RESEARCH ARTICLE

Bayesian structural time series for biomedical sensor data: A flexible modeling framework for evaluating interventions

Jason Liu[1,2‡], Daniel J. Spakowicz[3,4‡], Garrett I. Ash[5,6], Rebecca Hoyd[3], Rohan Ahluwalia[1,2], Andrew Zhang[1,2], Shaoke Lou[1,2], Donghoon Lee[7,8], Jing Zhang[9], Carolyn Presley[3], Ann Greene[10], Matthew Stults-Kolehmainen[11,12], Laura M. Nally[10], Julien S. Baker[13,14], Lisa M. Fucito[15,16,17], Stuart A. Weinzimer[10,18], Andrew V. Papachristos[19], Mark Gerstein[1,2,20,21] *

1 Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, United States of America, 2 Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, United States of America, 3 Division of Medical Oncology, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio, United States of America, 4 Department of Biomedical Informatics, The Ohio State University College of Medicine, Columbus, Ohio, United States of America, 5 Veterans Affairs Connecticut Healthcare System, West Haven, Connecticut, United States of America, 6 Center for Medical Informatics, Yale School of Medicine, New Haven, Connecticut, United States of America, 7 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, 8 Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, 9 Department of Computer Science, University of California, Irvine, California, United States of America, 10 Department of Pediatrics, Yale School of Medicine, New Haven, Connecticut, United States of America, 11 Digestive Health Multispecialty Clinic, Yale-New Haven Hospital, New Haven, Connecticut, United States of America, 12 Department of Biobehavioral Sciences, Teachers College, Columbia University, New York, New York, United States of America, 13 Faculty of Sports Science, Ningbo University, China, 14 Centre for Health and Exercise Science Research, Department of Sport, Physical Education and Health, Hong Kong Baptist University, Kowloon Tong, Hong Kong, 15 Department of Psychiatry, Yale School of Medicine, New Haven, Connecticut, United States of America, 16 Yale Cancer Center, Yale School of Medicine, New Haven, Connecticut, United States of America, 17 Smilow Cancer Hospital at Yale-New Haven, New Haven, Connecticut, United States of America, 18 Yale School of Nursing, West Haven, Connecticut, United States of America, 19 Department of Sociology, Northwestern University, Chicago, Illinois, United States of America, 20 Department of Computer Science, Yale University, New Haven, Connecticut, United States of America, 21 Department of Statistics & Data Science, Yale University, New Haven, Connecticut, United States of America

‡ These authors are co-first authors on this work.
* mark@gersteinlab.org

Data Availability Statement: The data can be found at https://github.com/gersteinlab/mhealthCI.

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, interpreting biomedical sensor data needs 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. This framework corrects for covariates to provide accurate assessments of the significance of interventions.

Competing interests: The authors have declared that no competing interests exist.

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, MhealthCI, around a specific implementation of the Bayesian structural time series framework that uniformly processes, prepares, and registers diverse biomedical data. We apply the software implementation of MhealthCI 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 and also compare the performance to other models. Specifically, 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.

Author summary

In this paper, we propose and describe a robust and flexible modeling framework called MhealthCI based on the Bayesian structural time series, for which we have found to excel at analyzing diverse biosensor data. While Bayesian modeling is often employed in various fields such as finance, marketing, and weather forecasting, it is rarely used in biomedicine, specifically for biosensor and wearable data relating to human health and behavior. We use and apply this framework with the goal of interpreting and quantifying the causal impact of an intervention, a widespread goal of biomedicine. We describe the diversity of data types to which it could apply, provide intuition to its mechanics, collect relevant data in various fields, provide a wrapper tool around well-known R packages that prepares and registers diverse biosensor data to be analyzed, and finally apply the method to showcase its strength in quantifying the impact of interventions.

This is a *PLOS Computational Biology* Methods paper.

Introduction

Background

As mobile technology advances rapidly, the global mobile healthcare market is projected to be over 90 billion USD in 2022 [1]. Investment is bolstered by the great potential for advancing precision 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 interventions. A survey of distinct types of data are presented in Fig 1.

