Quantifying the Causal Effect of Individual Mobility on Health Status in Urban Space

YUNKE ZHANG, Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University
FENGLI XU*, Knowledge Lab & Mansueto Institute for Urban Innovation, University of Chicago
TONG XIA, Department of Computer Science and Technology, University of Cambridge
YONG LI, Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University

How does individual mobility in the urban environment impact their health status? Previous works have explored the correlation between human mobility behaviour and individual health, yet the study on the underlying causal effect is woefully inadequate. However, the correlation analysis can sometimes be bewildering because of the confounding effects. For example, older people visit park more often but have worse health status than younger people. The common associations with age will lead to a counter-intuitive negative correlation between park visits and health status. Obtaining causal effects from confounded observations remains a challenge. In this paper, we construct a causal framework based on propensity score matching on multi-level treatment to eliminate the bias brought by confounding effects and estimate the total treatment effects of mobility behaviours on health status. We demonstrate that the matching procedure approximates a de-confounded randomized experiment where confounding variables are balanced substantially. The analysis on the directions of estimated causal effects reveals that fewer neighbouring tobacco shops and frequent visits to sports facilities are related with higher risk in health status, which differs from their correlation directions. Physical mobility behaviours and environment features have more significant estimated effects on health status than contextual mobility behaviours. Moreover, we embed our causal analysis framework in health prediction models to filter out features with superficial correlation but insignificant effects that might lead to over-fitting. This strategy achieves better model robustness with more features filtered out than L1-regularization. Our findings shed light on individual healthy lifestyle and mobility-related health policymaking.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**; • **Information systems →** *Data mining*.

Additional Key Words and Phrases: Urban mobility; Health status; Causal inference

ACM Reference Format:
Yunke Zhang, Fengli Xu, Tong Xia, and Yong Li. 2021. Quantifying the Causal Effect of Individual Mobility on Health Status in Urban Space. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 193 (December 2021), 30 pages. https://doi.org/10.1145/3494990

*This is the corresponding author: fenglixu@uchicago.edu

Authors' addresses: Yunke Zhang, zyk21@mails.tsinghua.edu.cn, Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing; Fengli Xu, Knowledge Lab & Mansueto Institute for Urban Innovation, University of Chicago, Chicago; Tong Xia, Department of Computer Science and Technology, University of Cambridge, Cambridge; Yong Li, Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing.

2474-9567/2021/12-ART193 $15.00
https://doi.org/10.1145/3494990

1 INTRODUCTION

As most countries around the globe witnessed rising life expectancy and ageing populations [47], people are increasingly conscious about maintaining healthy lifestyles, which are often associated with various mobility-related exercises and activities, such as having more physical movements and visiting places with healthy context [2, 10, 65]. To which extent is one's health condition determined by his/her mobility behaviours? It has been a long-standing research problem to understand the causal relationship between mobility behaviour and health status. A number of studies have shown that the physical and contextual characteristics of mobility behaviour are correlated with various indicators of health status, such as depressive moods [11], stress level [36], endocrine disorder [44] and all-cause health conditions [93]. Therefore, it is a paramount task to uncover the causal effects of mobility behaviour on health status, which might have important implications on wide-ranging applications and policy-making.

However, the conclusions drawn from correlation analyses could be not reliable and sometimes even contradictory. For instance, when we examine the association between park visit frequency and health status, we might find a misleading negative correlation because they are both affected by an individual's age [22]. To be specific, elder people tend to visit parks more frequently compared to the youngers, while the elders also tend to have higher risks of health conditions. Therefore, even if the correlation between health status and park visit frequency is positive within each age group, the overall correlation analysis will likely generate a negative outcome. Another example is one in the current work. When we calculate the association between the number of sports venues an individual has visited and her behaviour of hospital visit, the outcome supposes that more sports facilities visit is related to higher health risk, which is opposed to the intuitive characteristic of sports venues. A deeper probe into the relationship between age and sports visits reveals that seniors significantly visit more sports facilities than younger people. This suggests that the correlation of visits on sports and health status may be partially affected by their association with age and requires further consideration. The aforementioned counter-intuitive correlations are caused by the well-known confounding effect in experiment design [56], which needs to be adjusted with techniques in the area of causal inference. Specifically, when there are common causes that simultaneously affect two variables, their correlation is confounded by the common causes and thus is inappropriate to be interpreted as the causal effect. Therefore, directly applying the potentially biased conclusions of correlation analysis to health monitoring could cause critical consequences. Motivated by this challenge, we aim to remove the confounding bias in the correlation study and estimate the causal effect of individual mobility behaviour on health status.

In this paper, we launch the study on the relationship between urban mobility behaviour and health status from a causal perspective. We alleviate the confounding effect with a causal analysis framework that mainly consists of three components. First, we leverage a large-scale mobility dataset passively collected from mobile networks, which records individual's movements and stays in urban spaces with high spatial and temporal granularity [93]. In addition, the context of mobility behaviour is supplemented by a POI (Point of Interest) dataset. Each urban visit can be characterized by the distributions of various categories of neighbouring POIs. Moreover, the dataset is associated with the ground truth of all-cause health status, which is collected via a user survey. These datasets provide a unique chance to estimate the causal effects of mobility behaviour on health status. Second, beyond simple correlation analysis, we estimate the causal effects of each mobility behaviour pattern on health status. We first construct a causal diagram to depict causal relationships and confounding effects based on findings from previous works. Then we design a propensity score-based matching procedure that matches individuals with similar confounding variables to approximate a randomized experiment, relieving the confounding effect under our assumptions on causal relationships. The average treatment effects of mobility behaviour patterns on health status are calculated on the matched pairs of individuals. The significance and directions of estimated causal effects are further analyzed to understand mobility behaviour's relation with health status. An experiment of estimating mobility pattern's effect on hospital readmission is included to lend more credibility to the estimation of their effect on health status. Thirdly, we embed the causal analysis framework into health status prediction models. Since causal analysis provides each pattern's effect independently, the significance level of an input feature's estimated causal effects can serve as a guide for judging whether it is essential in the prediction or just superficial noise that might cause over-fitting. By selecting causally significant input features, we can achieve more robust models with better performance on prediting health status.

Through the above design-guided analysis and prediction, our main findings are fourfold. First, we discovered that some directions of estimated effects differ from confounded correlation directions after the confounding effect is relieved. For example, more neighbouring entertainment venues have effects on worse health outcome, while it is correlated with better health status. Visiting sports facilities may impact better health status, but the correlation analysis suggests an opposite association. The causal analysis provides more intuitive findings, which helps us to understand the role of mobility behaviour patterns in influencing health status. Second, the significance test on estimated causal effects reveals that the physical mobility behaviours and environment features have a more significant impact on health status than contextual mobility behaviours. Meanwhile, visiting sports and entertainment venues has a higher estimated impact on health than visiting other categories of POIs. Third, the causal framework can balance the distributions of confounding variables on matched pairs. We demonstrate that after the matching process, the differences between average values of confounding variables on an individual with a higher treatment value in a pair and an individual with a lower treatment value are reduced from over 70% to no more than 10%. This balancing property ensures that the matching is a good approximation of randomized experiments, validating the credibility of the estimations. Fourth, our causal-based feature selection method significantly improves the prediction performances of the health status prediction task based on a naive Bayes model($p<0.01$). Furthermore, causal-based feature selection in logistic regression achieves the same level of performance with L1 regularization with 65% more features filtered. The strategy's efficiency demonstrates that causal analysis can avoid misleading features included in prediction models, consequently improving the model's robustness and interpretability.

The current work's major contributions are summarized as follow.

- We are the first to analyze the causal effects of mobility behaviour patterns derived from large, passively collected mobility data set on general health status. It confirms the deficiency of the correlation analysis and opens the opportunity to conduct cost-efficiency, large-scale, and comprehensive studies on the effect of mobility on health status.
- We alleviate the confounding effects in the correlation analysis on mobility behaviour patterns and individual health status by a causal framework. New insights in understanding mobility patterns' impact on individual health are drawn by analyzing the significance and the directions of estimated causal effects.
- We leverage the results of causal analysis for feature selection in health prediction models. Experiment results demonstrate that the prediction performances are improved by 0.1% to 8% after filtering insignificant input features by causal-based feature selection.

After discussing some related works in Section 2, we introduce the mobility dataset and explain the motivation to conduct causal analysis in Section 3. Section 4 describes the approach to estimate causal effects. Section 5 provides the analysis and credibility checks on estimated causal effects and its application in the health prediction task. Implications and limitations are discussed in Section 6.

2 RELATED WORKS

2.1 Mobility Data Mining

Studies on mobility data have grown rapidly in recent years due to the abundant opportunity to access passively sensed large-scale mobility data owing to the widespread use of smartphones and cellular networks [58]. The studies can be categorized into modelling the physical attributes of mobility and the application of mobility data in urban problems. Classic models of human mobility include Lévy flight [8], gravity model [39], and radiation model [70]. González *et al.* [28] discovered that individual urban travel shows a high degree of spatial and temporal regularity. These models efficiently advance the prediction of urban mobility. Besides, mobility data are applied in various problems. On an individual level, Xu *et al.* [87] captured semantic features of living style by temporal modes detected in mobility trajectories. The semantic and physical attributes of urban mobility are jointly modelled to predict an individual's demographic [86]. On a regional level, Yuan *et al.* [89] use urban human mobility patterns to infer regional functions. Regional physical inactivity is studied by mobility data in [2]. The successful application of mobility data in various topics strengthens our motivation to study the relationship between mobility behaviour and individual health status.

2.2 Health Analysis with Mobility Data

Both mobility patterns and health indicators can be sensed passively without human interaction. This is consistent with that a significant proportion of research interests falls in health analysis among the various applications of mobility data. Relationships between mobility patterns and a diverse range of health indicators are studied. The mobility behaviours are shown to be indicative of individual health conditions by conducting prediction tasks in [44, 93]. Hillebrand *et al.* [33] combine mobile network data with app usage in inferring well-being status. Contextual mobility behaviour is exploited in [82] to predict chronic disease. As for mental health analysis, Canzian *et al.* [11] investigate the correlation between various mobility behaviours and an individual's degree of depression. Morshed *et al.* [52] collect passively sensed mobility and activity data to predict mood instabilities reliably. Mobility indicators of stress-resilience are identified and their association with mental health are studied in [1]. Mobility trajectories and POI visits help in analyzing social anxiety [36]. In addition, hospital readmission is evaluated and predicted by patient's mobility behaviours, such as sedentary behaviour [4] and step counts [75]. A common issue of these studies is that they are based on correlation analysis which may be inadequate to represent how mobility patterns impact health status because of the confounding effects in correlation studies. A potential improvement is to alleviate the confounding effects by methods in the field of causal inference. We aim to adopt these measures into studying the relationship between mobility and individual health status.

2.3 Causal Inference in Health Analysis

Causal inference has been long studied in various subjects. The goal of causal inference, different from the association, is to understand the causal relation and infer the beliefs when conditions are changed or intervened [56]. When we observed the occurrence of two phenomenons, association analysis focuses on how are they related, while causal analysis imagines what would happen if one phenomenon had not occurred. Current studies on studying the causal effect of a treatment on an outcome mostly follow two equivalent causal models - the Neyman-Rubin potential-outcome framework [63, 72] and Pearl's structural causal model [54]. Under the potential outcome framework, the potential outcome denotes what the outcome would be if an individual were to take a treatment. The causal effect of a treatment is then defined as the difference of the potential outcomes of taking and not taking the treatment. However, it is always impossible to observe all potential outcomes for a given individual [63]. This leads to the requirement of estimating counterfactual outcomes. Various approaches have been designed to deal with this problem and estimate causal effects, including grouped conditional outcome modelling [43], propensity score matching [61], inverse probability weighting [35], double machine learning [14], and causal forest [80].

Among the literature that is concerned with causal relationships, a great deal of effort has been devoted to the public health community. Understanding the causal effect of a health-related association may indicate the potential effectiveness of the intervention [27]. Currently, with the development in machine learning tools for causal inference [79] and the access to abundance data, observational causal studies have sprung up in the public health literature, employing data from various sources. Hasthanasombat *et al.* [31] collect open drug prescription and map data to inquire into the causal effects of neighbourhood built environment on public mental health on a district population level. In [45], the authors evaluate disease management's effect on congestive heart failure by constructing comparable control groups with propensity scores. Psychiatric drug's effect on individual psychopathology is studied with the self-reported medication drawn from social media data [64]. Previous works successfully reduce the confounding effects in observational studies by causal approaches to study the effect of various sources of potential causes on health indicators. The current paper is instructed by similar thoughts of de-confound correlation to understand mobility pattern's effect on health status using passively sensed data. Compared with the literature on causal-boosted health analysis, we mainly focus on the effect on general health status, which is represented by an individual's hospital visit behaviour. We do not concentrate on a single factor's effect on health outcome, instead, we construct a confounding structure with different mobility patterns and estimate each of their effects on health status. To sum up, we are the first to illustrate an approach to understand the confounding effects in the correlation between various mobility patterns and individual health using passively collected mobility record data.

