



OPEN

A multilevel scenario based predictive analytics framework to model the community mental health and built environment nexus

Sayanti Mukherjee^{1✉}, Emmanuel Frimpong Boamah², Prasangsha Ganguly¹ & Nisha Botchwey³

The built environment affects mental health outcomes, but this relationship is less studied and understood. This article proposes a novel multi-level scenario-based predictive analytics framework (MSPAF) to explore the complex relationships between community mental health outcomes and the built environment conditions. The MSPAF combines rigorously validated interpretable machine learning algorithms and scenario-based sensitivity analysis to test various hypotheses on how the built environment impacts community mental health outcomes across the largest metropolitan areas in the US. Among other findings, our results suggest that declining socio-economic conditions of the built environment (e.g., poverty, low income, unemployment, decreased access to public health insurance) are significantly associated with increased reported mental health disorders. Similarly, physical conditions of the built environment (e.g., increased housing vacancies and increased travel costs) are significantly associated with increased reported mental health disorders. However, this positive relationship between the physical conditions of the built environment and mental health outcomes does not hold across all the metropolitan areas, suggesting a mixed effect of the built environment's physical conditions on community mental health. We conclude by highlighting future opportunities of incorporating other variables and datasets into the MSPAF framework to test additional hypotheses on how the built environment impacts community mental health.

The physical and socio-economic aspects of the built environment impact population mental health outcomes of a community. The physical aspects include human-created infrastructure systems, such as transportation and housing infrastructure systems, which support the functioning of people within a community¹. The socio-economic aspects refer to the economic, racial and ethnic, and relational conditions that may influence a person's ability to function well, both physically and psychologically, within their communities. Studies have examined how such physical and socio-economic aspects of the built environment impact a community's overall health and well-being in terms of crime rates², educational performance, property values³, and various health outcomes such as obesity, heart disease, cancer, stroke, respiratory disease, diabetes, and suicide rates⁴⁻⁷. More specifically, understanding and predicting health outcomes as a function of the built environment is a significant focus among urban planning, public health and allied professionals. The SARS-COV-2 pandemic has further exacerbated the urgency to understand how such aspects of the built environment influence health outcomes. Examples include—studying the reasons for faster disease spread within the vulnerable and minority population based on the socio-economic conditions and physical setting of their surrounding built environment^{8,9}, investigating how the different conditions within the built environment (e.g., sanitation conditions and closed and open areas) or different types of physical surfaces (e.g., metal and other solid surfaces, and water) aid in the spread of the virus^{10,11}.

Mental health is one of the specific health outcomes impacted by the built environment. Mental illness or disorder contribute significantly to the global burden of disease, accounting for 32.4% of years lived with disability

¹Department of Industrial and Systems Engineering, School of Engineering and Applied Sciences, University at Buffalo - The State University of New York, Buffalo, NY 14260, USA. ²Department of Urban and Regional Planning, School of Architecture and Planning, University at Buffalo - The State University of New York, Buffalo, NY 14214, USA. ³School of City & Regional Planning, College of Design, Georgia Institute of Technology, Atlanta, GA 30332, USA. ✉email: sayantim@buffalo.edu

(YLDs) and 13.0% of disability-adjusted life-years (DALYs), globally¹². As of 2016, global estimates revealed that mental disorders (e.g., chronic depression, anxiety, substance use disorders) were significant contributors to disability in young adults; depressive and anxiety disorders were high among females, while substance use and autism spectrum disorders were high among males¹³. In the US, suicide ideation in adults is increasing, with 10.3 million adults diagnosed with severe thoughts of suicide¹⁴, and over two million youth having severe depression¹⁵. In fact, depression and hopelessness are the key predictors of suicide ideation and attempts in young adults⁶. Based on the commissioning report on mental health and the sustainable development goals (SDGs)¹⁶, mental health is considered a “global public good”, but both developed and developing countries struggle to understand and address the complex physical, social, and environmental influences that interact with genetic, neuro-developmental, and psychological processes driving the mental health and well-being of people¹⁶.