Example Human Activity Data

Example Human Behavioral Data

Example Environmental Sensor Data
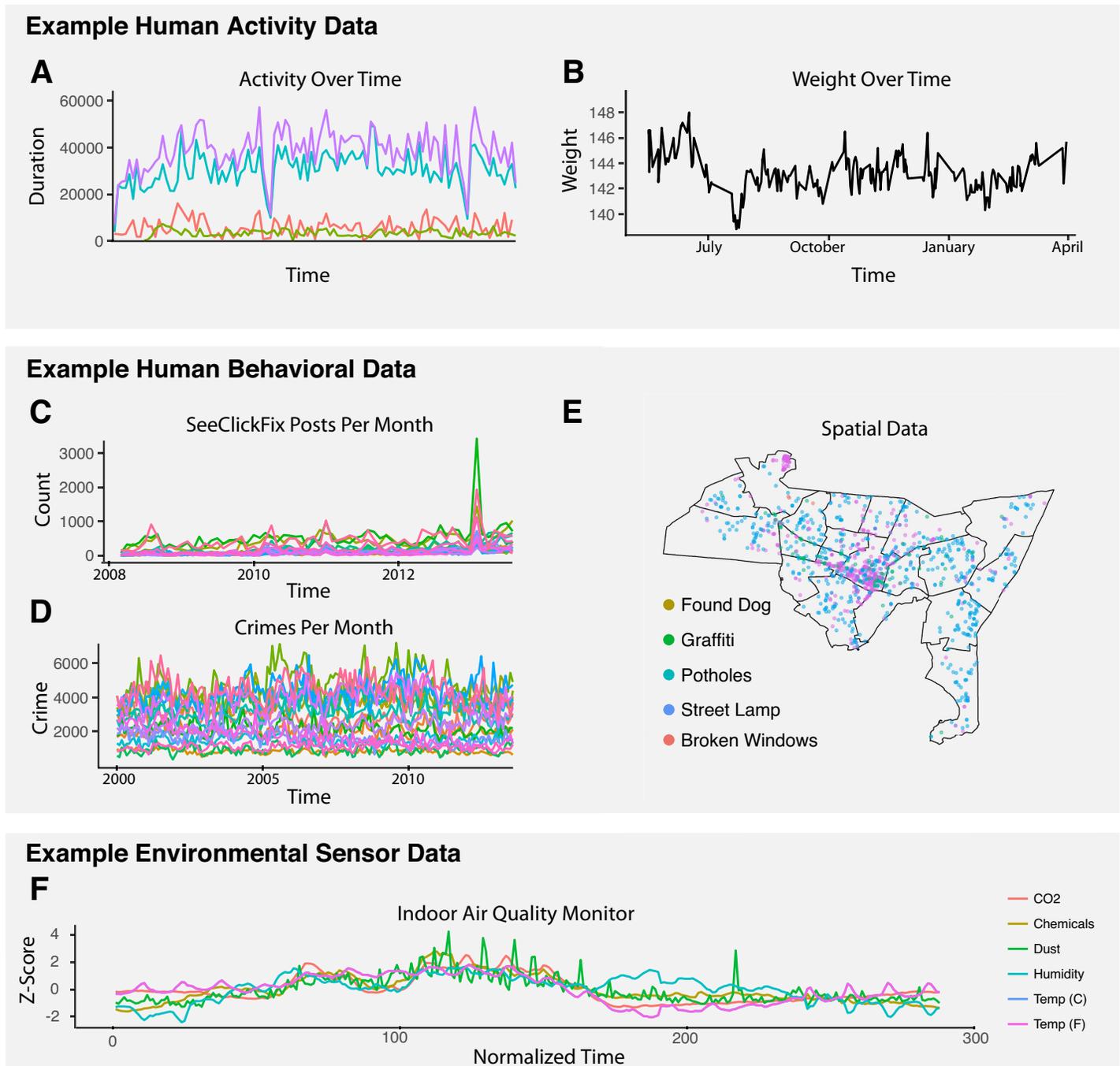
Fig 1. A) MOVES data set showing different exercise activities over time, with each color representing a different activity. B) Weight over time from the Withings data set. C) An example of human behavioral data, the use of SeeClickFix, over time. D) Various neighborhoods in New Haven and their crime data over time. E) Spatial map of New Haven illustrating the use of SeeClickFix and the associated geographical data. Open source map shapes can be found at *https://rdrr.io/github/CT-Data-Haven/cwi/man/neighborhood_shapes.html*. F) Various measurements taken from an indoor air quality sensor averaged over a day across a month.

https://doi.org/10.1371/journal.pcbi.1009303.g001

Though increasing amounts of sensor data is being released and publications highlighting their usefulness in personal health exist [8–10], there is still 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. In the simple case, t-tests can evaluate differences of averages before and after an intervention, but do not account for temporal aspects. Another type of method, auto regressive integrated moving average (ARIMA), is common for addressing longitudinally observed patterns that could inform the timing of an intervention but does not evaluate the impact and efficacy of such an intervention or provide dynamic p-values (more details in S1 Text). Specifically, the ARIMA model makes use of lagged forecast errors in order to evaluate the best fit and forecast future values, though it is generally not set up to forecast multiple, connected time points in the future. Other methods like cumulative sum control chart (CUSUM) are used to detect changes in time series data, though they do not evaluate the degree and length that a change may last, and are thus not suitable for evaluating long term impact of an intervention [11–13].

While the nature of data from sensors and wearables will vary depending on the context, most of them share certain properties, such as being densely sampled and longitudinal. A variety of such data can be seen in Fig 1. 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 [14]. The principles of this modeling framework stem from Bayesian inference and the analysis of time series data, which have been well established for decades [15]. While the idea of using Bayesian inference models has been extensively used in fields such as finance [16], ad campaigns [17], and marketing [18], they have been rarely used in the context of wearable and sensor data analysis to evaluate impact intervention over various time periods. A simple literature search for the application of Bayesian structural time series modeling of biosensor data yields few results. Therefore, here we apply this widely accepted statistical framework to the specialized context of biomedicine. We showcase the effectiveness of detecting the strength and duration of an intervention when applied to a variety of sensor data we collected and accurately assess the impact that various interventions have on individuals.

Modeling 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 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, such as a person's weight. This response variable could be a direct measurement taken from a sensor, or it could be a derived value–calculated from various underlying variables measured by a device–such as calories burned [19] or a "Nike Fuel Score" [20]. 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 [21], temperature [22], or diet [23].

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 its duration. There exist methods that detect intervention times [24,25] though to jointly assess the impact and time of an intervention would result in a significant disadvantage in statistical power. 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 pre-diet eating following several weeks of dieting) which is not addressed by our present model.