2.4 Propensity Score Matching

The causal inference framework in the current research is based on the procedure of propensity score matching, which is widely applied in observational causal inference studies in different fields, including economics [17], pedagogy [88], political science [77], and medical science. It serves as a strong technique to alleviate systematic bias in causal analysis based on observational data, especially when randomization trials are unethical or uneconomical. When evaluating the effect of one treatment variable on an outcome in observational studies, there might have covariates that affect both the treatment and the outcome, which negates the direct comparison between treatment and control group. By mapping the high-dimensional covariates to a scalar propensity score and matching units with almost identical scores, it is able to construct comparable pairs of units that have same the balanced covariates and only differ in the level of treatment.

Benefit from its strength in estimating treatment effect, propensity score matching is widely employed in the medical literature. The above-mentioned causal studies in Section 2.3 [31, 45, 64] all utilize propensity score matching-based methods. Although the subjects of these studies differ, researchers similarly identify covariates that affect both treatment and outcome to estimate propensity scores for the matching method. Considering the high-dimensional covariate in the current study, we choose to estimate causal effects based on propensity score matching. Meanwhile, this matching method also receives critiques [3, 40]. King and Nielsen [40] appraise that matching based on propensity score can yield imbalanced unit pairs that aggravate the bias, which requires researchers' prudent check on covariate balance. Following their instructions, we carefully probe into the degree of covariate balance after our matching procedure to ensure the credibility of our causal analysis.

3 DATASET AND PROBLEM STATEMENT

In this section, we first briefly introduce the mobility dataset we use in our research. Then we illustrate how the dataset is processed to obtain mobility behaviour patterns. Finally, motivated by the deficiency of correlation analysis, we formulate the research problem that aims to answer and show the collected data provide a unique angle on studying the causal effect of urban mobility on health.

3.1 Data Overview

The research dataset is collected from a healthy survey conducted in 2017 in Beijing. 1,056 outpatients chosen from 13 major hospitals filled in a medical experience survey and provided permission to collect their demographic of

Table 1. The basic information of the mobility dataset.

City	Collection Period	Sample Size	Average Time Length	Average Records per Day
Beijing	Jul. 5th ~Aug. 31st, 2017	2112	48.67 days	65.81 records

Table 2. The distributions of user demographics.

Demographic	Healthy Persons	Outpatients
gender	female(807, 76.4%)	female(549, 62.0%)
	male(249, 23.6%)	male(507, 38.0%)
age	0~30(360, 34.1%)	0~30(191, 18.1%)
	30~50(583, 55.2%)	30~50(505, 47.8%)
	50~99(113, 10.7%)	50~99(360, 34.1%)
income level	low(317, 30.0%)	low(299, 28.3%)
	medium(409, 38.7%)	medium(319, 30.2%)
	high(330, 31.3%)	high(438, 41.5%)

age, gender, income level and phone number. All outpatients are visiting a hospital to seek medical treatment. 1,056 randomly sample persons who report a good health condition supply the same information. The phone numbers are desensitized before being provided to the researchers. Mobility records ranging from July 5th to August 31st are gathered from the cellular network according to the phone number. Mobility records contain geographic coordinates, arrival times, and stay times for users' every visit across the metropolitan area of Beijing. The basic information of the mobility dataset is listed in Table 1. The mobility dataset provides sufficient mobility length and records for studying mobility behaviour's impact on health status. The distributions of user demographics on outpatients and healthy persons are listed in Table 2. We can observe that most of the users are female, and half of the users are aged from 30 to 50 years old. The outpatients have a higher proportion of elder people over 50 and people with the highest income level. In subsequent causal analysis, the demographics will play an important role in estimating causal effects of mobility behaviours.

In addition to the mobility dataset, we select the Point of Interest(POI) dataset collected from BaiduMap to demonstrate the context of a geographic location. It covers the location, category, and sub-category information of more than a million POIs in the city of Beijing. Here we mainly focus on POIs with semantics related to health, including four categories - food, entertainment, scenic spot, and sports, and two sub-categories - fast food and tobacco/liquor shop.

For ethical considerations about the privacy of mobility data, the following protocols are imposed to eliminate the privacy and ethical risk during the data analysis. First, all users are informed of the research purpose of the mobility data and health status. They provide their phone number and authorize data collectors to gather their mobility data from the cellular network. Users' real phone numbers are properly anonymized by the collector before being shared with researchers. Second, all researchers sign firm non-disclosure agreements before being permitted to view or cope with the mobility data. The research proposal is approved by the local institutional board. Finally, the anonymized mobility data are stored in a secure offline server where only approved researchers can operate it.

3.2 Hospital Visits and Mobility Behaviour

To study the causal relationship between mobility behaviour and health status, we extract quantitative patterns to depict both mobility behaviour and health status. Since the ground truth of hospital visits is assured by the healthy survey, we use whether a user visits the hospital as the indicator of her overall health status. To satisfy that the cause happens before the effect, we cut off the mobility records after each outpatient's first hospital visit. Then, we referred to the literature on health-mobility relation to design patterns potentially affect health status that can be extracted from the available mobility record data. First, measurements of the range and activeness of urban mobility are possibly influencing health outcomes. The relationship between these physical characteristics of mobility and health outcomes has been widely studied in [4, 11, 42, 62]. Second, it is possible to infer the residential place of individuals from their mobility records [95], and numerous previous works have shown that the living environment is significantly correlated with public health [6, 7, 26, 31, 49, 81], which enlightens us to model individual's living environment with the POI dataset and take it into consideration. Given that the neighbourhood is influencing an individual's health status, nonresidential activities should also be taken into account since people travel in the city daily for various purposes. Previous studies have shown that visiting non-residential places is associated with health outcomes such as weight or BMI [38, 90]. Based on these references, the environment of places visited by individuals is included in the current study. Thus, in a short summary, we can contextualize the patterns related to health that can be derived from mobility records as the following three aspects, physical mobility behaviours, environment features, and contextual mobility behaviours.

3.2.1 Physical mobility behaviour patterns. include the radius of gyration, the standard deviation of displacements, and the distribution entropy of places visited, denoted as ρ, σ, ϵ respectively. The radius of gyration depicts the user's mobility range, where a higher ρ represents a higher mobility range. It is shown in [62] that reduced physical mobility is associated with a higher risk of depression. Displacements are the distances between two consecutive points on the mobility trace. σ represents the user's mobility regularity. Users with regular mobility patterns have lower σ's. In [11], the authors show a strong correlation between σ and mental health conditions. The distribution entropy of places visited ϵ characterizes the diversity of locations that the user has visited. We segment the city into 300 meters by 300 meters grids and for each user, we count the number of visits to each grid. ϵ is defined as 1 minus the sum of the squares of the visit frequencies of all grids. Therefore, users who mostly visit a few locations will have a small ϵ.

3.2.2 Environment features. depict the user's living environment by calculating the area of green space and counting the numbers of POIs within a certain range around the user's home. The location of the home is estimated from the mobility records by choosing the place where the user mostly stays in the nighttime. $N(green)$ denotes the area of green space in the 500 meters by 500 meters region centered by the user's home. It measures the accessibility to nature, which is regarded to be linked with health [6]. $N(food)$, $N(ent)$, $N(sport)$, $N(scene)$, $N(fast)$, and $N(tob)$ represents the number of POIs from six categories or sub-categories within 500 meters around user's residence place. The (sub-)categories are food, entertainment, sport, scenic spot, fast food, and tobacco/liquor shop, respectively. We select these (sub-)categories based on the following reasons. [49, 81] study the relationship between food venue, fast food store, and health indices. Entertainment venues display an important influence in health monitoring [93]. The effect of neighbouring sports venues on depression is examined in [31]. Scenic spots serve a similar role as green space [7]. The density of tobacco stores will affect life expectancy according to [26].

3.2.3 Contextual mobility behaviour patterns. represent the semantic of one's mobility behaviour by averaging the POI distribution over each location on a mobility trajectory. Previous works have pointed out that where we visit is associated with our health status besides our living environment [41, 78]. For instance, visiting green spaces or one's favourite sites are positively associated with mental health and vitality [78]. Since the exact visiting semantic of each point on mobility records cannot be obtained exactly from the mobility record, we consider the POI counts within 500 meters around each location on the mobility trajectory as the context of that visit. Then we calculate the weighted average of the contexts of all points on the mobility trajectory, where

Table 3. Summary of mobility behaviour patterns and the directions of their correlation with health status.

Physical Mobility Behaviour Patterns	Environment Features	Contextual Mobility Behaviour Patterns
$\rho(-), \sigma(-), \epsilon(-)$	$N(green), N(food)(-), N(ent)(-), N(sport)(-), N(scene)(-), N(fast)(-), N(tob)(-)$	$V(food)(+), V(ent)(+), V(sport)(+), V(scene), V(fast)(+), V(tob)(+)$

the weight of each location is the distance to the residence place as distant visits are more representative of one's subjective visit willingness. The same six categories in environment features are considered and counted in the above approach. The contextual mobility behaviour patterns are denoted as $V(food)$, $V(ent)$, $V(sport)$, $V(scene)$, $V(fast)$, and $V(tob)$. To summarize, the mobility behaviour patterns are listed in Table 3.

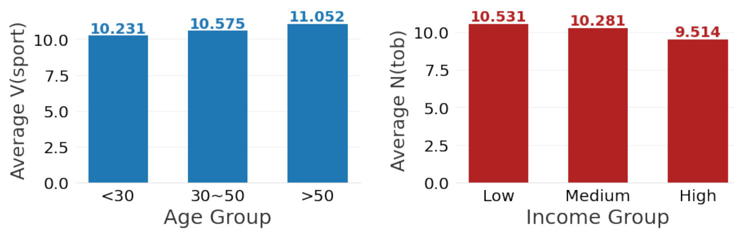
3.3 Motivation

The literature mainly adopts correlation analysis to study the association between mobility and health [11, 36, 44, 93]. It intuitively reveals the positive or negative direction of association. The Pearson correlation coefficients between binned-value of mobility behaviour patterns and health status and corresponding significance levels are listed in Table 4 and the directions of significant correlations are listed in Table 3. $N(green)$ and $V(scene)$ have insignificant correlations with health status, which suggest that they have no causal relationship. Therefore, we do not include these patterns in the causal estimation. Note that visiting a hospital is recorded as an outcome of 1. Thus, a positive coefficient implies that an increase in the value of mobility behaviour patterns is related to an increased probability of worse health status, vice versa.

However, the correlation analysis provides confusing results. For example, all environment features except the green space area in the neighbourhood have negative significant correlations with hospital visits, while all contextual mobility behaviours have positive correlations. Some of the "directions" contradict both our intuition and previous studies. For instance, tobacco store has been commonly regarded as negatively associated with health outcome [26], but more tobacco stores in the neighbourhood are associated with a lower risk in worse health in our analysis. Urban green spaces are considered to promote public health [84], however, we may interpret that living in a block with more green areas is health-harmful according to the correlation direction in Table 3.

These confusing correlations can be interpreted by confounding effects brought by *confounding variables*. They are features that potentially affect both the mobility behaviour pattern we are interested in and the individual health status. For example, individual demographics are confounding variables for mobility behaviours. Age significantly affects health status. At the same time, it affects the choice of living environment and the context of mobility. To intuitively explain this effect, we first calculate the average number of sport POIs on user's mobility trace in three user groups: users younger than 30, between 31 and 50, and over 50. The results are shown in Figure 1(a). We can observe that there are more sports venues around locations visited by older people on average. The average of the oldest group is significantly higher than the average of the youngest group ($p<0.05$). The positive correlation between age and $V(sport)$ implies that the positive correlation between $V(sport)$ and hospital visits is possibly produced by that age - the confounding variable is simultaneously correlated with $V(sport)$ and hospital visits positively.

Similar to the analysis above, we plot the average $N(tob)$(number of tobacco POIs in the neighbourhood) among different income groups in Figure 1(b). Wealthier people tend to live in an environment with fewer tobacco stores than people with the least income($p=0.096$). Since people with higher income are more likely to seek hospital treatment [91, 92, 94], this correlation can partially explain the negative correlation coefficient of $N(tob)$



(a) Age and $V(sport)$.

(b) Income and $N(tob)$.

Fig. 1. Correlation between mobility behaviours and demographics.

in Table 4. Above examples establish the flaw in simple correlation analysis. As a result, we are dedicated to assess the reliability of significant correlations by causal analysis in following sections.