Although the built environment impacts mental health, there are gaps in the literature about the complex and non-linear relationships between mental health outcomes and the built environment. Studies examining the socio-economic determinants of mental health have shown that poverty, childhood adversity, and violence are the key risk factors of mental health disorders¹⁷. Studies have also indicated a disparity among large and medium/small metropolitan areas’ suicide rates with the latter being higher than the former⁷. Variations in population demographics and socio-economic factors such as unemployment rates, household income, and climate are the key factors associated with such disparities⁷. Other studies have also looked at the link between mental health and low quality of care for mental health disorders as well as human rights abuses¹⁸. However, there are gaps in the literature about the link between the various aspects of built environment and mental health outcomes. In what they delineate as the “neighborhood domain”, the commissioning report on the SDGs and mental health indicates that poorly planned or deteriorating neighborhoods (e.g., housing vacancy and declining quality of housing and community infrastructure) pose mental health challenges on individual-level biological markers¹⁶.

Various studies attempt to explain the link between population mental health and the built environment. For example, one study found that adolescents living in physically deteriorated neighborhoods had more health problems, including depression and anxiety than those living in ordered neighborhoods¹⁹. Another research project that studied 1355 residents in the New York City found that populations living in poor quality neighborhoods had a greater likelihood of experiencing chronic depression, after controlling for their income, race/ethnicity, age, and neighborhood-level income²⁰. A cross-sectional study of adults (16 years and above) residing in north London showed that the prevalence of depression had a statistically significant relationship with living in areas characterized by deck access homes (i.e., abundant with graffiti and without shared recreational spaces), after adjusting for individuals’ internal characteristics of their dwellings and socio-economic status²¹. A systematic review of 45 studies reveals that 37 reported at least one built environment characteristic associated with depression or depressive symptoms²².

However, despite some advancements in understanding the interplay between the built environment and mental health, there are limited methodological frameworks to parse the nonlinear relationships between the built environment and mental health outcomes²³, which this article aims to contribute. For instance, apart from genetic, lifestyle, and physio-psychological factors, mental health is influenced by a complex interplay of the physical and socio-economic aspects of the built environment (e.g., neighborhood decline, transportation conditions, unemployment, income, race, age, social capital, education). The complex, non-linear relationship between the built environment factors and mental health outcomes constrains the traditionally-used linear and static models’ explanatory and predictive abilities²⁴. Even though these complex interactions are acknowledged, studies are yet to leverage recent advances in big data analytics to explore such complexities.

In this study, we demonstrate a novel data-driven approach to study the complex associations between mental health of adults within a metropolitan community and the physical and socio-economic aspects of the built environment, thus guiding how properly planned neighborhoods may improve the overall mental health outcomes of the adult population within a community. Specifically, we develop and employ a novel methodological framework, a multi-level scenario-based predictive analytics framework (MSPAF), to explore the complex relationships between mental health outcomes and conditions in the built environment. The MSPAF combines rigorously validated interpretable machine learning algorithms and scenario-based sensitivity analysis to test several hypotheses on how the built environment affects mental health outcomes across the largest metropolitan areas in the US. The scenario-based analysis predicts how the community mental health outcomes in these metropolitan areas change under plausible perturbations of various built environment factors.

Results

Predictive performance of interpretable machine learning models and model selection. This study leveraged a library of supervised interpretable machine learning models to assess the associations between community mental health outcomes and, the built environment’s physical and socio-economic aspects. Interpretable machine learning models, ranging from parametric, semi-parametric to non-parametric models, vary widely in their degree of complexity, robustness, flexibility, and interpretability (discussed further in “[Overview of statistical learning](#)” section)^{25,26}. The statistical learning techniques are used in different research areas, such as energy demand modeling²⁷, infrastructure vulnerability assessment²⁸ or crime prediction². The parametric modeling technique (e.g., linear regression models), where a parametric function is fitted to the training data (e.g., via mechanisms such as least-squares), is the most popular modeling approach in healthcare research. Although such models are simple and easier to interpret, they often fail to approximate the true function since real relationships are often not linear. On the other hand, non-parametric data-driven models do not make any unrealistic assumptions about the functional form, thereby better approximating the true functional form. However, flexibility comes at the cost of interpretability²⁵. Although such data-driven non-parametric algorithms have seen their wide application in various domains of risk assessment such as crime risk modeling², energy