Our goal is to model the pre-intervention period including covariates to most accurately assess the behavior of 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 Fig 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 and include a graphical schematic in Fig 2.
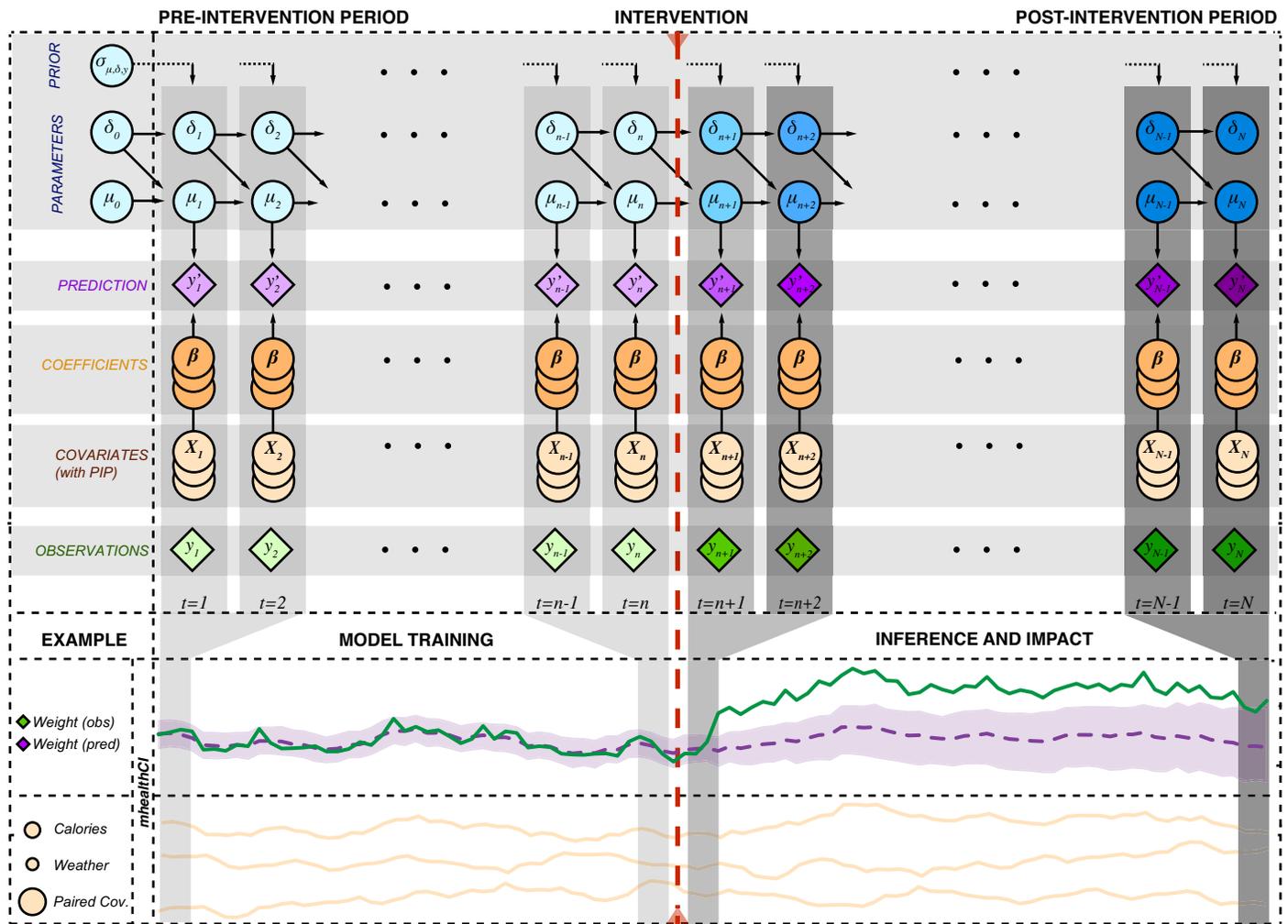
Fig 2. Schematic and illustration of the Bayesian structural time series. Latent parameters and hyperparameters are shown in blue, observations are shown in green, covariates in yellow (with a property of posterior inclusion probability), coefficients in orange, and predictions in purple. Predictions have an associated credible 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.

We assume for a given time point, t, there exists an observation y_t, linked to a variety of other parameters, $a_t : \{\mu_t, \delta_t, \boldsymbol{x}_t, \eta_t^y, \eta_t^\mu, \eta_t^\delta\}$. Here, a_t represents all of the state parameters at time t. The parameters are defined below. As time progresses from t to $t+1$, the other parameters also progress due to their time dependency.

We use the following equations to represent our time series data [14,26].

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

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

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

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 $N(0, \sigma_t^2)$. Specifically, the parameters $\{\sigma_\mu, \sigma_\delta, \sigma_y\}$ correspond to the standard deviations of the distribution that each of $\{\eta_t^\mu, \eta_t^\delta, \eta_t^y\}$ are drawn from, respectively. In general, the $\{\sigma_\mu, \sigma_\delta, \sigma_y\}$ variables are initially sampled based on the prior distribution, usually a $\mathrm{Gamma}\left(\frac{v}{2}, \frac{s}{2}\right)$.

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 number of covariates 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 [27]. The PIP gives a ranking measure to show how favorable the inclusion of a particular covariate is. The variable μ_t, representing the local trend of the model, 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. 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 Figs A-E in S1 Text.

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 $t = 1,\ldots,n$, while the post-intervention period is defined as $t = n+1,\ldots,N$. Furthermore, we define observations as y and predictions of the model as y'. Therefore, the set of observations in the pre-intervention period, $y_{1,\ldots,n}$, serve as the training data, based on all covariates. At each time point $t = 1,\ldots,n$, we update our parameter set of the model to better fit $y_{1,\ldots,n}$ 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, $y'_{n+1,\ldots,N}$. This is done by using the parameter set defined in the pre-intervention period and the covariates from the post-intervention period. In particular, because a Bayesian approach coupled with MCMC is used to estimate the state parameter distributions, predicted values in the post-intervention are estimated based on distributions of parameters. This enables the

counterfactual predictions to have reliable and accurate credible intervals, unlike many other time series methods. Additionally, because the counterfactual predictions are described as a distribution, these credible intervals can be propagated for the whole post-intervention period (which gives rise to the cone shape credible interval). Specifically, the counterfactual output value itself is the mean of the distribution and the credible interval is determined from the bounds (i.e. 95%) of the counterfactual distribution. Due to the stochastic nature of MCMC, small variations between runs of the framework may arise, though this can be mitigated with more iterations. Given $y'_{n+1,...,N}$ has an associated credible interval, we can calculate a significance p-value associated with the difference between $y'_{n+1,...,N}$ and $y_{n+1,...,N}$. 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 ($t = n+1,...,N$) is known as the *cumulative impact*. It should be noted that the credible interval associated with $y'_{n+1,...,N}$ 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, MhealthCI, around a specific Google implementation of the Causal Impact model, which makes use of the bsts R package. Altogether, our wrapper 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 $t = \{1,...,N\}$, there exists a set of variables (observation and covariates), y_t and 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 [14,28].

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 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. Second, we apply the Bayesian structural time series framework on data collected from an Android phone sensor in order to give intuition for–and demonstrate several key features of–our model. In this case, a known intervention is implemented, and allows for model comparison to ARIMA and model performance analysis. 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 quantify 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 and believe it is the first report of such a finding using data of this type. While these data are informative, they currently lack several attributes that we expect future biomedical data to have. In particular, we anticipate 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.

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 (Fig 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 [29], 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). Fig 3A shows the CO_2 measurements 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.

