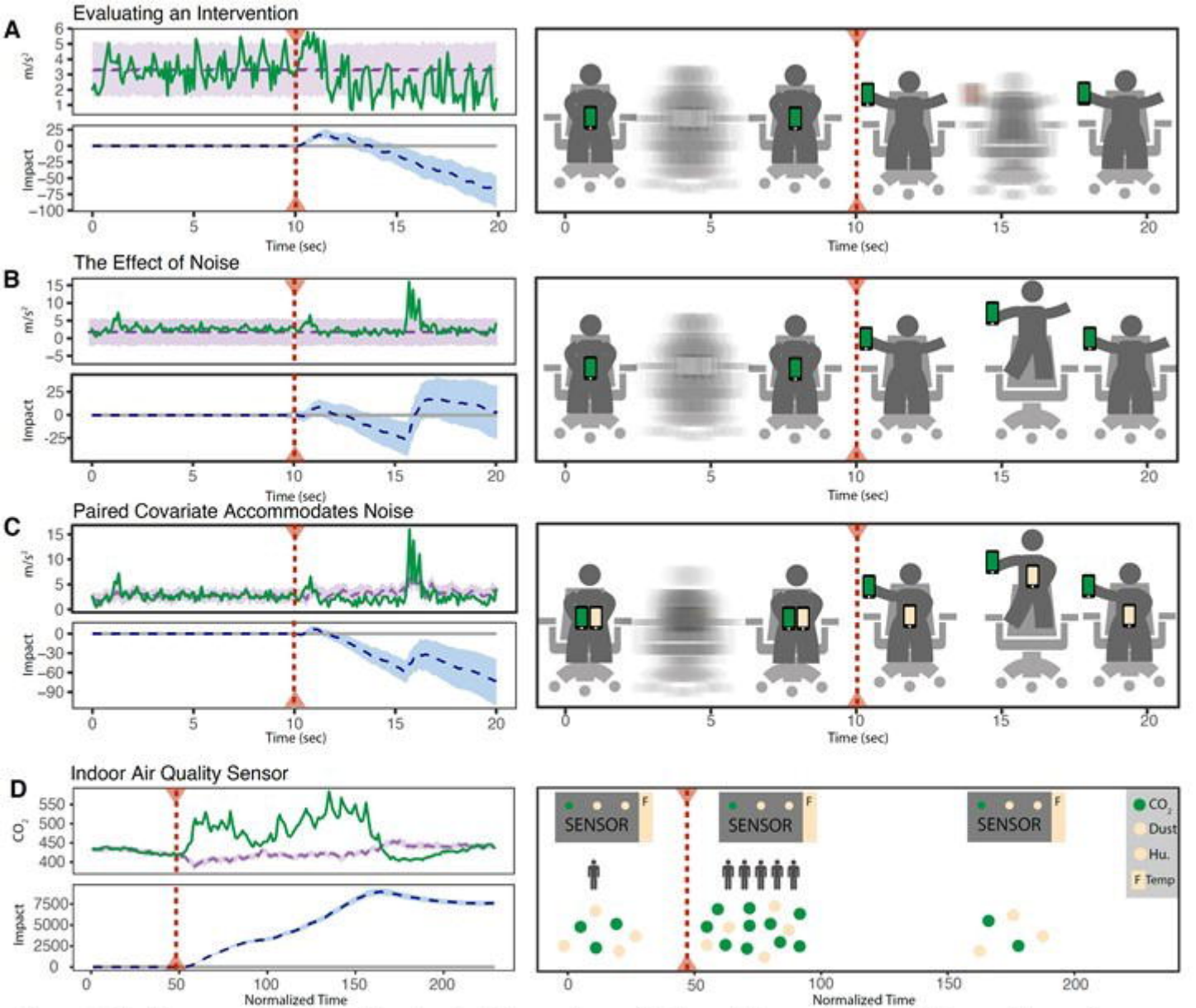

Figure 3. Performance of the Bayesian structural time series model in model experiments with known interventions. Accelerometer measurements of (A) a person spinning in a chair holding sensor near body and then extended to arm's length at the intervention; (B) with simulated "noise" produced by a hop during the intervention period and (C) with a paired covariate (second sensor) that is not affected by the intervention but experiences the "noise" hop. D) Using an indoor air quality sensor, CO2 is measured with a variety of other covariates. The intervention (arrival of office members) causes an increase in CO2 and is determined to be impactful using the model.