3.4 Problem Statement

Since confounding effects bring bias to correlation analysis, we are motivated to quantify the de-biased impact of urban mobility behaviour on individual health status. We adopt the framework of causal inference and formally define the research problem as follows. Given the user's mobility behaviour patterns and whether the user visits the hospital after the collection period of the mobility trace, we aim to estimate the *causal effect* of each mobility behaviour pattern on individual health status, which is indicated by the possibility of a hospital visit. The estimated causal effect should disclose the direction and significance of mobility behaviour pattern's impact on health status.

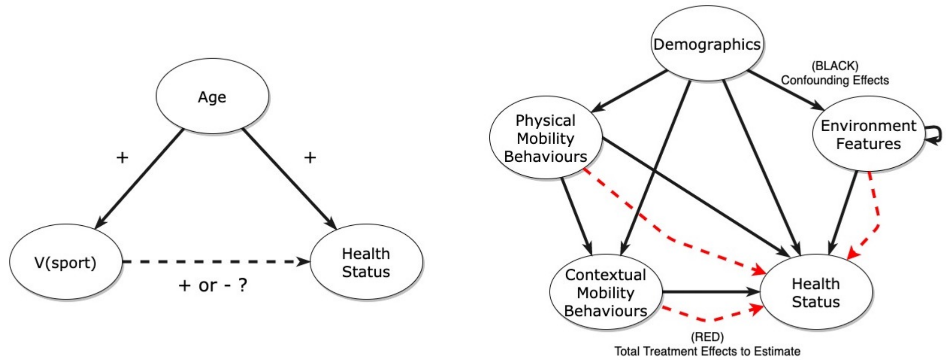
4 METHODS

In the above analysis, we demonstrate that the correlation analysis can hardly obtain the impact of the change of one specific mobility behaviour pattern on health because changes always simultaneously occur on various variables. Therefore, we adopt tools in causal inference scenarios to separate each pattern's *causal effect* on health status. In this section, we first briefly introduce several concepts in the causal inference that helps to express the problem from a causal perspective. Then we propose a series of approaches to estimate the causal effects of mobility behaviour patterns.

4.1 Basic Concepts

We are interested in studying the effect of specific mobility behaviour patterns on health status, which is denoted by hospital visits. In our causal analysis, we adopt basic concepts in Neyman-Rubin causal model [63, 72]. When we investigate the causal effect of a specific pattern, that pattern is called a *treatment*. The possibility of the individual visiting hospital is the *outcome*. From the perspective of causal inference with binary treatment, the *potential outcome* of treatment is what the outcome would be if we apply the treatment to a sample. The *treatment effect* is the difference between the potential outcome of applying and not applying the treatment. It answers the question of "what would the outcome be if we take or not take the treatment". This question is often regarded as *counterfactual* because we can always only observe only one outcome of applying or not applying the treatment but never both of them.

Unlike binary treatments, the mobility behaviour patterns in our data set are continuous or discrete values. The extension of binary treatments to multi-level treatments is discussed in various works [34, 37, 68]. When analyzing the causal effect of specific treatment, we will transfer the treatment into binned values which lowers



(a) The confounding effect of age on $V(sport)$ and health status.

(b) The constructed causal diagram.

Fig. 2. Illustrating the confounding effect and the causal diagram of mobility behaviour patterns and health status.

the number of treatment levels while preserving the treatment value's order. The transferred treatment level is also called a *dose*. The *potential outcome* under a specific dose is what the outcome would be if the treatment is applied with that dose. Further definition of the *treatment effect* or the *causal effect* of a treatment is the change of the potential outcome would be if we raise the dose with one level. Here we generally hypothesize a linear relation between potential outcome and dose level to avoid over-fitting that potentially brought by complex assumptions. In another word, whenever the dose raises one level, the change in the potential possibility of a hospital visit equals the *treatment effect*.

4.2 Confounding Variables

In Section 3.3, we establish that demographics such as age and income to some extent affect the correlation between mobility behaviour patterns and health status. A *confounding variable* in the causal model is the variables that have a causal effect on both the treatment and the outcome. Under a causal view, the analysis of the confounded correlation between visited sports venues and health status discussed in Section 3.3 can be illustrated by a causal diagram shown in Figure 2(a). Age, the confounding variable, has a positive causal effect on both visited sports venues and individual health status. This confounding effect negates the correlation analysis and we require a de-confounding method to estimate the causal effect of visited sports venues.

A critical prerequisite to estimate the unbiased treatment effect of mobility behaviour on health status is to determine the causal structure. Here we propose a possible causal structure, as shown in the black arrows in Figure 2(b). Our proposed causal relationships in the diagram are hypothesized with literature support. First, previous works reveal the association between demographics and individual health [16, 19, 60], which correspond with the black arrow pointing from demographics to health status. Meanwhile, previous works have demonstrated that age [12, 22], gender [12, 51], and income level [53] influence physical and contextual mobility behaviours. Thus, causal relations from demographics to contextual and physical mobility behaviours are added. Furthermore, the impact of demographics on living choice is studied in [24, 48, 67]. Therefore, we consider the demographics as common confounding variables for all mobility behaviours. Second, the spatial distributions of different categories of POIs are correlated [5]. This implies that the living environment features may have tangled effects. When estimating the treatment effect of specific environmental features, we treat all other environment features as confounding variables. Finally, physical mobility behaviours may exert an influence on contextual mobility behaviours. We hypothesize a causal relation flowing from physical behaviours to contextual behaviours. The causal relationships are summarized as the black arrows in Figure 2(b).

Given the causal relationships between variables, we want to derive each mobility pattern's *treatment effect* or *causal effect* on health status. These effects are represented as the red dotted lines in Figure 2(b). Since there are two paths from physical mobility behaviours to health status in the causal graph, the treatment effect of physical mobility behaviours on health status is also known as the *total treatment effect* according to Pearl [55]. The total treatment effect is the collective effect of multiple casual paths. Anyhow, it is identical with the treatment effect introduced in Section 4.1, representing the difference between potential outcomes when the treatment varies.

There are points on the selection of confounding variables and the decision of causal diagram requires further discussion. We discuss these topics in Section 6.2.2.

4.3 Matching and Propensity Score

Under the definition of the treatment effect, which is the difference between potential outcomes, an ideal approach to obtain the treatment effect is to do a randomized controlled trial(RCT), where we randomly apply a level of dose to a proportion of the population and apply another level of dose to another proportion and compare their outcomes. Due to the randomness of the assignment of treatment, the distributions of other variables are identical on the treatment and control group, which ensures the comparability between the two groups. However, this method is obviously quixotic based on our confounded observations. We have to seek a substitution for RCT that can simulate the identical distributions of confounding variables on both groups based on our observations.

A classical method for simulating an RCT is matching. In binary treatment cases, for each sample in the control group, select a sample in the treatment group with identical values for all confounding variables to match with it. Since the only difference is the value of treatment, these two samples can be considered as a counterfactual pair. Under this setting, the causal effect will be the average difference between matched samples. However, an identical match on confounding variables could be difficult or infeasible. To address this problem, the matching method based on propensity score is a widely-used substitute to achieve balanced confounding variables [61].

The propensity score projects the high-dimensional confounding variables to a numerical value and the matching is conducted based on this value. To ensure that the confounding variables are balanced on matched pair, the propensity score should be a balancing score $b(X)$, which is a function of confounding variables X that the conditional distribution of X given $b(X)$ is the same for each treatment level T,

$$X \perp T | b(X). \tag{1}$$

For binary treatment scenario, the propensity score is defined as the probability of taking the treatment given confounding variables $e(X) = P(T = 1|X)$, it is proved that the confounding variables can still be balanced by matching on the propensity score [61]. To extend the propensity score on multi-level treatments, we need to find a proper value that satisfies the balancing property in Equation 1.

There are various approaches that attempt to achieve the balancing property, including generalized propensity score [34] and generalized covariate balancing propensity score [25]. We refer to the method introduced in [31] and adopt a propensity score derived from the ordered logistic regression model [50]. Ordinal regression is designed to predict an ordinal variable, which satisfies the binned treatment level in our data. The ordinal regression model can be formulated as

$$P(T \leq d|X) = \sigma(\theta_d - \mathbf{w}^T X), \tag{2}$$

where d is the dose level, $\mathbf{w}$ and θ_d are parameters, and $\sigma(\cdot)$ is the sigmoid function. Note that the distribution of dose level given confounding variables only depends on $b(X) = \mathbf{w}^T X$. Thus, $P(T|X) = P(T|b(X))$. Furthermore, we can prove that

$$P(T, X|b(X)) = P(T|X, b(X))P(X|b(X)) = P(T|X)P(X|b(X)) = P(T|b(X))P(X|b(X)). \tag{3}$$

This equation proves the independence between the treatment level and confounding variables conditioning on $b(X)$, which corresponds with the balancing property in Equation 1. Therefore, we treat $b(X) = \mathbf{w}^T X$ as the propensity score of each sample with confounding variables X.

4.4 Estimating Causal Effect

With the formerly introduced propensity score, we are able to design a matching approach to estimate the causal effect of each mobility behaviour pattern. For each pattern, we first transferred it into a binned value with 6 levels according to their percentile. Then, we fit an ordinal regression model that estimates the treatment level by confounding variables. The selection of confounding variables is discussed in Section 4.2. We adjust the regularization term of the regression model to achieve the best prediction accuracy. The estimated propensity score for an individual with confounding variables X is given by $\hat{\mathbf{w}}^T X$, where $\hat{\mathbf{w}}$ is the model's fitted parameter.

Having obtained the estimated propensity score, we conduct matching on individuals to create balanced confounding variable distributions between matched individuals. The propensity score matching method for binary treatment minimizes the difference between the propensity scores of matched samples, and the matched pair should be selected from different groups. For treatment with multi-level doses, we follow the modified version of propensity score distance in [46]. The matched pairs are proposed to have large differences between treatment levels and small differences between estimated propensity scores. The distance $d_{i,j}$ between two individuals i and j is defined as

$$d_{i,j} = \begin{cases} \dfrac{|\hat{\mathbf{w}}^T X_i - \hat{\mathbf{w}}^T X_j|}{|T_i - T_j|}, & T_i \neq T_j. \\ \infty, & T_i = T_j. \end{cases} \tag{4}$$

In the binary treatment scenario, only samples from different groups are permitted to be matched. Thus, matching for binary treatment is an optimal bipartite graph matching problem. As for multi-level dose in our data, the matching is converted to an optimal weighted graph matching problem, where nodes represent individuals, edges are weighted by the distance $d_{i,j}$ between two connected nodes. We adopt the algorithm introduced by Edmonds [18] to find a matching with minimized weight.

Now that the optimal matching $M = \{(i, j)\}$ is obtained, where individual i and individual j are paired, the individual treatment effect per level of treatment for matched pair (i, j) is the difference between outcome Y's divided by difference between treatment level T's. The average treatment effect (ATE) per level of treatment can be estimated by the average of individual treatment effects

$$ATE = \frac{1}{|M|} \sum_{(i,j) \in M} \frac{Y_i - Y_j}{T_i - T_j}. \tag{5}$$

We calculate the ATE per level of treatment for each mobility behaviour pattern. To quantify the significance level of the treatment effect, we conduct a two-sided t-test [73] on the observed individual treatment effects to calculate the p-values. The results are shown in Table 4.

5 RESULTS

5.1 Credibility of Causal Estimation

Unlike prediction or classification tasks with given labels, a crucial problem in evaluating the estimated treatment effects is that we do not have their ground-truth value. Thus, we have to check the credibility of propensity score matching before making any conclusions. In this section, we follow King's guidance [40] of clarifying the balance of confounding variables before analyzing the estimated treatment effects.