Wearable sensor timeseries with a controlled, defined intervention

To evaluate the performance of the BSTS modeling framework we designed an experiment with an intervention of defined size. In addition, we used this framework to demonstrate key features of the flexible covariate structure that set the BSTS framework apart from more commonly used methods. We used the Google Science Journal app on an Android phone and collected data for a person spinning and then extending the phone to arm's length while maintaining the same spin rate (intervention = change in centripetal force, see schematic in Fig 3B–3D). In the simplest case a single user's longitudinal data was used to estimate the effect of the intervention (Fig 3B). The model successfully detected the state change associated with the intervention, as shown by the cumulative change relative to the predicted values. For comparison, we also modeled these data using the more commonly used Autoregressive Integrated Moving Average (ARIMA) framework, which did not confidently detect an invention (p-value was greater than 0.05) due to larger confidence intervals (Fig 3B, left-middle panel). To further demonstrate the performance difference between the BSTS and ARIMA models, the size of the intervention was computationally altered (Fig 3B, middle panel). By scaling the effect size to 10 times larger than was experimentally generated (the equivalent of holding the sensor 30 ft away from the body while maintaining a constant rate of spin), we were able to compare the performance of two methods and show that our modeling framework confidently detected the intervention in half of the models when the effect size matched the experimental condition, while ARIMA required a 3x increase in the effect size to reach the same confidence.
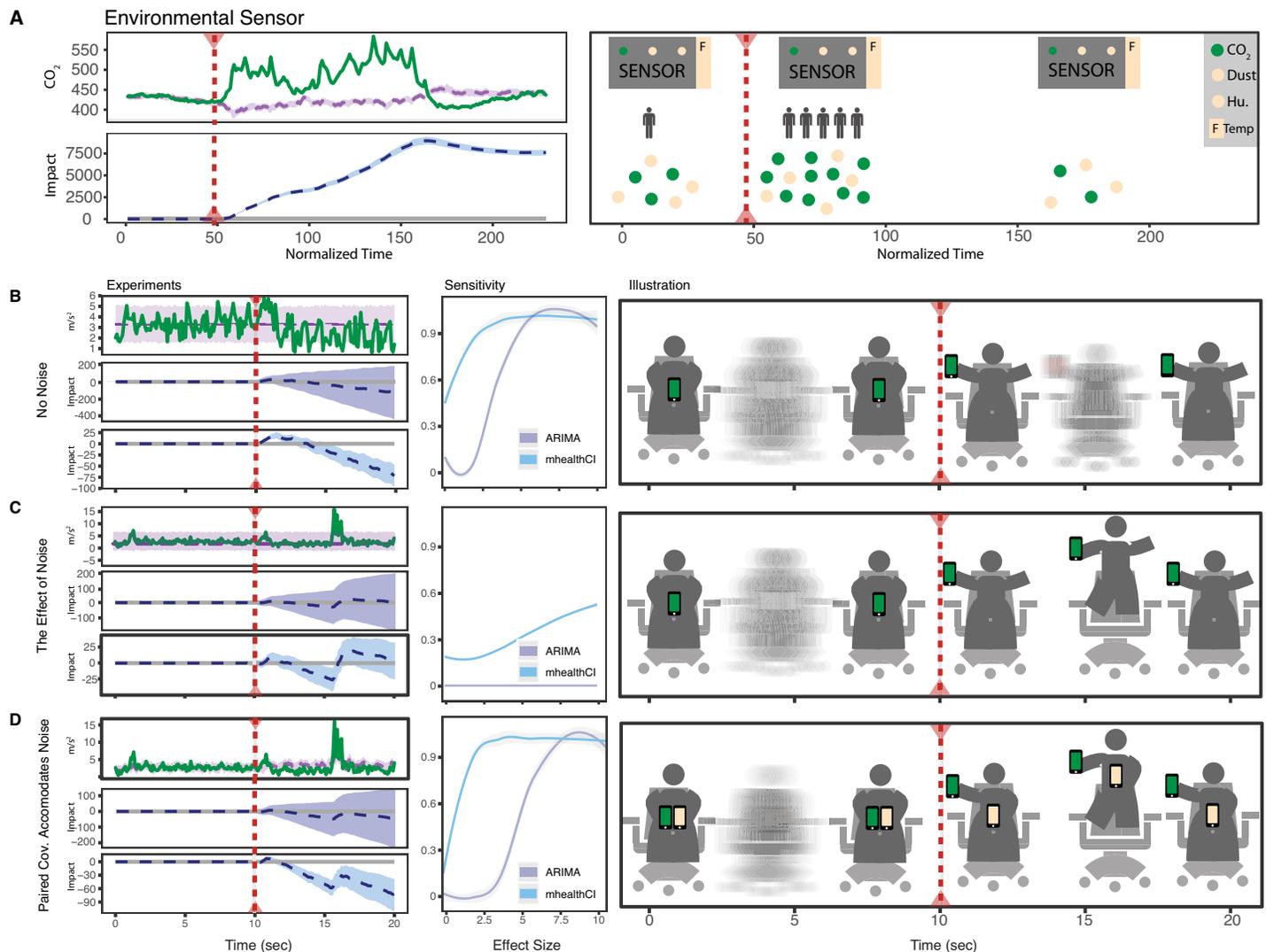
Fig 3. Performance of the Bayesian structural time series model in model experiments with known interventions. A) Using an indoor air quality sensor, CO_2 is measured with a variety of other covariates. The intervention (arrival of office members) causes an increase in CO_2 and is determined to be impactful using the model. B) Accelerometer measurements of a person spinning in a chair holding sensor near body and then extended to arm's length at the intervention; C) with simulated "noise" produced by a hop during the intervention period and D) with a paired covariate (second sensor) that is not affected by the intervention but experiences the "noise" hop. Sensitivity analysis shows comparison of mhealthCI to ARIMA, with the vertical axis as the fraction of the intervention that correctly identified a non-zero impact.

https://doi.org/10.1371/journal.pcbi.1009303.g003

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 (Fig 3C). 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. Neither BSTS, nor ARIMA confidently identified that an intervention occurred in this condition (Fig 3C, middle panel).

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 (Fig 3D). In this case two sensors were held while spinning, but only one was subjected to the

intervention (extended away from the chest). Both, however, experience the noise (hop), which the model is then able to control. Here again the BSTS model correctly identifies the intervention similar to the case without noise (Fig 3B). The difference in the performance of the ARIMA model is greater even than in the simplest case, with a five times difference in effect size needed to detect an intervention with the same confidence as the BSTS model.