Model	R^2	In-sample		Out-of-sample	
		RMSE	MAE	RMSE	MAE
Generalized linear model	0.987	0.387	0.295	0.389	0.295
Ridge regression	0.971	0.584	0.457	0.588	0.459
Lasso regression	0.962	0.661	0.517	0.665	0.519
Generalized additive model	0.991	0.315	0.238	0.319	0.240
Multivariate adaptive regression splines [MARS]	0.983	0.441	0.328	0.442	0.328
MARS [degree=2]	0.982	0.454	0.343	0.456	0.344
MARS [degree=3]	0.982	0.454	0.342	0.456	0.343
MARS [degree=3; penalty=2]	0.981	0.454	0.361	0.456	0.343
Random forest	0.996	0.199	0.139	0.493	0.347
Gradient boosting method	0.994	0.261	0.197	0.309	0.282
Bayesian additive regression trees	0.997	0.182	0.136	0.221	0.159
Null	NA	3.382	2.773	3.386	2.774

Table 1. Model performance comparison.

supply inadequacy risk^{27,29–34}, infrastructure risk assessment³⁵, natural disaster risk assessment^{28,36,37}, among others, these methods are significantly under-explored in healthcare research despite their robustness and flexibility. To bridge this gap, this study assessed the predictive performance of eight interpretable machine learning models ranging from parametric to non-parametric—generalized linear model (GLM)³⁸, ridge regression (RR)³⁹, lasso regression (LR)⁴⁰, generalized additive model (GAM)⁴¹, multivariate adaptive regression splines (MARS)⁴², gradient boosting method⁴³, random forest (RF)⁴⁴, and Bayesian additive regression tree (BART)⁴⁵. Leveraging an 80–20 randomized percentage holdout cross-validation technique, we estimated the generalization performances of the models and selected the model that outperformed all the other models in terms of both in-sample goodness-of-fit and out-of-sample predictive accuracy (see “[Overview of statistical learning](#)” section). The model performances of the various models are depicted in the Table 1. Our results indicate that BART outperformed all the other models which is, thus, leveraged for the relevant statistical inferencing (see “[Key factors attributing to socio-economic and physical aspects of the built environment](#)” section).

Key factors attributing to socio-economic and physical aspects of the built environment. We leveraged the variable importance plot (VIP) (see Supplementary Information) and the partial dependence plots (PDPs) to identify the key built environment predictors of mental health outcomes, and evaluated their associated relationships (see “[Overview of statistical learning](#)” section for mathematical details of the VIP and PDP). For our analysis, we also controlled for behavioral and underlying health conditions (e.g., smoking habit, principal components of underlying physical health conditions) that significantly influence mental health outcomes. Since this study focuses on the built environment factors, our subsequent discussions will focus on the built environment’s physical and socio-economic aspects, which remain under-explored and are central to this article.