Table 4. Comparison in correlation coefficients and estimated average treatment effects.
***: p<0.0001, **: p<0.01, *: p<0.05

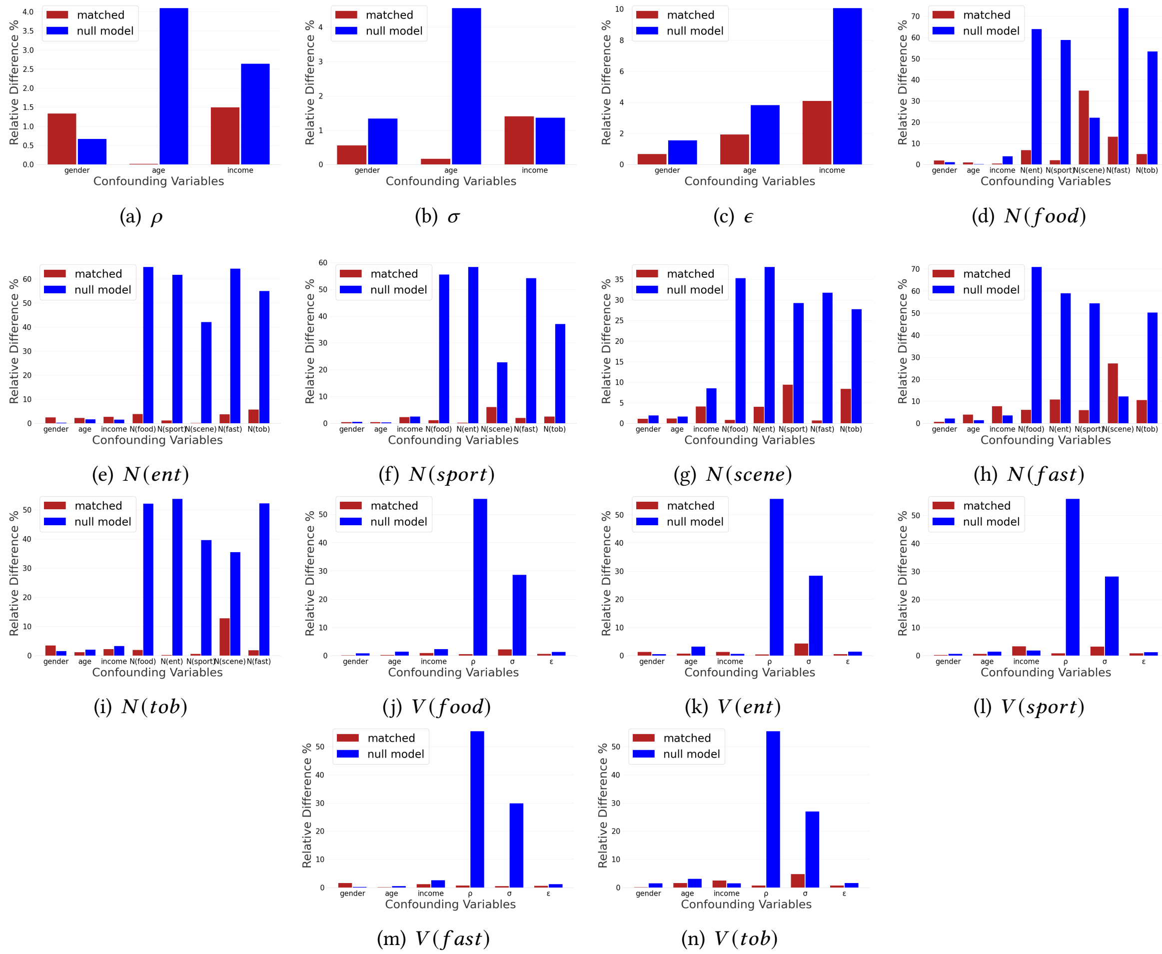
Category	Mobility Behaviour Pattern	Correlation Coefficient	Estimated Causal Effect
Physical Mobility Behaviour	ρ	-0.1669 ***	-0.02796 ***
	σ	-0.2041 ***	-0.04904 ***
	ϵ	-0.1335 ***	-0.02291 ***
Environment Features	N(green)	-0.0184	—
	N(food)	-0.0884 ***	-0.02537
	N(ent)	-0.0403 *	0.08778 ***
	N(sport)	-0.0899 ***	-0.02822 **
	N(scene)	-0.0948 ***	-0.02127 **
	N(fast)	-0.0945 ***	0.02440
	N(tob)	-0.0449 *	0.04660 ***
Contextual Mobility Behaviour	V(food)	0.0887 ***	-0.00166
	V(ent)	0.1126 ***	0.02623 **
	V(sport)	0.0815 ***	-0.02173 *
	V(scene)	0.0044	—
	V(fast)	0.0854 ***	0.00221
	V(tob)	0.1037 ***	0.00052

Unlike in binary treatment settings, where distributions of confounding variables on the controlled and treatment groups can be directly compared, we investigate their similarity on the high-dose group and the low-dose group. The high-dose group consists of samples with higher treatment levels in each matched pair and the low-dose group consists of samples with lower treatment levels. We compare the first-order moments of confounding variables in two groups. The relative differences between average values are shown in Figure 3. To establish the efficiency of the propensity score-based matching procedure, we also provide the relative differences given by a null model in Figure 3. The null model corresponds with naive correlation analysis. It compares the average values of confounding variables on two halves of the population - the half with a higher treatment level and the other half with a lower treatment level. The null model's relative differences are shown by blue columns, whereas the matched pairings' relative differences are represented by red columns. Note that $N(green)$ and $V(scene)$ have an insignificant correlation with health outcome, we no longer take them into consideration.

For most treatment variables, we can observe that matching has greatly improved the balance between high-dose and low-dose groups. Differences in confounding variables are dramatically reduced. Several differences seem to increase after the matching procedure, for instance, income as confounding variable of treatment ρ. However, t-test [83] shows that both the original difference and the difference after matching are insignificant. All relative differences are lower than 10% except for $N(food)$ and $N(fast)$. Both treatments do not achieve balance on the confounding variable of $N(scene)$, the relative differences are over 20%. Therefore, the estimated treatment effects of $N(food)$ and $N(fast)$ are incredible and omitted in the following discussions. For other mobility patterns, the reliability of the matching procedure is proved.

5.2 Analyzing Estimated Treatment Effects

The estimated ATEs represent the increase in the potential probability of hospital visits when the treatment level increase by one level. The balance check depicts that confounding effects in correlation analysis are relieved by the matching procedure. We can understand the direction and the extent of mobility behaviour pattern's



(a) ρ (b) σ (c) ϵ (d) $N(food)$

(e) $N(ent)$ (f) $N(sport)$ (g) $N(scene)$ (h) $N(fast)$

(i) $N(tob)$ (j) $V(food)$ (k) $V(ent)$ (l) $V(sport)$

(m) $V(fast)$ (n) $V(tob)$

Fig. 3. The relative differences of average confounding variables on high-dose and low-dose group for each treatment variable.

impact on health outcome under an unconfounded scenario from the average treatment effects. For mobility behaviour patterns with positive ATE, they are regarded as unhealthy that a higher pattern value may impact the individual having a higher health risk. On the other hand, ones with negative ATE are considered healthy - the risk of unhealthy outcome decreases as the pattern's value increases.

From another perspective, we can interpret the extent of mobility behaviour pattern's causal impacts on health status by evaluating the absolute value of the average treatment effect. We list the significance levels of ATEs in Table 4.

In the following sections, we analyze the extent and direction of mobility behaviour pattern's estimated causal effects by their category. The significance and directions of estimated causal effects provide new insights in understanding mobility behaviour's impact on health outcome.

5.2.1 Physical Mobility Behaviours. As shown in Table 4, all physical mobility behaviour patterns have significant negative ATEs, which is consistent with the correlation analysis. The radius of gyration ρ's negative ATE implies that a larger covering area is associated with better health status. This corresponds with former conclusions that lower mobility is associated with a higher risk of health problems [11, 93]. The standard deviation of displacement σ and the distribution entropy of location visited ϵ illustrate the regularity and diversity of individual mobility. The estimated negative effects suggest that leaving from our daily routine for some improvised visits once a while is probable to achieve a better health outcome. To summarize, the estimated causal effects of physical mobility behaviour patterns are consistent with the former study and can be leveraged in individual health management.

5.2.2 Environment Features. The associations between living environment and health are broadly studied in the literature [11, 26, 31, 49, 81]. According to Table 4, simple correlation analysis shows confusing results that the POI densities of six categories in the neighbourhood are all associated with better health status. The estimations of causal impacts suggest that $N(ent)$, $N(sport)$, $N(scene)$, and $N(tob)$ have significant estimated ATEs and the direction of $N(ent)$ and $N(tob)$'s effects is positive, which differs from the direction of correlation. Meanwhile, the confounding variables of $N(food)$ and $N(fast)$ are not balanced by the matching procedure. Their estimated ATEs are invalid.

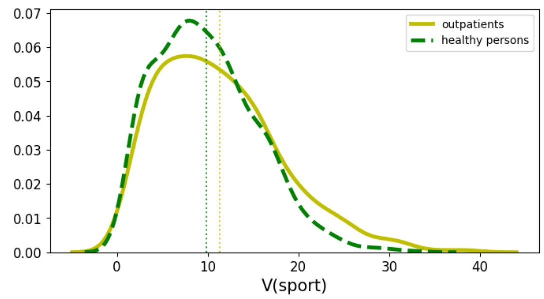
First, The significant negative treatment effects of sports and scenic POIs correspond with former studies [31, 69] and subjective impressions. A community with more accessibility to sports facilities or nature are a wholesome neighbourhood.

Second, two unhealthy environment features are $N(ent)$ and $N(tob)$. This suggested that living in a neighbourhood with too many entertainment venues could affect worse health status. This effect may be brought by noise or various pollution produced by entertainment POIs. Moreover, the positive ATE of tobacco and liquor shop with a significance level $p<0.001$ is consistent with the traditional perception of tobacco and liquor shops' healthy-harmful characteristics. A probable assumption is that the availability of tobacco and alcohol could imply more potential consumption of this health-unfriendly merchandise.

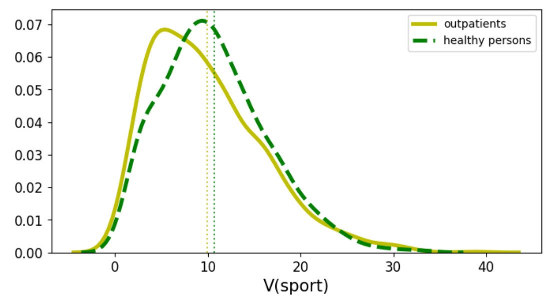
Analyses above demonstrate that the living environment possibly affects our health status. It is critical to living in a health-friendly environment that helps to lower health risk - an environment with more convenience to sports facilities and scenic parks but less entertainment, and tobacco/liquor POIs could help. We can also notice that after the matching procedure that balances the confounding effect among environment features, the estimated causal effects have various directions that differ from directions of correlation analysis. Under the scenario of our supposition on confounding relationships, the estimated causal effects are more explainable and compatible with our intuition.

5.2.3 Contextual Mobility Behaviours. The contextual mobility patterns depict the environment of locations that an individual has visited. Among all five categories of POIs having a significant correlation with health status, the number of entertainment and sports POIs around individual's visit locations $V(ent)$ and $V(sport)$ have significant ATEs. $V(ent)$ has an identical direction with $N(ent)$, indicating that spend time in entertainment venues is also linked with worse health status. The estimated treatment effect of $V(sport)$ is negative, which differs from the direction of correlation while corresponding with the estimated causal effect of $N(sport)$. This finding indicates that doing exercise in non-neighbouring sports facilities can also enhance our health status. Other contextual mobility patterns have insignificant treatment effects in this observational analysis.

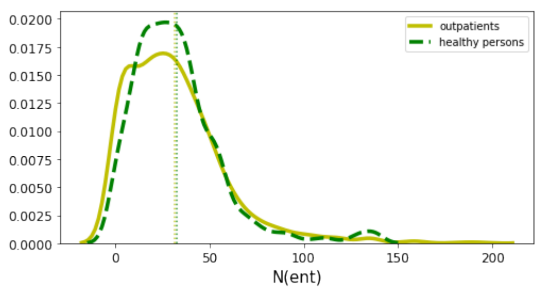
To sum up, there are nine mobility behaviour patterns with significant estimated ATEs on health status - physical mobility behaviours, four environment features and two contextual mobility behaviours. We may be able to adopt the results of causal analysis into health management and monitoring by following the instructions given by the direction of estimated causal effects. However, the results deserve deeper discussions. Potential unobserved confounding variables or the data set may bring errors to the estimation. Despite the possible errors, the causal analysis helps us to understand confounding effects within correlation studies. Next, we will demonstrate how causal analysis estimates effects differ from correlation analysis in direction.



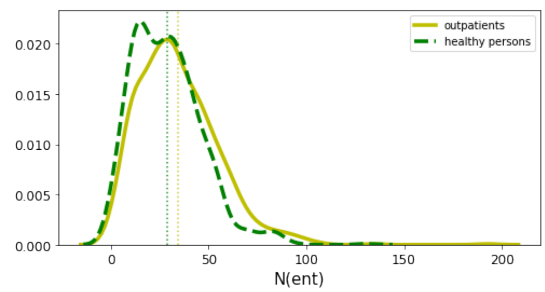
(a) The distributions of $V(sport)$ on unmatched populations.



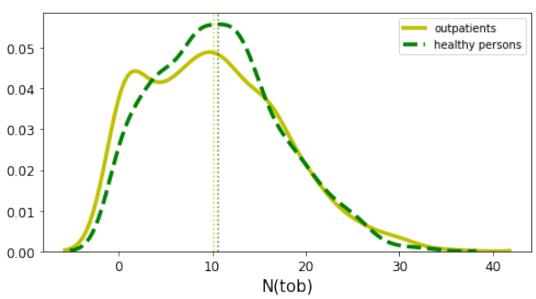
(b) The distributions of $V(sport)$ on matched populations.