A real-world analogy comparable to this simple experiment could be sleep data (accelerometer displacement) and the effect of changing one's pillow. Having no other information than one individual's sleep data is equivalent to the first experiment (Fig 3B). Next, the "noise" that we showed with a hop 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, the paired covariate that we demonstrated with a second phone 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 that uses a different, relevant data stream. For example, a non-paired covariate could be a device that measures 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). Both paired and non-paired covariates are effective, 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.

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 [30,31]. People with type 1 diabetes are advised to intensively monitor and 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.

Fig 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 [32]. In general, values above the target range are predictive of long-term organ damage, while values below the target range lead to acute hypoglycemia [33], 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 [34]).

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 (intensity and duration), **2)** individual characteristics (endogenous insulin sensitivity, and the effect of physical fitness to increase insulin sensitivity), and **3)**
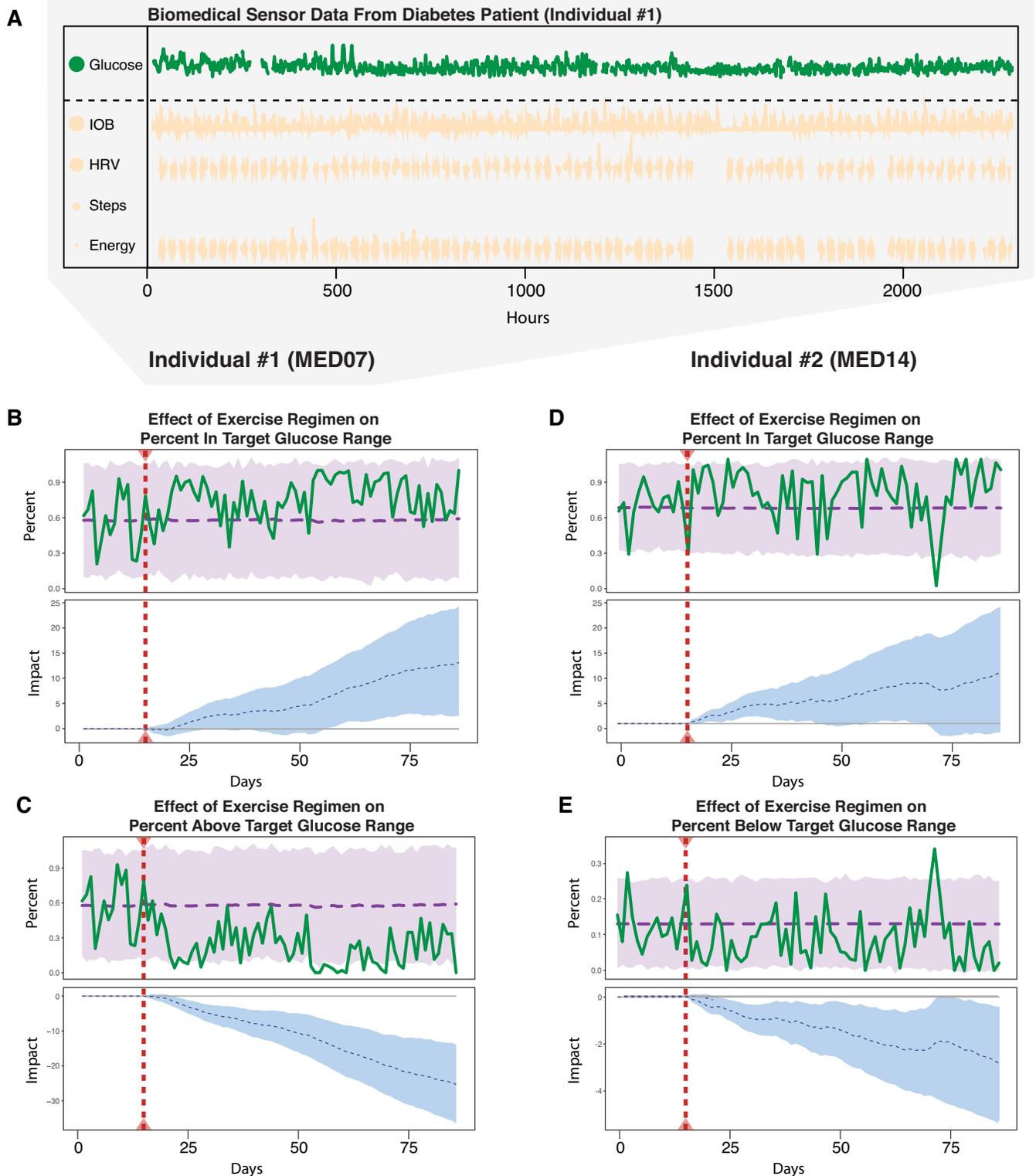
Fig 4. A) Continuous glucose, estimated insulin on board (IOB), and Apple watch data, including heart rate variability (HRV), daily steps taken, and energy expended, over 12 weeks for the participant of interest. B) Analysis of a 10-week exercise regimen's causal impact on the percentage of daily glucose readings in the target range (percent-in-target) of individual 1. 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) for individual 1. HRV, heart rate variability. IOB, insulin on board. D) Percent-in-target range analysis for individual 2. E) Percent-below-target range analysis for individual 2.

https://doi.org/10.1371/journal.pcbi.1009303.g004

contextual factors (pre-exercise blood glucose level, insulin- and carbohydrates-on-board, and concentration of counter regulatory hormones) [30].

We applied our Bayesian structural time series framework to two different participants. Participant #1 entered the study with excessive blood glucose time above-target indicated by baseline glycosylated hemoglobin (HbA1c) 8.6% compared to recommended goal 7.0% [35]. Her HbA1c at the end of the exercise program reached the goal (6.9%), consistent with the lowering effect that exercise has on blood glucose. Participant #2 entered the study with extremely little blood glucose time above-target (baseline HbA1c 5.2%) which did not change by the end of the exercise program (HbA1c 5.3%). Our developed method accurately predicted both of these situations. Specifically, for participant #1, we found 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 Fig 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. For participant 2, we found that the exercise regimen did not coincide with increased percent-in-target, which is supported by our model in Fig 4D and 4E ($p = 0.076$ and 0.323 respectively). We also compared the model to a simple aggregation style test, which our model outperformed (see S1 Text). To further check the validity of our results, we also performed a negative control analysis to showcase no significant impact was detected when no intervention occurred (see S1 Text). Thus, our model accurately reflected that there was minimal or insignificant difference in this biological variable with the introduction of exercise. Furthermore, of all the covariates we used, insulin on board (IOB) was found to have the highest PIP. This is reasonable since insulin is essential to glucose control. In summary, the Bayesian structural time series framework was effective in determining the change for both of the participants analyzed.