Partial dependence plots (PDPs) of the socio-economic aspects of the built environment, are depicted in Fig. 1. The PDP of poverty, shown in Fig. 1a, indicates a strong positive correlation with mental health outcomes. This relationship suggests that as the percentage of families below the poverty level increases from 10 to 80%, the percentage of adults (> 18 years) reporting poor mental health (mental health not good for ≥ 14 days) increases from 12.8 to 13.8% in the community on average. The narrow confidence interval (represented by the shaded grey area) indicates that the estimates are associated with less uncertainty. Other significant factors in this category include economic variables such as median family income and change in the unemployment rate (2005–2014). The partial dependence plot of median family income (Fig. 1b) shows a negative correlation. More specifically, we observe that as the median family income decreases from around \$130,000 to \$20,000, the percentage of adults reporting mental health disorders increases from 13.0 to 13.3% on average in a community. However, the wider confidence interval around the larger income values indicates that the estimated mental health outcomes for adults in the higher income range significantly vary. On the other hand, the relationship between unemployment changes and the percentage of adults reporting poor mental health is relatively uncertain Fig. 1c. Besides the economic status of a community, access to medical insurance plays a major role in predicting the community mental health outcomes. The PDP of the percentage of families with no health insurance (Fig. 1d) shows that it has a strong positive correlation with mental health outcomes. It is observed that as the percentage of families with no health insurance in a community increases from 5 to 35%, the percentage of adults reporting poor mental health increases from 12 to 14.5%. The narrow confidence interval indicates lower uncertainty and variations in the estimated relationship across the US metropolitans. Our results also suggest that the insurance type plays a major role in influencing mental health of adults within a metropolitan community. The PDP of insurance type, representing the ratio of percentage of families with public health insurance to private health insurance, is plotted in Fig. 1e. From the plot, we observe that as the proportion of families having public health insurance compared to that having private health insurance approximately doubles, the percentage of adults reporting poor mental health declines from 13.4 to 13.0% on average. The decreasing trend indicates that increased access to public health insurance is associated with decreased mental health disorders reported by adults in a metropolitan community on average.

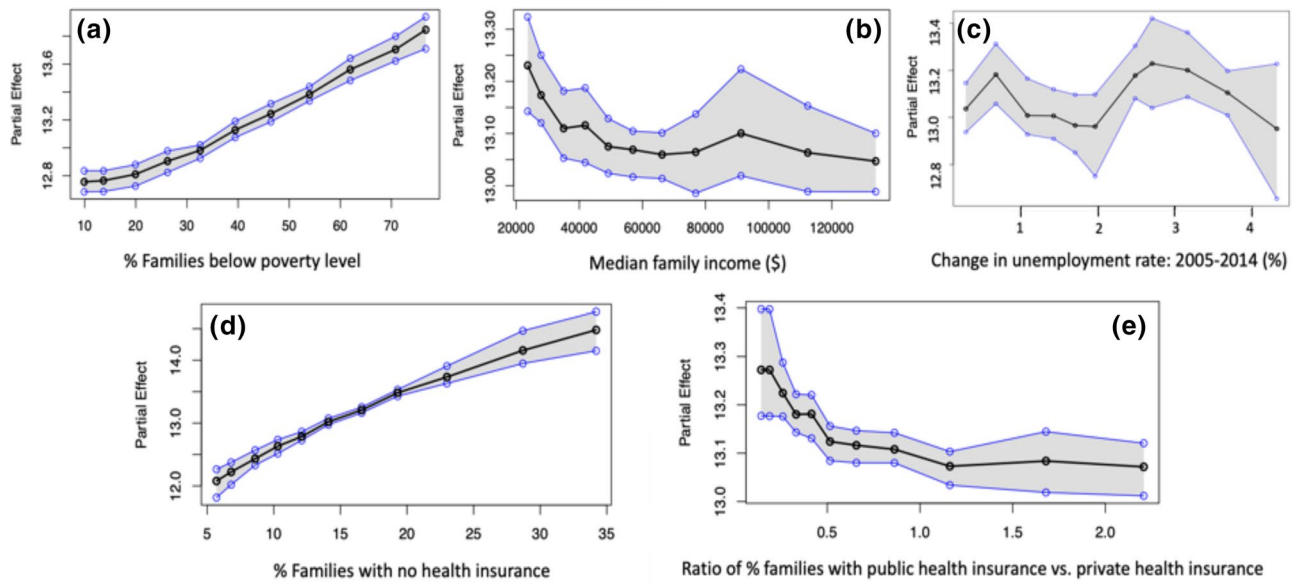


Figure 1. Partial Dependence Plots (PDPs) of the key socio-economic aspects of the built environment: (a) % of families below poverty level, (b) median family income, (c) % change in unemployment rate, (d) % families with no health insurance, (e) ratio of % of families with public health insurance vs. private health insurance. In each sub-figure, on the *x-axis*, the values of the particular independent variable is plotted, and on the *y-axis*, the partial effect of the independent variable on the response variable is depicted. The black curve is the average partial effect of the predictor variable and blue lines indicate the 95% confidence intervals.