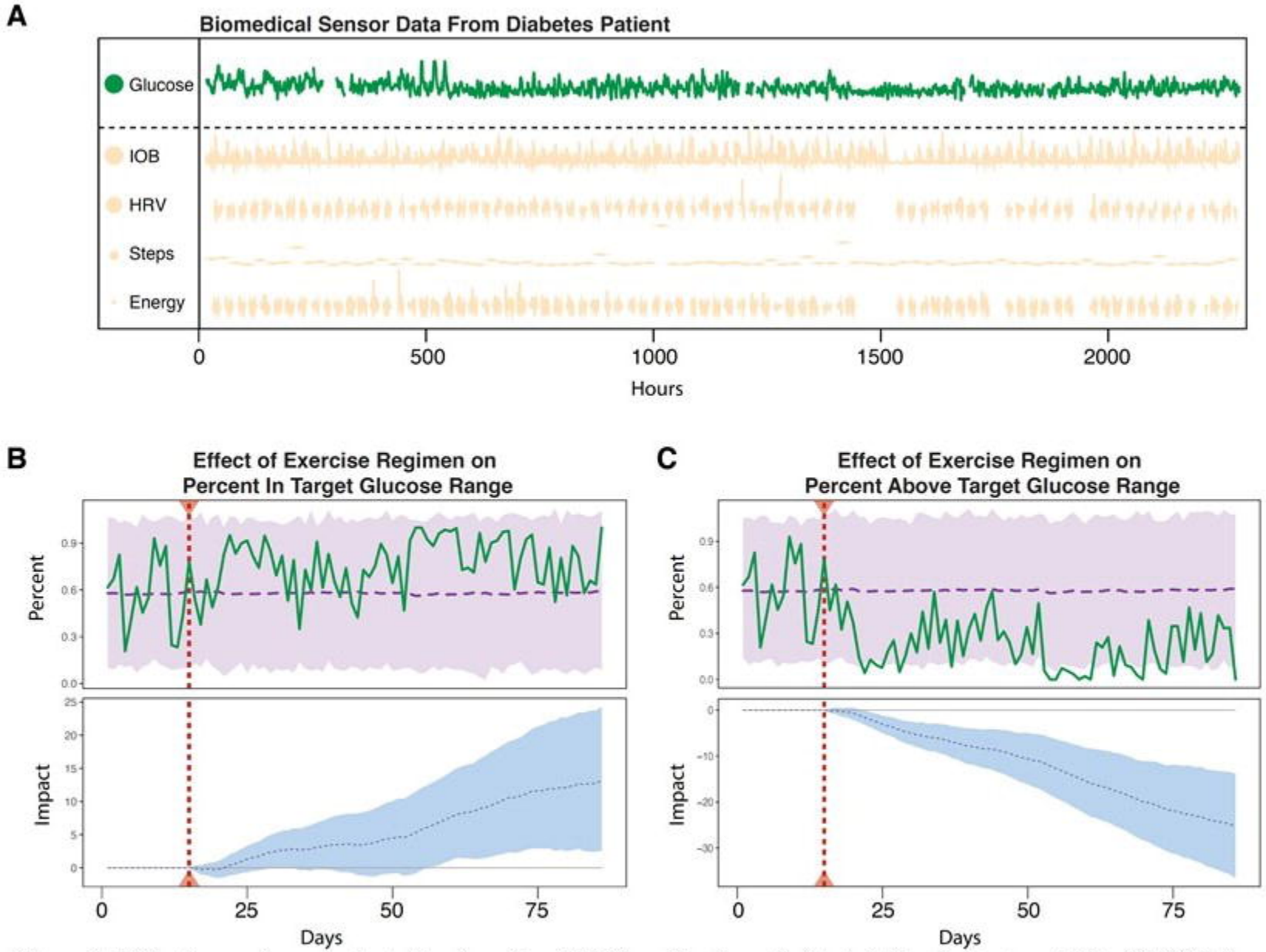

Figure 4. A) Continuous glucose, estimated insulin on board (IOB), and Apple watch data, including heart rate variability (HRV), daily steps taken, and energy expended, over 12 weeks for the participant of interest. B) Analysis of a 10-week exercise regimen's causal impact on the percentage of daily glucose readings in the target range (percent-in-target). C) Analysis of a 10-week exercise regimen's causal impact on the percentage of daily glucose readings above the target range (percent-above-target). HRV, heart rate variability. IOB, insulin on board.