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–as we expect in the future. 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 information 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 [36] and can be linked to many other features such as census data of household incomes or education [37] (analogous to covariates of movement such as age, physical fitness, etc. [38]). 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, and promoting collective efficacy [39].

Following the intervention (creation of SeeClickFix) there was no detectable decrease in crime across the entire area (Fig 5A), nor in particular neighborhoods (Fig 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 SeeClick-Fix posts by neighborhood (Fig 5C) and then modeled the crime using a neighborhood that was not affected by the intervention as a control (i.e. a neighborhood without SeeClickFix posts), but did experience any city-wide initiatives that might affect crime (Fig 5D). Neighborhoods with heavy SeeClickFix use showed an effect of the intervention on crime when controlling for unobserved factors with synthetic paired covariates (Fig 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 also demonstrated better performance as compared to other commonly used time series methods such as ARIMA. We successfully evaluated the impact of interventions in a variety of examples and additionally showcase how a rich dataset with complex covariates can benefit from the framework.

While some studies demonstrating the success of generalized linear models for mobile health data do exist [24,25,40–42], there is a lack of emphasis on considerations for the temporal aspect of interventions and the effect-size of interventions at each time point in the post-intervention period. Specifically, generalized linear modeling frameworks lack the flexibility to evaluate the intervention's impact cumulatively in the post-intervention period. In this study, we illustrate the benefit of using a Bayesian structural time series framework for 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 leverage such data for precision medicine and health in the future, we must understand how specific timings and types of interventions impact individual patients [2–4].

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
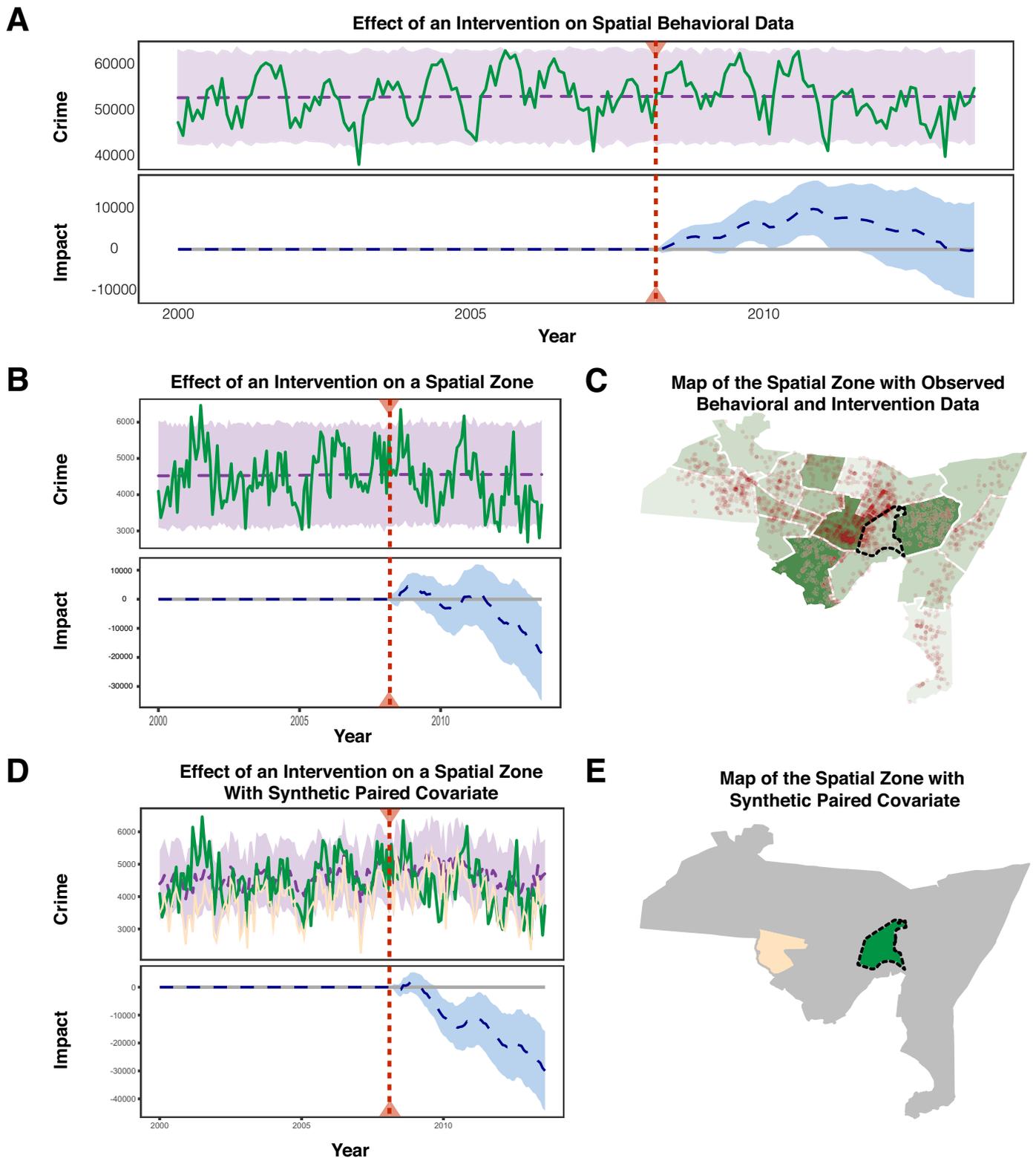
Fig 5. A) Impact of intervention (use of SeeClickFix) on all of New Haven. B) Impact of intervention on only one neighborhood, Wooster Square. C) Spatial data of New Haven showing variety of SeeClickFix usage throughout various neighborhoods. D) Impact of intervention on Wooster Square crime using West River as a paired covariate. E) Illustration showing the observed data, Wooster Square, versus the paired covariate, West River. All open source map shapes can be found at *https://rdrr. io/github/CT-Data-Haven/cwi/man/neighborhood_shapes.html*.