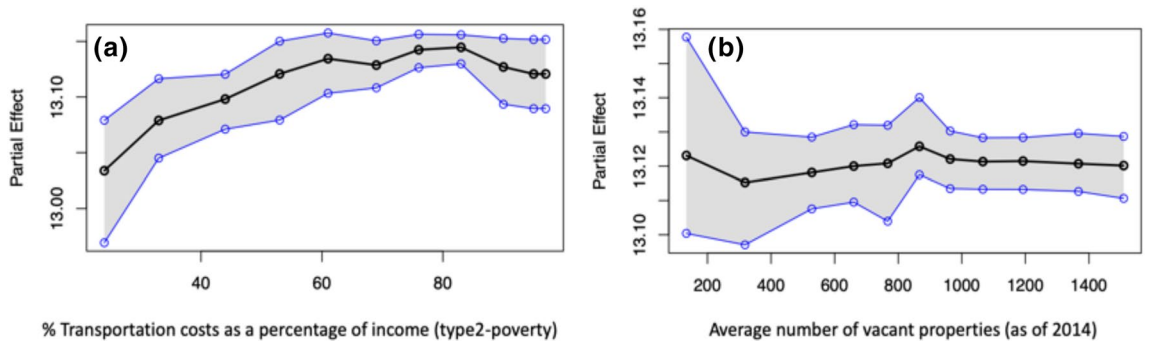


Figure 2. Partial Dependence Plots (PDPs) of the key physical aspects of the built environment: (a) % of the transportation cost as a percentage of income and (b) average number of vacant properties. In each sub-figure, on the *x-axis*, the values of the particular independent variable is plotted, and on the *y-axis*, the partial effect of the independent variable on the response variable is depicted. The black curve is the average partial effect of the predictor variable and blue lines indicate the 95% confidence intervals.

Our result shows that transportation or commuting cost (percentage of household income spent on transportation) and the average number of vacant properties, which constitute the built environment’s physical aspect, are the key predictors of community mental health. The PDP of transportation cost shows a positive correlation with poor mental health (Fig. 2a). More specifically, we observe that as household transportation expenditures increase from 20 to 100% on average, the percentage of adults reporting poor mental health increases from 13.0 to 13.10% on average. Although this increment seems small, it should be noted that these numbers only indicate the national average of large metropolitan communities, with some US states experiencing much higher negative impacts than others. Our scenario based sensitivity analysis (refer to “Data and methods” section) emphasizes such variations across the various metropolitan areas in the US states. However, the relationship between average number of vacant properties and community mental health is uncertain (Fig. 2b). We observe a slightly increasing trend in the percentage of adults reporting poor mental health as the average number of vacant properties increases. However, as the trend reaches the threshold point around 900 vacant properties in a community on average, the association flattens, that is, the number of reported mental health issues becomes insensitive to changes in vacant properties at a certain threshold level.

Independent variable category	Perturbation scenario and implications	Hypothesis index	Hypothesis statement
Economic characteristics	<i>Economic degradation</i> increase in % families below poverty level and unemployment rate by 1σ ; decrease in median household income by 1σ	H1	Economic degradation leads to degradation in community mental health or increase in K
	<i>Economic improvement</i> decrease in % families below poverty level and unemployment rate by 1σ ; increase in median household income by 1σ	H2	Economic improvement leads to improvement in community mental health or decrease in K
Unavailability of health insurance	<i>Less unavailability of health insurance</i> % families with no health insurance decreases by 1σ	H3	Less unavailability of health insurance leads to improvement in community mental health or decrease in K
	<i>More unavailability of health insurance</i> % families with no health insurance increases by 1σ	H4	More unavailability of health insurance leads to degradation in community mental health or increase in K
Category of health insurance	<i>Decreased access to public health insurance</i> Ratio of % families with public vs. private insurance decreases by 1σ	H5	Decreased access to public insurance leads to degradation in community mental health or increase in K
	<i>Increased access to public health insurance</i> Ratio of % families with public vs. private insurance increases by 1σ	H6	Increased access to public insurance leads to improvement in community mental health or decrease in K
Transportation cost	<i>Cheaper mode of travel and/or shorter commuting distance to work</i> annual transportation cost decreases by 1σ	H7	Decreased transportation cost leads to improvement in community mental health or decrease in K
	<i>Expensive mode of travel and/or longer commuting distance to work</i> annual transportation cost increases by 1σ	H8	Increased transportation cost leads to degradation in community mental health or increase in K
Average number of vacant properties	<i>People moving into the metropolitan areas (community expanding)</i> average number of vacant properties decreases by 1σ	H9	Decrease in vacancy would lead to improvement in community mental health or decrease in K
	<i>People moving out of the metropolitan areas (community shrinking)</i> average number of vacant properties increases by 1σ	H10	Increase in vacancy would lead to degradation in community mental health or increase in K