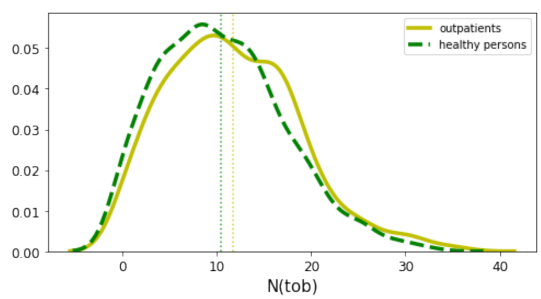
(c) The distributions of $N(ent)$ on unmatched populations.



(d) The distributions of $N(ent)$ on matched populations.



(e) The distributions of $N(tob)$ on unmatched populations.



(f) The distributions of $N(tob)$ on matched populations.

Fig. 4. Comparisons between causal analysis and correlation analysis.

5.2.4 Comparison of Causal and Correlation Analysis. To understand the origin of the differences between causal relations and correlations, we conduct a case study on the estimated causal effect and correlation coefficient of $V(sport)$. When we calculate the Pearson correlation coefficient between $V(sport)$ and health status, all observed individuals are taken into consideration. The distribution of $V(sport)$ on outpatients is depicted by the yellow curve in Figure 4(a), and the distribution of $V(sport)$ on healthy persons is depicted by the green curve. The average values on two populations, presented as vertical dotted lines, reveal that outpatients have a higher average value in $V(sport)$ ($p<0.01$). However, as discussed in Section 4.2, this naive comparison is affected by the confounding effect. To remove the confounding effect, we match pairs of individuals with similar confounding variables to simulate a randomized controlled test. The estimated ATE is then calculated by Equation 7. Since only pairs with different observed outcomes contribute to the ATE, we only consider individuals from matched healthy-unhealthy(outpatient) pairs to plot the distributions of $V(sport)$ in Figure 4(b). We can observe that after the matching procedure, outpatients have a lower average $V(sport)$ than healthy persons (p=0.06). Aside from $V(sport)$, environment features $N(ent)$ and $N(tob)$ also have different directions on estimated treatment effect and correlation. We conduct same experiments to compare correlation and causal analysis. As shown in Figure 4, healthy persons have a slightly higher average value in $N(ent)$(p=0.36) and $N(tob)$(p=0.17). After the matching procedure, both patterns have a higher average value in matched outpatients(p<0.01). These comparisons explain how the matching procedure provides causal directions different from the correlation analysis.

5.3 Mobility Pattern's Effect on Readmission

As discussed in Section 5.2, one error source of the observational causal analysis is potential unobserved confounding variables. They may remain unbalanced after the matching procedure, potentially biasing the estimated ATEs. One possible unobserved confounding variable is the individual's existing health status. Some individuals could have been predisposed to bad health before the period of mobility record collection, which will affect their observed mobility patterns and health outcome. Despite that our mobility traces contain solely pre-visit mobility records, individuals could have various existing health status that is infeasible to be detected. To rule out the confounding effect brought by existing health statuses, we re-examine the mobility record data and estimate the impact of mobility patterns on hospital readmission. In this supplementary experiment, the existing health status is controlled as unhealthy.

Although existing health status remains unknown in the original mobility data, the unhealthy individuals' hospital visits are verifiable. The outpatients' data are collected from a survey on medical experience, the data collectors have verified that the outpatients are visiting a hospital for medical treatment. We can select a session of mobility starting with a hospital visit for each outpatient, ensuring that the existing health status is constantly unhealthy. The outpatients are differentiated into two groups - outpatients with readmission and outpatients without readmission. Hospital visits are detected from mobility traces to distinguish outpatients with readmission. We identify 843 outpatients who re-visit the hospital two weeks after their first visit, and 843 outpatients without readmission after a visit. For readmission individuals, their mobility patterns are extracted from the session between two hospital visits. For non-readmission individuals, their mobility patterns are extracted from the session after their first hospital visit. The existing health status of all individuals is controlled to be unhealthy. Therefore, the confounding effect caused by existing health status is removed.

We apply the propensity matching procedure introduced in Section 4 on the readmission dataset to estimate mobility patterns' causal effect on hospital readmission. The estimated ATEs of mobility patterns on readmission and health status are listed in two columns in Table 5 respectively. Corresponding significance levels are marked alongside.

Previous works studying the association between mobility patterns and readmission have similar conclusions with the literature on understanding mobility pattern's relation with overall health status. Restriction in mobility range and mobility level is correlated with a higher risk of readmission [20, 23]. Contextual factors centered on residence are predictive of readmission rate according to [13, 71]. This similarity implies that mobility patterns may have alike mechanisms to influence readmission and overall hospital visit. It can be observed from Table 5 that the directions and significance of estimated ATEs on readmission are comparable with the ones on health outcome. Specifically, physical mobility behaviours all have negative significant ATEs. Environment features have identical directions in estimated ATEs, while tobacco shops in neighbourhood are not significant in affecting readmission and fast food stores have a larger impact. The credibility of their estimated ATEs on overall health status is therefore decreased. As for contextual mobility behaviours, the estimated ATE of visiting entertainment venues remains positive significant. However, the ATE of $V(sport)$ is positive insignificant. A possible explanation is that visiting non-residential sports facilities is not that effective for recovery from health problems comparing

Table 5. Comparison between estimated average treatment effects in two experiments.
***: p<0.0001, **: p<0.01, *: p<0.05

Category	Mobility Behaviour Patterns	Estimated ATE on Readmission	Estimated ATE on Health Status
Physical Mobility Behaviour	ρ	-0.06882 ***	-0.02796 ***
	σ	-0.05682 ***	-0.04904 ***
	ϵ	-0.05607 ***	-0.02291 ***
Environment Features	N(food)	-0.02542	-0.02537
	N(ent)	0.02463 *	0.08778 ***
	N(sport)	-0.04330 ***	-0.02822 **
	N(scene)	-0.04673 ***	-0.02127 **
	N(fast)	0.02932 *	0.02440
	N(tob)	0.01774	0.04660 ***
Contextual Mobility Behaviour	V(food)	-0.01189	-0.00166
	V(ent)	0.02148 *	0.02623 **
	V(sport)	0.00869	-0.02173 *
	V(fast)	0.03304 ***	0.00221
	V(tob)	0.00493	0.00052

with its effect in keeping fit. Meanwhile, visiting fast food may significantly impact on a higher risk of readmission. Most importantly, the directions of significant estimated ATEs are identical, which strengthen the credibility of their estimated ATEs on health status.

From the above discussion, the credibility of the experiment on readmission is assured since the existing health status' confounding effect is eliminated. Based on the hypothesis that readmission and overall health have similar mechanisms and mobility patterns affect them in the same direction, the results in the experiment on readmission can lend credibility to estimated ATEs on health outcome.

5.4 Estimated Effect's Robustness to Data Quality

Another challenge to the robustness of the estimated ATE is the possible error in the label of "healthy" and "unhealthy" outcomes. Several measures are taken to overcome this.

First, as introduced in Section 3.1, the "unhealthy" labelled individuals have the ground-truth unhealthy condition. This half of the labels are verified. Second, the "healthy" labelled individuals report no health problem when asked to donate their mobility records for research purposes. However, they might have latent unhealthy conditions afterwards but not visiting the hospital. This potential error should be considered in the experiment.

To check the effect of this potential error of data quality on estimated ATEs, we repeatedly change 5% of the label of "healthy" individuals into "unhealthy", representing the "healthy" labelled individuals with an underlying health problem. For each new population, we run the estimation of ATE introduced in Section 4 to obtain estimated treatment effects. This experiment is repeated 20 times. We concentrate on the mobility patterns with significant ATE estimated from the original population evaluate their robustness to data quality. In Table 6, we list the count of experiments where each mobility pattern has significant estimated ATEs and their corresponding directions.

From the results, we find that the directions of significant estimated ATEs remain unchanged. Among the mobility patterns, physical mobility behaviours have negative ATEs for all experiments. Living environment patterns are more robust than contextual mobility patterns, with $N(ent)$ and $N(sport)$ remaining significant in

Table 6. Robustness check of estimated ATEs on error in data quality.

Mobility Pattern	Number of Experiments with Significant ATE	Directions of Significant ATEs	Original Estimated ATE
ρ	20	-	-0.02796
σ	20	-	-0.04904
ϵ	20	-	-0.02291
N(ent)	15	+	0.08778
N(sport)	10	-	-0.02822
N(scene)	4	-	-0.02127
N(tob)	5	+	0.04660
V(ent)	5	+	0.02623
V(sport)	3	-	-0.02173

over half of the experiments. The above observations prove that the significant estimated treatment effects are robust to a small number of mislabels.

5.5 Estimating Causal Effects on Synthetic Data

A typical standard for evaluating the credibility of causal estimation is conducting experiments on synthetic data [31, 74]. We are able to specify the treatment effect when generating outcomes from confounding variables and treatment, and then assess how true effects can be retrieved by causal estimations.

For each mobility pattern as the treatment variable, we conduct the following experiment. First, we generate 5,000 individuals with the confounding variables of this treatment. The distributions of confounding variables are set to resemble the true distribution in our dataset. Second, we generate levels of treatment variables from the ordinal distribution in Equation 2. Distribution's parameters are also estimated from the true dataset. Third, we generate the outcome from confounding variables and the treatment level. The outcome follows a Bernoulli distribution, where the expectation is the sum of a linear term of the treatment level and a sigmoid function of confounding variables. The linear coefficient here denotes "the true causal effect". Here we set this value as the corresponding estimated ATE in the true dataset, listed in Table 4. Finally, we estimate causal effects on the synthetic data. The above procedure is repeated one hundred times for each treatment.

The confounding variables are generated based on their distributions in the real data. For each variable, we fit its distribution by maximum likelihood estimation. The fitted distributions are selected from four families - Bernoulli, gamma, exponential, and exponential normal. Here we plot the true distributions and the generated distribution of confounding variables(demographics, physical mobility behaviours, and living environments) in Figure 5.

The generation of treatment is based on the ordinal distribution in Equation 6. T is the treatment level and X represents confounding variables. The distribution's parameters θ_d and $\mathbf{w}$ are estimated from the real data. Then, for each generated individual, we generate its treatment level according to its cumulative distribution function decided by confounding variables and estimated parameters.

$$P(T \leq d|X) = \sigma(\theta_d - \mathbf{w}^T X). \tag{6}$$

The outcome of each individual is then determined by its confounding variables and treatment level. Equation 7 shows the expectation of the outcome conditioned on confounding variables and the treatment level. It consists of a function $f(\cdot)$ of confounding variables and a linear term of the treatment level. The linear coefficient stands

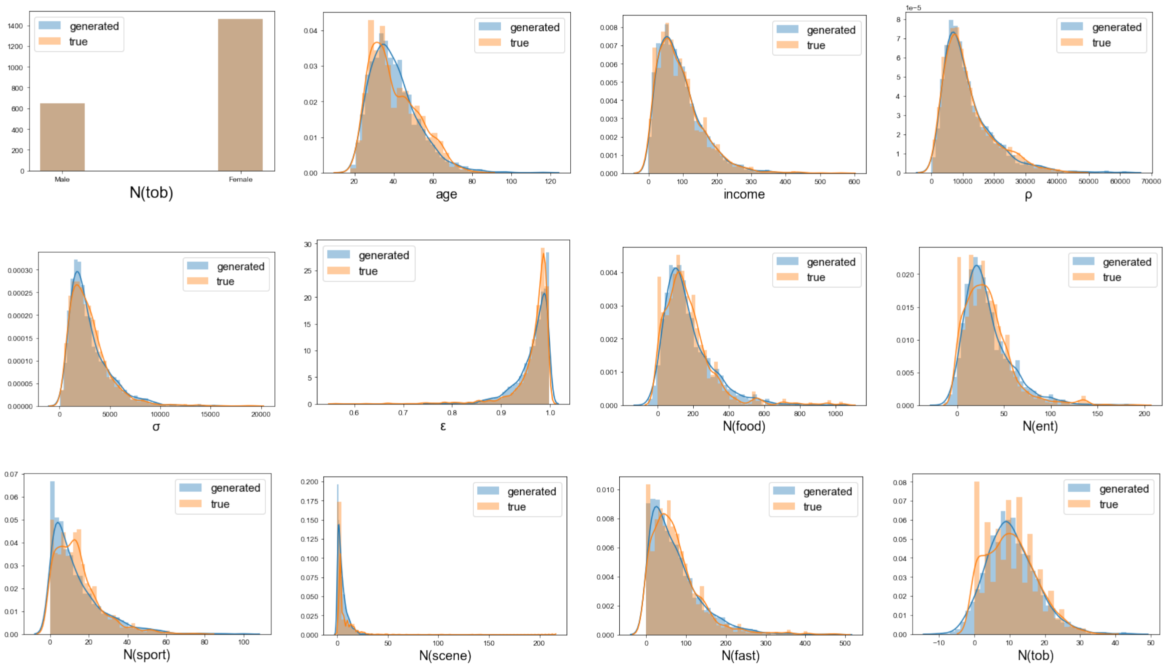


Fig. 5. Comparisons between real distributions and generated distributions of confounding variables.

for the true treatment effect, which can be set personalized. It should be acknowledged that the underlying mechanism of the causal structure is simplified into several manually selected functions. The goal is to generate reasonable data that can reflect the real observation data.

$$P(Y = 1|T, X) = f(X) + ATE \cdot T. \tag{7}$$

Following the above procedure of generation, we obtain 5,000 mimic individuals with confounding variables, treatment level, and outcome. Then we run the propensity score matching procedure to estimate ATEs on the synthetic dataset. For each treatment, we set the "true treatment effect" in Equation 7 as the estimated treatment effect from observational data, and run the experiment 100 times. We list the average values of estimated ATEs, minimum value, maximum value, 25% quartile, and 75% quartile for each treatment in Table 7.