https://doi.org/10.1371/journal.pcbi.1009303.g005

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 [43]. However, these prior reports utilized HbA1c measured every 3–6 months as a chronic indicator of blood glucose control. We are the first study to our knowledge measuring changes in continuously measured blood glucose over several months, thus allowing a Bayesian time series approach that permitted analysis and adjustment for covariates at the individual participant level. We identified respective individuals whose blood glucose responded favorably and neutrally to the exercise intervention, after accounting for the blood glucose confounders of insulin dosing, physical activity outside of structured exercise, and autonomic nervous system function. This information would be useful feedback for the individuals and their healthcare providers since it can be otherwise challenging to predict whether exercise will increase or decrease blood glucose in type 1 diabetes.

It is also important to consider the contexts in which this modeling framework does not yield significant advantage over other frameworks. For example, 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.

Additionally, 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, and the strength of the results are based on such an assumption. In some cases, a synthetic paired covariate may be difficult to find. 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.

Overall, 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.

Google science journal data collection. 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.

AWAIR collection. 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 two participants 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 (NCT04204733) [44]. Both were delivered through a customized mobile digital application and supported by a coach internally certified in exercise for diabetes (GlucoseZone[TM], Fit-script[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%) [45]. 2). Insulin was delivered according to each participant's usual prescribed therapy. The participants 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 [46]. 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%) [47] and heart rate variability (standard deviation of interbeat intervals, validated against electrocardiography with intraclass correlation coefficient 0.98) [48] 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%) [49]. 5) Diabetes control was measured by glycosylated hemoglobin. Participant #1 used the DCA Vantage Analyzer (Bayer, Tarrytown, NY) at baseline and PTS Diagnostics A1cNow+ (Indianapolis, IN) at 10 weeks. Participant #2 used the AccuBase A1c Home Test Kit (DTI Laboratories, Thomasville, GA).

Computational Method. A computational method was developed in order to accurately determine the response to the exercise study. There were three phases to the algorithm: 1) align the imported data to the today study time period 2) match apple watch, insulin and glucose data to five-minute intervals 3) causal impact analysis. This alignment is required to test impact from the covariates. Following alignment, the casual impact analysis took covariates, target range, intervention period, and other features to predict the total time in range, above range, and below range. (more details in S1 Text).

Clinical Outcomes. Participant #1 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). Participant #2 was a 53-year-old white non-Hispanic female with type 1 diabetes for 35 years, receiving 22 units/day of insulin (0.3 units/kg body weight/day) and performing no regular exercise at baseline. During the 10-week intervention she received 19 units/day of insulin (0.2 units/kg body weight/day) and exercised on average 4.3 days per week, 40 minutes per session, at moderate to hard intensity (Borg rating of perceived exertion 4.0 / 10). The exercise routines were dynamic, interval-based, and equally emphasized all major muscle groups.

Processing and analysis of all data was done in R and Python and our packaged tool can be found at https://github.com/gersteinlab/mhealthci.

Supporting information

S1 Text. Supplemental information text file with additional details for manuscript. Fig A. Generated Data. **Fig B.** Using a linear model with no covariates to model the generated data. **Fig C.** Using a linear model with covariates to model the generated data. **Fig D.** Using the Bayesian structural time series model on the generated data. **Fig E.** Using the Bayesian structural time series with no covariates. **Fig F.** Spatial information from human behavior data set. **Fig G.** Determining causal impact on behavior data with no paired covariates. **Fig H.** Determining causal impact on behavior data with a paired covariate. **Fig I.** Flowchart and workflow for modeling. **Fig J.** Correlation Matrices For Daily Transformations of Variables Against Daily Percent-Glucose-In-Target-Range And Each Other. **Fig K.** Control for Diabetes dataset. **Fig L.** Effect of length of pre-intervention on statistical assessment of impact. (DOCX)

Author Contributions

Conceptualization: Jason Liu, Daniel J. Spakowicz, Mark Gerstein.

Data curation: Jason Liu, Daniel J. Spakowicz, Garrett I. Ash, Andrew V. Papachristos.

Formal analysis: Jason Liu, Daniel J. Spakowicz, Rebecca Hoyd, Rohan Ahluwalia, Andrew Zhang, Mark Gerstein.

Funding acquisition: Daniel J. Spakowicz, Garrett I. Ash, Mark Gerstein.

Investigation: Jason Liu, Daniel J. Spakowicz, Mark Gerstein.

Methodology: Jason Liu, Daniel J. Spakowicz, Rohan Ahluwalia, Shaoke Lou, Donghoon Lee, Jing Zhang, Mark Gerstein.

Project administration: Jason Liu, Daniel J. Spakowicz, Matthew Stults-Kolehmainen, Laura M. Nally, Julien S. Baker, Lisa M. Fucito, Stuart A. Weinzimer, Mark Gerstein.

Resources: Jason Liu, Daniel J. Spakowicz, Mark Gerstein.

Software: Jason Liu, Daniel J. Spakowicz, Rebecca Hoyd, Rohan Ahluwalia.

Supervision: Jason Liu, Daniel J. Spakowicz, Mark Gerstein.

Validation: Jason Liu, Daniel J. Spakowicz, Garrett I. Ash, Rebecca Hoyd, Rohan Ahluwalia, Mark Gerstein.

Visualization: Jason Liu, Rebecca Hoyd, Andrew Zhang.

Writing – original draft: Jason Liu, Daniel J. Spakowicz, Garrett I. Ash, Rohan Ahluwalia, Mark Gerstein.

Writing – review & editing: Jason Liu, Daniel J. Spakowicz, Garrett I. Ash, Rebecca Hoyd, Rohan Ahluwalia, Shaoke Lou, Donghoon Lee, Jing Zhang, Carolyn Presley, Ann Greene, Matthew Stults-Kolehmainen, Laura M. Nally, Julien S. Baker, Lisa M. Fucito, Stuart A. Weinzimer, Andrew V. Papachristos, Mark Gerstein.