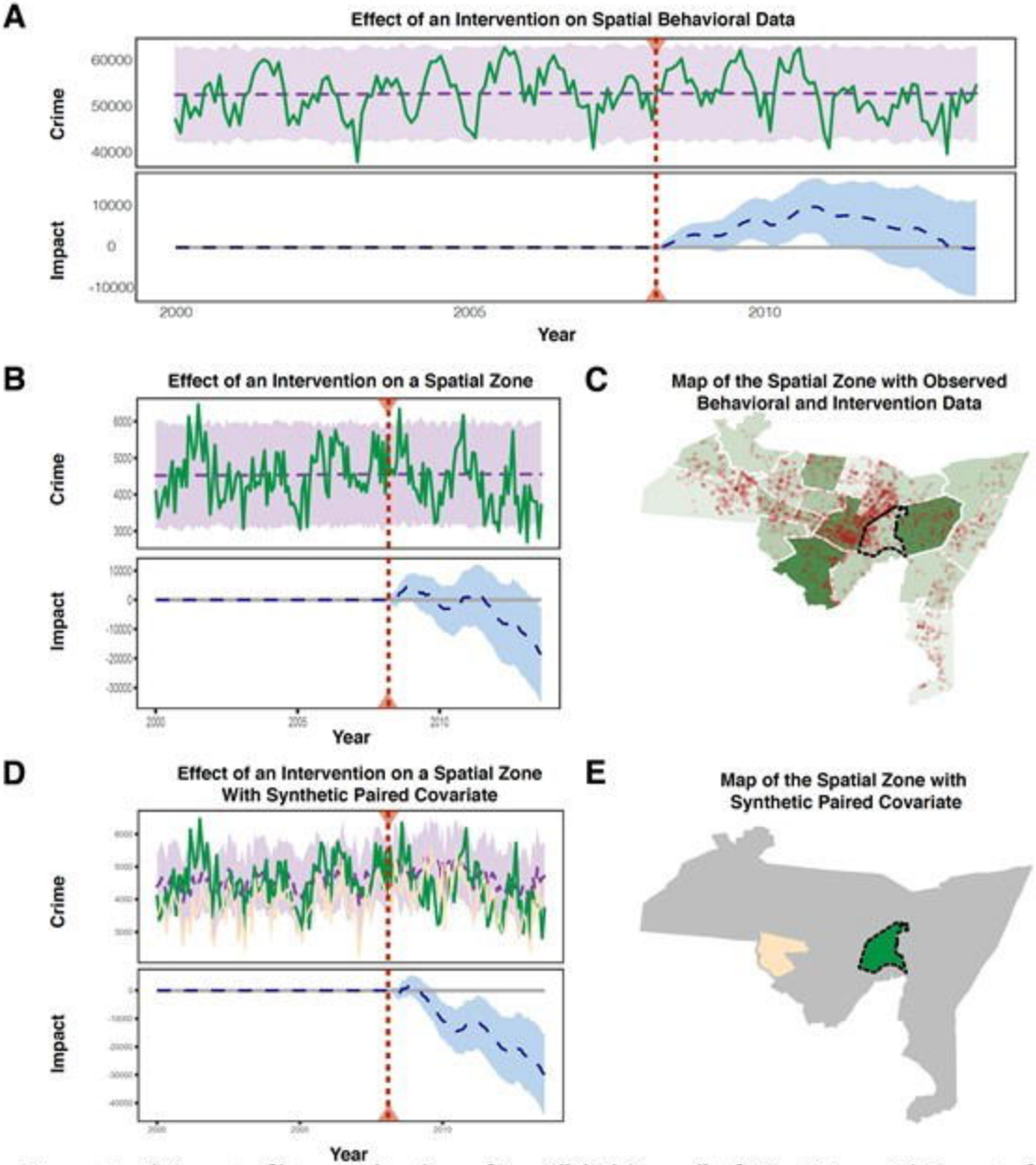

Figure 5. A) Impact of intervention (use of SeeClickFix) on all of New Haven. B) Impact of intervention on only one neighborhood, Wooster Square. C) Spatial data of New Haven showing variety of SeeClickFix usage throughout various neighborhoods. D) Impact of intervention on Wooster Square crime using West River as a paired covariate. E) Illustration showing the observed data, Wooster Square, versus the paired covariate, West River.