We can observe from the results that all treatments have an average estimated ATE similar to the "true treatment effect" set in the generation procedure. Estimated ATEs on synthetic data have same directions with true ATEs. For treatment with significant ATE on observational data, the range of their estimated ATE on synthetic data always lies in the correct direction except for $N(scene)$ and $V(sport)$. Their maximum estimated ATEs in 100 experiment are positive numbers but with small magnitude. As for other patterns with insignificant "true treatment effects", their estimated ATEs on synthetic data span in both negative and positive directions. Typically, the range of $N(food)$ and $N(fast)$'s estimated ATEs extend over 0.12, while other patterns with similar range of "true treatment effect" less than 0.06. The estimations on synthetic data consolidate the effectiveness of our approach on estimating treatment effects. The directions of estimated ATEs from observational data are strengthened.

Table 7. Estimated treatment effects on synthetic data.

Treatment	True ATE	Average Estimated ATE	Minimum	25% Quartile	75% Quartile	Maximum
ρ	-0.02796	-0.02878	-0.04437	-0.03483	-0.02466	-0.01259
σ	-0.04904	-0.04967	-0.06820	-0.05589	-0.04493	-0.03451
ϵ	-0.02291	-0.02391	-0.03743	-0.02929	-0.02024	-0.00507
N(food)	-0.02537	-0.02825	-0.09392	-0.04859	-0.00865	0.03565
N(ent)	0.08778	0.08921	0.00269	0.06097	0.11183	0.16679
N(sport)	-0.02822	-0.02958	-0.05718	-0.03562	-0.02352	-0.00074
N(scene)	-0.02127	-0.02307	-0.05245	-0.03206	-0.01669	0.00228
N(fast)	0.02440	0.02234	-0.05274	-0.00486	0.04440	0.07968
N(tob)	0.04660	0.04549	0.02425	0.03846	0.05407	0.07598
V(food)	-0.00166	-0.00150	-0.03234	-0.00903	0.00639	0.03040
V(ent)	0.02623	0.02424	0.00049	0.01911	0.03024	0.04356
V(sport)	-0.02173	-0.02287	-0.05316	-0.02841	-0.01824	0.01058
V(fast)	0.00221	0.00212	-0.02437	-0.00598	0.01107	0.03374
V(tob)	0.00052	0.00067	-0.03477	-0.00748	0.00675	0.02938

5.6 Feature Selection for Prediction Model

We are interested in leveraging the results of causal analysis in health monitoring to improve the accuracy of the prediction of health status. For prediction models under supervised learning, the accuracy is affected by various factors. One factor is the noise in the training set leads to over-fitting. Input features with a small impact on model output but a large noise may be mistakenly grasped by the trained model. A common approach to prevent over-fitting brought by insignificant input features is L1 regularization [76]. The model is penalized by the sum of the learned weight's absolute values. L1 regularization can sparsify the model, where fewer input features are assigned with non-zero weight in the model when the penalty grows. Therefore, the performance of the L1-regularized model on the test set is usually better than the model with no regularization. The L1-regularization is often fused with regression models, while some other popular feature selection approaches preprocess features before training machine learning models. Univariate feature selection models choose features based on a univariate statistical test. ANOVA F-value, mutual information and simple correlation coefficients are commonly adopted as test metrics, measuring features' linear dependencies with the outcome variable [30]. By selecting features with the best above-mentioned test metric, noisy features can be filtered out to improve the prediction performance.

Inspired by the thought of model sparsification, we are curious if we can prevent over-fitting by removing the input features with insignificant estimated causal effects from the prediction model. We treat the mobility behaviour patterns as input features for prediction models and conduct 5-fold cross-validation on the dataset. We evaluate the average prediction accuracy, F1-score, and ROC AUC score [59] on test sets. The results confirm the effectiveness of our method.

5.6.1 Significance-based Feature Selection. The significance level of mobility behaviour pattern's estimated causal effects is listed in Table 4. They are determined by the p-value given by the t-test conducted on the individual treatment effects to evaluate to which extent the average treatment effect deviates from zero. Since the treatment effect represents the increase in health risk when the treatment increase by one level, we consider the input

Table 8. Performance comparison in prediction models with all features and selected features.

Prediction Method	Accuracy			F1-Score			ROC AUC		
	All Features	Selected Features	Increase	All Features	Selected Features	Increase	All Features	Selected Features	Increase
Naive Bayes	0.6279	0.6638	5.26%	0.5833	0.6330	8.52%	0.6902	0.7265	5.26%
Logistic Regression (Default)	0.6756	0.6766	0.30%	0.6657	0.6673	0.24%	0.7448	0.7420	-0.38%

features with a p-value greater than 0.05 as insignificant features which bring more noise than semantic in the prediction model.

We implement the causal analysis introduced in Section 4 on each training set in the 5-fold cross-validation. For each mobility behaviour pattern, it is discarded from the prediction model if the p-value of the significance test is greater than 0.05. The absolute estimated ATEs of each input feature in five training folds are presented in Figure 6. Input features with a p-value lower than 0.05 are highlighted by red frames and kept in the prediction model. For instance, $N(scene), V(food), V(ent), V(sport), V(scene)$ are filtered out when we conduct prediction on the first fold. On average, 6.6 input features are filtered out from each training set in the 5-fold cross-validation. To evaluate the strength of our causal-based feature selection, we examine the performances of the prediction model with all input features and the prediction model with selected input features. Here the prediction models are two widely-used models - naive Bayes classifier and logistic regression. All model's parameters are set to default values in the python package scikit-learn [57]. The performances are shown in the first two rows of Table 8.

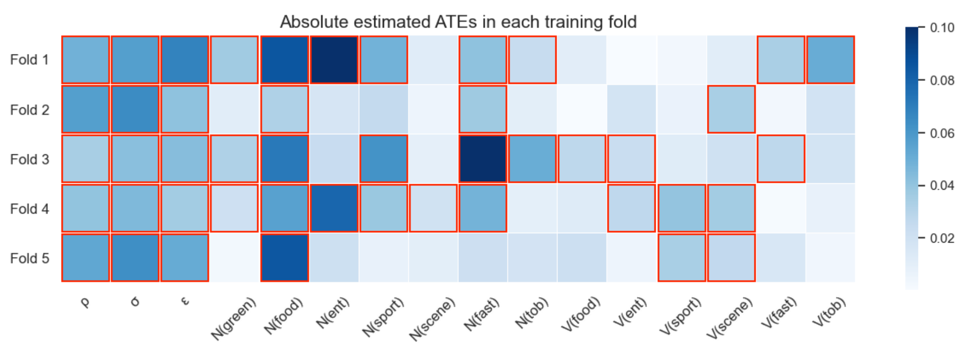


Fig. 6. A sketch map on the procedure of input feature selection.

5.6.2 *Analysis of Prediction Performance.* Here we can observe that most prediction metrics are improved by our causal-based feature selection. Welch's t-test shows that all metrics of the naive Bayes model are improved with a significant level with p-values lower than 0.01. The accuracy and F1-score of the logistic regression model are improved insignificantly(p=0.9). The comprehensive advances in prediction performances corroborate the effectiveness of causal-based feature selection.

As discussed above, L1-regularization and univariate feature selection methods are powerful approaches to prevent over-fitting and improve prediction performance. We are interested in the comparison between the effect of our causal-based feature selection and their effects. First, we remove features with insignificant estimated causal effect from the input and train a naive Bayes model and a logistic regression model respectively with no regularization. Then we train a default L1-regularized logistic regression model with all features as input and use the features chosen by L1-regularization to train a naive Bayes model. As for univariate feature selection methods,

Table 9. Performance comparison in prediction models with various feature selection methods.

Prediction Method	Feature Selection Method	Accuracy	F1-Score	ROC AUC
Naive Bayes	L1-Regularization	0.6738	0.6618	0.7419
	ANOVA-based	0.6747	0.6657	0.7424
	Mutual Information-based	0.6534	0.6401	0.7220
	Correlation-based	0.6742	0.6641	0.7426
	Causal-based	**0.6804**	**0.6718**	**0.7426**
Logistic Regression	L1-Regularization	0.6439	0.5992	0.7109
	ANOVA-based	0.6302	0.5857	0.6985
	Mutual Information-based	0.6368	0.5990	0.6963
	Correlation-based	0.6368	0.5943	0.7018
	Causal-based	**0.6638**	**0.6330**	**0.7265**

we select 80% best features based on ANOVA F-test and mutual information respectively, and select features with significant correlation coefficient($p<0.05$) with the outcome to train naive Bayes and logistic regression models.

The prediction results are shown in Table 9. Our causal-based features selection achieves the best prediction performance on each metric and both prediction models. For the naive Bayes model, the accuracy and F1-score of our method are significantly higher than all other feature selection approaches($p<0.05$), while the ROC AUC is higher than L1-regularization with $p=0.064$. For the logistic regression model, our method does not significantly outperform other methods. From the perspective of model sparsity, the L1-regularization filters out 4.0 input features from each training set on average. Ours performs similarly with L1-regularization with 65% more features filtered. This provides strong evidence that the causal-based feature selection can be as effective as L1-regularization in preventing over-fitting. To conclude, we adopt the casual approach introduced in Section 4 in prediction models to filter out features with insignificant effects on the outcome. By this measure, the model's robustness and interpretability are improved.

6 DISCUSSION

6.1 Implications and Applications

To deal with the confounding effect brought by the correlation analysis, we go beyond correlation to the causal relation between urban mobility and health. We leverage a propensity score-based matching method to simulate a randomized controlled trial. A confounding structure is put forward to help understanding causal effects. The estimated treatment effects represent the elevation in the potential health risk when we intervene in a mobility behaviour pattern to raise one level. We analyze the effect of propensity matching and estimated causal direction and significance of mobility patterns given by the causal framework. Furthermore, a causal significance-based input feature selection method is put forward for robust health prediction models. Various implications for researchers and the health-related community can be drawn from these findings and applied to multiple aspects.

Before discussing any inferred implications or applications, we should reiterate that thorough privacy protection protocols have been enforced in the current research on observational data. The causal approach introduced in Section 4 relies on the accessibility user's demographic and mobility data, which are obtained under the user's approval. The data are carefully desensitized and are protected under non-disclosure agreements. In actual health-related applications or systems that require user mobility data, more strict considerations should be taken to minimize the invasion of user privacy as much as possible. Data anonymization, user's authorization, and a safe data container are necessary measures. Moreover, privacy protection mechanisms for mobility data have been proposed in recent works, including federated learning [21], differential privacy [85], and k-anonymity [29]. Designed systems should embed one of the privacy protection methods on user's mobility and location data.

Our causal analysis framework has important design implications for ubiquitous computing. First, in the simple correlation analysis, the negative correlations between environment features and health risk are counter-intuitive and confounded by variables that simultaneously affect health and environment features. The causal analysis we proposed can relieve the confounding effects in correlation analysis by balancing confounding variables. Second, the causal-based feature selection method removes input features with insignificant estimated causal effects on the predicted value in prediction models. This strategy can alleviate the noise brought by input features and improve the accuracy and ability of the generalization of prediction models. These findings imply the feasibility of leaping from association to causation in broader ubiquitous computing topics, not just restricted in health monitoring. With the improvement in ubiquitous computing, signals from various sources are collected by wearable devices and mobile phones. Under cautious consideration of privacy and ethic, the extracted patterns may have different contributions in downstream applications such as mental health monitoring and emotion interpretation. Studying the causal effect of patterns on the predicted value can provide a more accurate understanding of the pattern's characteristics. In addition, before the input features are fed for subsequent prediction tasks, a causal-based feature selection might be a profitable strategy for both better accuracy and a lightweight model.