References

1. Lee SM, Lee D. Healthcare wearable devices: An analysis of key factors for continuous use intention. Service Business 2020; 12/01; 14(4):503–531.

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. https://doi.org/10.1186/s12916-019-1366-x PMID: 31311528

3. Sim I. Mobile devices and health. N Engl J Med 2019; 381:956–968. https://doi.org/10.1056/NEJMra1806949 PMID: 31483966

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. https://doi.org/10.2196/10291 PMID: 30309834

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. https://doi.org/10.1016/j.outlook.2019.05.011 PMID: 31248630

6. Fitbit's 100+ Billion Hours of Resting Heart Rate User Data Reveals Resting Heart Rate Decreases After Age 40. Business Wire. 2018. Available from: https://www.businesswire.com/news/home/20180214005548/en/Fitbit%E2%80%99s-100-Billion-Hours-of-Resting-Heart-Rate-%C2%A0User-Data%C2%A0Reveals-Resting-Heart-Rate-Decreases-After-Age-40.

7. Seshadri DR, Li RT, Voos JE, Rowbottom JR, Alfres CM, Zorman CA, et al. Wearable sensors for monitoring the internal and external workload of the athlete. NPJ Digit Med 2019; 2:71. https://doi.org/10.1038/s41746-019-0149-2 PMID: 31372506

8. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza Rose SM, et al. Digital health: Tracking physiomes and activity using wearable biosensors reveals useful health-related information. PLoS Biol 2017; Jan 12; 15(1):e2001402. https://doi.org/10.1371/journal.pbio.2001402 PMID: 28081144

9. Li J, Li X, Zhang S, Snyder M. Gene-environment interaction in the era of precision medicine. Cell 2019; Mar 21; 177(1):38–44. https://doi.org/10.1016/j.cell.2019.03.004 PMID: 30901546

10. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. Per Med 2018; Sep; 15(5):429–448. https://doi.org/10.2217/pme-2018-0044 PMID: 30259801

11. Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. Stat Methods Med Res 2003 Mar; 12(2):147–170. https://doi.org/10.1177/096228020301200205 PMID: 12665208

12. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 2020 Dec; 4(12):1208–1220. https://doi.org/10.1038/s41551-020-00640-6 PMID: 33208926

13. Page ES. Continuous Inspection Schemes. Biometrika 1954 Jun; 41(1):100–115.

14. 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.

15. Fienberg SE. When did Bayesian inference become "Bayesian"? Bayesian Anal 2006; 1:1–40.

16. 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. https://doi.org/10.1093/oxfordhb/9780199559084.013.0010

17. Jin Y, Wang Y, Yunting S, Chan D, Koehler J. Bayesian methods for media mix modeling with carryover and shape effects. 2017. Available from: https://research.google/pubs/pub46001.pdf

18. Rossi P, Allenby G. Bayesian statistics and marketing. Marketing Science 2003; 22:304–328.

19. 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. https://doi.org/10.1097/00005768-199805000-00021 PMID: 9588623

20. What is Nikefuel? Nike. 2014. Available from: https://news.nike.com/news/what-is-nikefuel.

21. Schoeller DA. The effect of holiday weight gain on body weight. Physiol Behav 2014; Jul; 134:66–69. https://doi.org/10.1016/j.physbeh.2014.03.018 PMID: 24662697

22. 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. https://doi.org/10.1086/698372 PMID: 30205024

23. Geddes R, Harvey JD, Wills PR. The molecular size and shape of liver glycogen. Biochem J 1977; May 1; 163(2):201–209. https://doi.org/10.1042/bj1630201 PMID: 869923

24. 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. https://doi.org/10.1016/j.addbeh.2017.10.026 PMID: 29121530

25. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, Hekler EB, Murphy SA. 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. https://doi.org/10.1093/abm/kay067 PMID: 30192907

26. Scott SL, Varian HR. Predicting the present with Bayesian structural time series. SSRN. 2013. Available from: http://dx.doi.org/10.2139/ssrn.2304426

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

28. Scott SL. bsts: Bayesian Structural Time Series. R package version 0.9.5 [Internet]. 2020. Available: https://CRAN.R-project.org/package=bsts.

29. 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.

30. 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. https://doi.org/10.3389/fendo.2019.00326 PMID: 31258513

31. 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. https://doi.org/10.1016/S2213-8587(17)30014-1 PMID: 28126459

32. 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. https://doi.org/10.2337/dci19-0028 PMID: 31177185

33. 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. https://doi.org/10.1056/NEJM199309303291401 PMID: 8366922

34. 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. https://doi.org/10.2337/diacare.10.5.617 PMID: 3677982

35. American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in Diabetes-2020. Diabetes Care 2020 Jan; 43(Suppl 1):S66–S76. https://doi.org/10.2337/dc20-S006 PMID: 31862749

36. 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.

37. 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.

38. 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. https://doi.org/10.1186/s40798-018-0173-9 PMID: 30617718

39. Mergel I. Distributed democracy: SeeClickFix.com for crowdsourced issue reporting. 2012;1/27/2012;1. Available from: https://ssrn.com/abstract=1992968

40. 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. https://doi.org/10.1093/ntr/ntx201 PMID: 29059349

41. 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. https://doi.org/10.1111/biom.12792 PMID: 29023626

42. 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. https://doi.org/10.1093/ntr/ntx225 PMID: 29059413

43. Wu N, Bredin SSD, Guan Y, Dickinson K, Kim DD, Chua Z, Kaufman K, 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):https://doi.org/10.3390/jcm8020253 PMID: 30781593

44. Ash GI, Nally LM, Stults-Kolehmainen M, De Los Santos M, Jeon S, Brandt C, et al. Personalized big data for type 1 diabetes exercise support. SportRxiv: 34vdc [Preprint]. 2021 [cited 2021 May 21]. Available from: https://doi.org/10.31236/osf.io/34vdc

45. 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. https://doi.org/10.1177/1932296818813581 PMID: 30453772

46. 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. https://doi.org/10.1056/NEJMoa1907863 PMID: 31618560

47. 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. https://doi.org/10.2196/mhealth.7043 PMID: 28302596

48. 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. https://doi.org/10.3390/s18082619 PMID: 30103376

49. 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. https://doi.org/10.3390/jpm7020003 PMID: 28538708