Our causal analysis reveals that mobility behaviour patterns have various impacts on health outcome. The direction of a mobility pattern's estimated causal effect implies its role in influencing health status. The estimation results demonstrate that the increase in range, activeness, and diversity of mobility can lower health risks. Living environments with fewer entertainment places and tobacco/liquor shops, more scenic parks and sports facilities are assumed beneficial for one's health. More sports activity, fewer visits to entertainment venues are proposed positive for better health. Moreover, the significance test shows that the contextual mobility behaviours are not as significant as the environment of residence in influencing health outcome. These findings can give advice to health monitoring. On an individual level, since the estimated causal effect is representing the change in outcome under the intervention on treatment, the directions of estimated causal effects can serve as guidance for fostering a healthier personal lifestyle. The causal analysis on the living environment and contextual mobility behaviours can be applied in choosing a health-friendly neighbourhood to reside in or planning fewer visits harmful to our health. Even if our living environment is not health-supportive, we can cultivate healthy mobility habits such as a diverse moving pattern and more fitness activities. On a community level, our analysis can be applied for building residential areas with a healthy environment. Under the concern of community health, urban planners could evaluate the health benefits of the current living environment in residential areas according to the distribution of neighbouring POIs to assign more health-beneficial venues, such as sports facilities and parks for neighbours with low health benefits. Policymakers could also intervene in a district's health benefit by prohibiting the excess construction of health-harmful POIs.

6.2 Discussion on Experiment Settings

6.2.1 Discussion on Treatment Stratification. As introduced in Section 4, when estimating the causal effect of a specific mobility pattern, we stratify continuous treatment variable into six discrete levels. This stratification strategy is also used in previous literature investigating the causal effects of continuous treatments [31]. The number of stratification levels is chosen based on the quantity of our dataset and distributions of variables. In [31], the authors binned the treatment into four levels with 625 samples in total. Since we have over two thousand individuals, we raise the bin number to six.

In order to clarify stratification's influence on the estimated ATEs, we conduct further experiments with different choices on the number of bin levels. For stratification with a level of 3 to 8, mobility patterns with significant estimated ATE under specific stratification strategy are listed in Table 10. The directions of ATEs

Table 10. Mobility patterns with significant estimated ATE under different stratification strategies.

Number of binned levels	**Mobility patterns with significant estimated ATEs**
3	ρ(-), σ(-), ϵ(-), N(food)(+), N(ent)(+), N(sport)(-), N(fast)(+), V(sport)(-)
4	ρ(-), σ(-), ϵ(-), N(food)(+), N(ent)(+), N(sport)(-), N(scene)(-), N(fast)(+), N(tob)(+), V(sport)(-)
5	ρ(-), σ(-), ϵ(-), N(sport)(-), N(scene)(-) N(fast)(+), V(sport)(-)
6	ρ(-), σ(-), ϵ(-), N(ent)(+), N(sport)(-), N(scene)(-), N(tob)(+), V(ent)(+), V(sport)(-)
7	ρ(-), σ(-), ϵ(-), N(food)(+), N(fast)(+), V(food)(-), V(fast)(-)
8	ρ(-), σ(-), ϵ(-), N(food)(+), N(ent)(+), N(scene)(-), N(fast)(+), V(ent)(+), V(fast)(-)

are marked in the brackets. From the results, we can observe that the directions of significant treatment effects remain identical when the number of stratification levels varies, although ATEs' magnitudes alter with the change of stratification strategy. The stability in ATE direction clarifies the effectiveness of the propensity score matching procedure in removing confoundedness under certain confounding assumptions.

6.2.2 Discussion on Confounding Variables. In Section 4.2, we introduce reasons for the decision of causal structure and confounding variables. It can be observed that we select confounding variables based on subjective knowledge and previous works that established associations between variables. One may challenge that there is not sufficient evidence of the causal relationships we assume. In fact, the problem of constructing causal relationships from a set of variables has extremely high computational complexity. A possible approach is a causal discovery based on graphical models [32] that find the most likely causal graph on given features under several assumptions and restrictions. To avoid taking excessive assumptions, the best alternative to determine confounding variables is to use subject recognition and related literature. We demonstrate that confounding variables selected by premised knowledge are fairly balanced by the matching procedure, which promises credible estimation of treatment effect under our assumptions.

Another problem is that there might have potential unobserved confounding variables for mobility behaviours. The dataset used in our work contains limited demographics under privacy concerns. Potential confounding variables include occupation [15], employment [66], education [9]. Despite this limitation, we note that the causal edge between the potential confounding variables may not have enough effect that is worth being considered in the causal structure. We only need to take the key variables into consideration.

6.3 Future Works

The current study may not comprehensively consider every detail, however, it does illustrate an approach to explore the causal effects of mobility patterns on individual health using passively collected mobility record data for the first time. In the future, we plan to incorporate more categories of contextual behaviours so as to consider broader possible causes on health status, instead of restricted in the six categories in the current work that has been proved health-correlated in previous studies. More detailed individual health reports with a longer duration can be adopted as the data source of health status, relieving the confounding effect brought by unspecified health status. Besides, since our analysis is limited to the city of Beijing, we would prefer to conduct the causal analysis on broader populations and populations from multiple cities, where new insights may be possibly drawn. Based on the analyses on populations covering various cultures and scales, we would like to be devoted to design an efficient causal-aware health monitoring model based on mobility behaviour that embeds the awareness of their causal effects on health status.

7 CONCLUSION

In this paper, we quantify the causal effect of urban mobility behaviours on individual health status. We collect 2,112 individuals' mobility traces from the cellular network for two months and their corresponding health conditions. Sixteen mobility behaviour patterns are extracted from the mobility traces, including physical mobility behaviours, environment features, and contextual mobility behaviours. Case studies show that the correlation directions may be confounded by confounding variables. We propose a framework based on the propensity score matching method to estimate the causal effect of mobility on health status by removing the bias brought by confounding effects. We prove that the matching procedure balances the distributions of confounding variables on treatment groups to simulate a randomized controlled trial. Our results provide new insight into understanding the role of mobility behaviour in influencing health status. Physical mobility behaviours and living environment have significant estimated effects on health status. The directions of estimated causal effects imply that a larger mobility range, higher visiting diversity, more sports facilities and scenic parks, and fewer entertainment venues and tobacco shops in the neighbourhood can be beneficial to better health status. We also observe that some causal directions are different from correlation directions. For instance, visiting more sports facilities are approximated to be health-friendly, while the correlation direction is opposite. A supplementary experiment on mobility pattern's effect on hospital readmission demonstrates that most patterns have a similar effect on overall health status and readmission. This result lends more credibility to the estimated effects on health status. Furthermore, we utilize the significance of estimated causal effect as a guideline for feature selection in health prediction tasks. Our strategy of filtering out causally insignificant input features to prevent over-fitting. Prediction results of naive Bayes prediction models increase significantly by this selection strategy, and it outperforms other widely-used feature selection approaches. These improvements validate the efficiency of our selection method. Our work can be adapted for various applications. The thought of de-confounding and causal-based feature selection for the prediction can be adopted for broader health analysis. The estimated effects of mobility patterns can be considered in urban policy planning and individual health monitoring.

ACKNOWLEDGMENTS

This work was supported in part by The National Key Research and Development Program of China under grant 2020AAA0106000, the National Natural Science Foundation of China under U1936217, 61971267, 61972223, 61941117, 61861136003.

REFERENCES

[1] Daniel A Adler, Vincent W-S Tseng, Gengmo Qi, Joseph Scarpa, Srijan Sen, and Tanzeem Choudhury. 2021. Identifying Mobile Sensing Indicators of Stress-Resilience. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–32.

[2] Tim Althoff, Jennifer L Hicks, Abby C King, Scott L Delp, Jure Leskovec, et al. 2017. Large-scale physical activity data reveal worldwide activity inequality. *Nature* 547, 7663 (2017), 336–339.

[3] Peter C Austin. 2008. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in medicine* 27, 12 (2008), 2037–2049.

[4] Sangwon Bae, Anind K Dey, and Carissa A Low. 2016. Using passively collected sedentary behavior to predict hospital readmission. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 616–621.

[5] Shan Bao-yan, Wang Jun-ning, Liu Yang-yang, and Zhang Zhi-xuan. 2020. Spatial Correlation Analysis of Residential Area and Service Industry Distribution Based on POI Big Data. In *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics*

(ICCCBDA). IEEE, 453–458.
[6] Jo Barton, Rachel Bragg, Carly Wood, and Jules Pretty. 2016. *Green exercise: Linking nature, health and well-being.* Routledge.
[7] Heidi M Blanck, Diana Allen, Zarnaaz Bashir, Nina Gordon, Alyson Goodman, Dee Merriam, and Candace Rutt. 2012. Let's go to the park today: The role of parks in obesity prevention and improving the public's health. *Childhood Obesity (Formerly Obesity and Weight Management)* 8, 5 (2012), 423–428.
[8] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. 2006. The scaling laws of human travel. *Nature* 439, 7075 (2006), 462–465.
[9] Giorgio Brunello, Margherita Fort, Nicole Schneeweis, and Rudolf Winter-Ebmer. 2016. The causal effect of education on health: what is the role of health behaviors? *Health economics* 25, 3 (2016), 314–336.
[10] Giovanna Calogiuri and Lewis R Elliott. 2017. Why do people exercise in natural environments? Norwegian adults' motives for nature-, gym-, and sports-based exercise. *International journal of environmental research and public health* 14, 4 (2017), 377.
[11] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 1293–1304.
[12] Carl J Caspersen, Mark A Pereira, and Katy M Curran. 2000. Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Medicine & Science in Sports & Exercise* 32, 9 (2000), 1601–1609.
[13] Marcelo Cerullo, Faiz Gani, Sophia Y Chen, Joseph K Canner, and Timothy M Pawlik. 2016. Readmission after major surgery: effect of the postdischarge environment. *Journal of Surgical Research* 205, 2 (2016), 318–326.
[14] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. Double/debiased machine learning for treatment and structural parameters.
[15] Sukyung Chung, Marisa E Domino, Sally C Stearns, and Barry M Popkin. 2009. Retirement and physical activity: analyses by occupation and wealth. *American journal of preventive medicine* 36, 5 (2009), 422–428.
[16] Amanda Deeks, Catherine Lombard, Janet Michelmore, and Helena Teede. 2009. The effects of gender and age on health related behaviors. *BMC Public Health* 9, 1 (2009), 1–8.
[17] Rajeev H Dehejia and Sadek Wahba. 2002. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics* 84, 1 (2002), 151–161.
[18] Jack Edmonds. 1965. Maximum matching and a polyhedron with 0, 1-vertices. *Journal of research of the National Bureau of Standards B* 69, 125-130 (1965), 55–56.
[19] Susan L Ettner. 1996. New evidence on the relationship between income and health. *Journal of health economics* 15, 1 (1996), 67–85.
[20] Roya Fathi, Peter Bacchetti, Mary N Haan, Thomas K Houston, Kanan Patel, and Christine S Ritchie. 2017. Life-space assessment predicts hospital readmission in home-limited adults. *Journal of the American Geriatrics Society* 65, 5 (2017), 1004–1011.
[21] Jie Feng, Can Rong, Funing Sun, Diansheng Guo, and Yong Li. 2020. PMF: A privacy-preserving human mobility prediction framework via federated learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–21.
[22] Rogerio Fermino, Rodrigo Reis, Pedro C Hallal, and Andrew T Kaczynski. 2015. Who are the users of urban parks? A study with adults from Curitiba, Brazil. *Journal of Physical Activity and Health* 12, 1 (2015), 58–67.
[23] Steve R Fisher, Yong-Fang Kuo, Gulshan Sharma, Mukaila A Raji, Amit Kumar, James S Goodwin, Glenn V Ostir, and Kenneth J Ottenbacher. 2013. Mobility after hospital discharge as a marker for 30-day readmission. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 68, 7 (2013), 805–810.
[24] Friederike Fleischer. 2007. "To Choose a House Means to Choose a Lifestyle." The Consumption of Housing and Class-Structuration in Urban China. *City & Society* 19, 2 (2007), 287–311.
[25] Christian Fong, Chad Hazlett, Kosuke Imai, et al. 2018. Covariate balancing propensity score for a continuous treatment: application to the efficacy of political advertisements. *The Annals of Applied Statistics* 12, 1 (2018), 156–177.
[26] Panagis Galiatsatos, Cynthia Kineza, Seungyoun Hwang, Juliana Pietri, Emily Brigham, Nirupama Putcha, Cynthia S Rand, Meredith McCormack, and Nadia N Hansel. 2018. Neighbourhood characteristics and health outcomes: evaluating the association between socioeconomic status, tobacco store density and health outcomes in Baltimore City. *Tobacco control* 27, e1 (2018), e19–e24.
[27] Thomas A Glass, Steven N Goodman, Miguel A Hernán, and Jonathan M Samet. 2013. Causal inference in public health. *Annual review of public health* 34 (2013), 61–75.
[28] Marta C. González, César A. Hidalgo, and Albert-László Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (jun 2008), 779–782.
[29] Marco Gramaglia and Marco Fiore. 2015. Hiding mobile traffic fingerprints with glove. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. 1–13.
[30] Isabelle Guyon and André Elisseeff. 2003. An introduction to variable and feature selection. *Journal of machine learning research* 3, Mar (2003), 1157–1182.
[31] Apinan Hasthanasombat and Cecilia Mascolo. 2019. Understanding the effects of the neighbourhood built environment on public health with open data. In *The World Wide Web Conference*. 648–658.
[32] David Heckerman, Christopher Meek, and Gregory Cooper. 1999. A Bayesian approach to causal discovery. *Computation, causation, and discovery* 19 (1999), 141–166.

[33] Martin Hillebrand, Imran Khan, Filipa Peleja, and Nuria Oliver. 2020. MobiSenseUs: Inferring Aggregate Objective and Subjective Well-being from Mobile Data. (2020).
[34] Keisuke Hirano and Guido W Imbens. 2004. The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives* 226164 (2004), 73–84.
[35] Daniel G Horvitz and Donovan J Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association* 47, 260 (1952), 663–685.
[36] Yu Huang, Haoyi Xiong, Kevin Leach, Yuyan Zhang, Philip Chow, Karl Fua, Bethany A Teachman, and Laura E Barnes. 2016. Assessing social anxiety using GPS trajectories and point-of-interest data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 898–903.
[37] Kosuke Imai and David A Van Dyk. 2004. Causal inference with general treatment regimes: Generalizing the propensity score. *J. Amer. Statist. Assoc.* 99, 467 (2004), 854–866.
[38] Sanae Inagami, Deborah A Cohen, Brian Karl Finch, and Steven M Asch. 2006. You are where you shop: grocery store locations, weight, and neighborhoods. *American journal of preventive medicine* 31, 1 (2006), 10–17.
[39] Woo-Sung Jung, Fengzhong Wang, and H Eugene Stanley. 2008. Gravity model in the Korean highway. *EPL (Europhysics Letters)* 81, 4 (2008), 48005.
[40] Gary King and Richard Nielsen. 2019. Why propensity scores should not be used for matching. *Political Analysis* 27, 4 (2019), 435–454.
[41] Kalevi M Korpela and Matti Ylen. 2007. Perceived health is associated with visiting natural favourite places in the vicinity. *Health & Place* 13, 1 (2007), 138–151.
[42] Kurt Kroenke, Robert L Spitzer, Janet BW Williams, and Bernd Löwe. 2010. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *General hospital psychiatry* 32, 4 (2010), 345–359.
[43] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences* 116, 10 (2019), 4156–4165.
[44] Zongyu Lin, Shiqing Lyu, Hancheng Cao, Fengli Xu, Yuqiong Wei, Hanan Samet, and Yong Li. 2020. HealthWalks: Sensing Fine-grained Individual Health Condition via Mobility Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–26.
[45] Ariel Linden, John L Adams, and Nancy Roberts. 2005. Using propensity scores to construct comparable control groups for disease management program evaluation. *Disease Management & Health Outcomes* 13, 2 (2005), 107–115.
[46] Bo Lu, Elaine Zanutto, Robert Hornik, and Paul R Rosenbaum. 2001. Matching with doses in an observational study of a media campaign against drug abuse. *J. Amer. Statist. Assoc.* 96, 456 (2001), 1245–1253.
[47] Wolfgang Lutz, Warren Sanderson, and Sergei Scherbov. 2008. The coming acceleration of global population ageing. *Nature* 451, 7179 (2008), 716–719.
[48] Laura Macdonald, Steven Cummins, and Sally Macintyre. 2007. Neighbourhood fast food environment and area deprivation—substitution or concentration? *Appetite* 49, 1 (2007), 251–254.
[49] Kate E Mason, Neil Pearce, and Steven Cummins. 2018. Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank. *The Lancet Public Health* 3, 1 (2018), e24–e33.
[50] Peter McCullagh. 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)* 42, 2 (1980), 109–127.
[51] Carme Miralles-Guasch, Montserrat Martínez Melo, and Oriol Marquet. 2016. A gender analysis of everyday mobility in urban and rural territories: from challenges to sustainability. *Gender, Place & Culture* 23, 3 (2016), 398–417.
[52] Mehrab Bin Morshed, Koustuv Saha, Richard Li, Sidney K D'Mello, Munmun De Choudhury, Gregory D Abowd, and Thomas Plötz. 2019. Prediction of mood instability with passive sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–21.
[53] Apostolos Papagiannakis, Ioannis Baraklianos, and Alexia Spyridonidou. 2018. Urban travel behaviour and household income in times of economic crisis: Challenges and perspectives for sustainable mobility. *Transport policy* 65 (2018), 51–60.
[54] Judea Pearl. 1995. Causal diagrams for empirical research. *Biometrika* 82, 4 (1995), 669–688.
[55] Judea Pearl. 2014. Interpretation and identification of causal mediation. *Psychological methods* 19, 4 (2014), 459.
[56] Judea Pearl et al. 2009. Causal inference in statistics: An overview. *Statistics surveys* 3 (2009), 96–146.
[57] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.
[58] Jacob Poushter et al. 2016. Smartphone ownership and internet usage continues to climb in emerging economies. *Pew research center* 22, 1 (2016), 1–44.
[59] Sebastian Raschka. 2014. An overview of general performance metrics of binary classifier systems. *arXiv preprint arXiv:1410.5330* (2014).
[60] Nina Rautio, Svetlana Filatova, Heli Lehtiniemi, and Jouko Miettunen. 2018. Living environment and its relationship to depressive mood: a systematic review. *International journal of social psychiatry* 64, 1 (2018), 92–103.

[61] Paul R Rosenbaum and Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 1 (1983), 41–55.

[62] Babak Roshanaei-Moghaddam, Wayne J Katon, and Joan Russo. 2009. The longitudinal effects of depression on physical activity. *General hospital psychiatry* 31, 4 (2009), 306–315.

[63] Donald B Rubin. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* 66, 5 (1974), 688.

[64] Koustuv Saha, Benjamin Sugar, John Torous, Bruno Abrahao, Emre Kıcıman, and Munmun De Choudhury. 2019. A social media study on the effects of psychiatric medication use. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 440–451.

[65] William A Satariano, Jack M Guralnik, Richard J Jackson, Richard A Marottoli, Elizabeth A Phelan, and Thomas R Prohaska. 2012. Mobility and aging: new directions for public health action. *American journal of public health* 102, 8 (2012), 1508–1515.

[66] Hendrik Schmitz. 2011. Why are the unemployed in worse health? The causal effect of unemployment on health. *Labour economics* 18, 1 (2011), 71–78.

[67] Calvin Schnure and Shruthi Venkatesh. 2015. Demographic and financial determinants of housing choice in retirement and the rise of senior living. *Available at SSRN 2588026* (2015).

[68] Patrick Schwab, Lorenz Linhardt, Stefan Bauer, Joachim M Buhmann, and Walter Karlen. 2020. Learning counterfactual representations for estimating individual dose-response curves. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 5612–5619.

[69] Chanuki Illushka Seresinhe, Tobias Preis, and Helen Susannah Moat. 2015. Quantifying the impact of scenic environments on health. *Scientific reports* 5, 1 (2015), 1–9.

[70] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. 2012. A universal model for mobility and migration patterns. *Nature* 484, 7392 (2012), 96–100.

[71] Erica S Spatz, Susannah M Bernheim, Leora I Horwitz, and Jeph Herrin. 2020. Community factors and hospital wide readmission rates: Does context matter? *PloS one* 15, 10 (2020), e0240222.

[72] Jerzy Splawa-Neyman, Dorota M Dabrowska, and TP Speed. 1990. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* (1990), 465–472.

[73] Student. 1908. The probable error of a mean. *Biometrika* (1908), 1–25.

[74] Wei Sun, Pengyuan Wang, Dawei Yin, Jian Yang, and Yi Chang. 2015. Causal inference via sparse additive models with application to online advertising. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

[75] Tetsuya Takahashi, Megumi Kumamaru, Sue Jenkins, Masakazu Saitoh, Tomoyuki Morisawa, and Hikaru Matsuda. 2015. In-patient step count predicts re-hospitalization after cardiac surgery. *Journal of Cardiology* 66, 4 (oct 2015), 286–291.

[76] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 1 (1996), 267–288.

[77] Sebastián Valenzuela, Arturo Arriagada, and Andrés Scherman. 2014. Facebook, Twitter, and youth engagement: A quasi-experimental study of social media use and protest behavior using propensity score matching. *International Journal of Communication* 8 (2014), 25.

[78] Magdalena Van den Berg, Mireille van Poppel, Irene van Kamp, Sandra Andrusaityte, Birute Balseviciene, Marta Cirach, Asta Danileviciute, Naomi Ellis, Gemma Hurst, Daniel Masterson, et al. 2016. Visiting green space is associated with mental health and vitality: A cross-sectional study in four european cities. *Health & place* 38 (2016), 8–15.

[79] Mark J Van der Laan and Sherri Rose. 2011. *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media.

[80] Stefan Wager and Susan Athey. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* 113, 523 (2018), 1228–1242.

[81] May C Wang, Soowon Kim, Alma A Gonzalez, Kara E MacLeod, and Marilyn A Winkleby. 2007. Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *Journal of Epidemiology & Community Health* 61, 6 (2007), 491–498.

[82] Yingzi Wang, Xiao Zhou, Cecilia Mascolo, Anastasios Noulas, Xing Xie, and Qi Liu. 2018. Predicting the Spatio-Temporal Evolution of Chronic Diseases in Population with Human Mobility Data. IJCAI.

[83] Bernard L Welch. 1947. The generalization ofstudent's' problem when several different population variances are involved. *Biometrika* 34, 1/2 (1947), 28–35.

[84] Jennifer R Wolch, Jason Byrne, and Joshua P Newell. 2014. Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and urban planning* 125 (2014), 234–244.

[85] Yonghui Xiao and Li Xiong. 2015. Protecting locations with differential privacy under temporal correlations. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 1298–1309.

[86] Fengli Xu, Zongyu Lin, Tong Xia, Diansheng Guo, and Yong Li. 2020. Sume: Semantic-enhanced urban mobility network embedding for user demographic inference. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–25.

[87] Fengli Xu, Tong Xia, Hancheng Cao, Yong Li, Funing Sun, and Fanchao Meng. 2018. Detecting popular temporal modes in population-scale unlabelled trajectory data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–25.

[88] Hiroyuki Yamada and Anthony S Bryk. 2016. Assessing the first two years' effectiveness of Statway®: A multilevel model with propensity score matching. *Community College Review* 44, 3 (2016), 179–204.
[89] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 186–194.
[90] Shannon N Zenk, Amy J Schulz, Stephen A Matthews, Angela Odoms-Young, JoEllen Wilbur, Lani Wegrzyn, Kevin Gibbs, Carol Braunschweig, and Carmen Stokes. 2011. Activity space environment and dietary and physical activity behaviors: a pilot study. *Health & place* 17, 5 (2011), 1150–1161.
[91] Xin Zhang, Qunhong Wu, Yongxiang Shao, Wenqi Fu, Guoxiang Liu, and Peter C Coyte. 2015. Socioeconomic inequities in health care utilization in China. *Asia Pacific Journal of Public Health* 27, 4 (2015), 429–438.
[92] Yuyang Zhang and Peter C Coyte. 2020. Inequality of opportunity in healthcare expenditures: evidence from China. *BMC health services research* 20 (2020), 1–11.
[93] Yunke Zhang, Fengli Xu, Tong Li, Vassilis Kostakos, Pan Hui, and Yong Li. 2021. Passive Health Monitoring Using Large Scale Mobility Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–23.
[94] Dawei Zhu, Na Guo, Jian Wang, Stephen Nicholas, and Li Chen. 2017. Socioeconomic inequalities of outpatient and inpatient service utilization in China: personal and regional perspectives. *International journal for equity in health* 16, 1 (2017), 1–10.
[95] Qingru Zou, Xiangming Yao, Peng Zhao, Heng Wei, and Hui Ren. 2018. Detecting home location and trip purposes for cardholders by mining smart card transaction data in Beijing subway. *Transportation* 45, 3 (2018), 919–944.