Table 2. Summary of the hypotheses.

Projected community mental health burden under plausible perturbations. Having identified the key built environment factors associated with mental health outcomes, we employed a scenario-based sensitivity analysis to understand how mental health outcomes may change under different built environment scenarios. Plausible future scenarios are captured through perturbations of the socio-economic and physical aspects of the built environment. Traditionally, in modern epidemiological studies, the sensitivity and uncertainty analyses for any disease burden and risk factor estimates are conducted using different weighting mechanisms and discount rate techniques⁴⁶. However, due to large degrees of uncertainties associated with value judgments and built environment conditions, the choice of discount rates is challenging and often cannot capture the wide range of future uncertainties⁴⁷. To overcome these challenges, we limited our analysis to statistical perturbation. The statistical perturbation consists of three significant steps described as follows: (1) we statistically perturbed the socio-economic and physical aspects of the built environment, which may lead to increase (e.g., economic growth) or decrease (e.g., economic recession) in the independent variable under consideration; (2) following a general intuition, we hypothesized whether the increase (decrease) in the independent variable leads to better (worse) mental health outcomes and vice-versa; and, (3) leveraging our predictive model, we verify if our hypothesis is valid nationally or only for certain US states (see “Data and methods” section for details on creating scenarios and list of hypotheses summarized in Table 2).

For illustration purpose, consider K represents the community mental health (response variable in our analysis), measured in terms of “% adults aged > 18 suffering from poor mental health for > 14 days”. Hence, *improvement in mental health* is depicted by a *decrease in K* , and *deterioration of community mental health* is observed when there is an *increase in K* . The predictor or independent variables under consideration for scenario-based sensitivity analysis are grouped into five categories, viz., (i) economic factors consisting of median household income, % of population below the poverty level and unemployment rate (ii) percentage of families with no health insurance, (iii) proportions of families having public insurance compared to private insurance, (iv) percentage of transportation cost spent as a % of household income, and (v) the average number of vacant properties (as of 2014). The first three categories of independent variables capture the socio-economic characteristics of the built environment and the last two categories represent the physical aspects of built environment. We assume, two hypothetical scenarios at a given time in our study: (1) the mean of the distributions of the socio-economic parameters (i.e., economic conditions, access to health insurance, and type of health insurance) and the physical aspects (i.e., travel cost and housing vacancy) of built environment of a community shifts by +1 standard deviation from their historical mean, which represents the base case or *as-is* scenario; and, (2) the mean of those distributions shift by −1 standard deviation from their historical mean. Note that, these statistical perturbations help to provide important insights regarding the trends of community mental health outcomes under plausible scenarios. However, it should be noted that our framework is generalized enough that it can be used to predict how the community mental health outcomes may change in the future, given the forecasted data on socio-economic aspects and built environment is available. Our study presents the framework illustrating how future mental health outcomes might be affected under various future scenarios. Furthermore, to understand whether such shifts results in a favorable outcome (improved mental health) or not, we compared the projections with the base case scenario of mental health outcomes by constructing ten hypotheses (see Table 2 in “Data and methods” section). Finally, we validated our hypotheses based on our model results and outcomes.

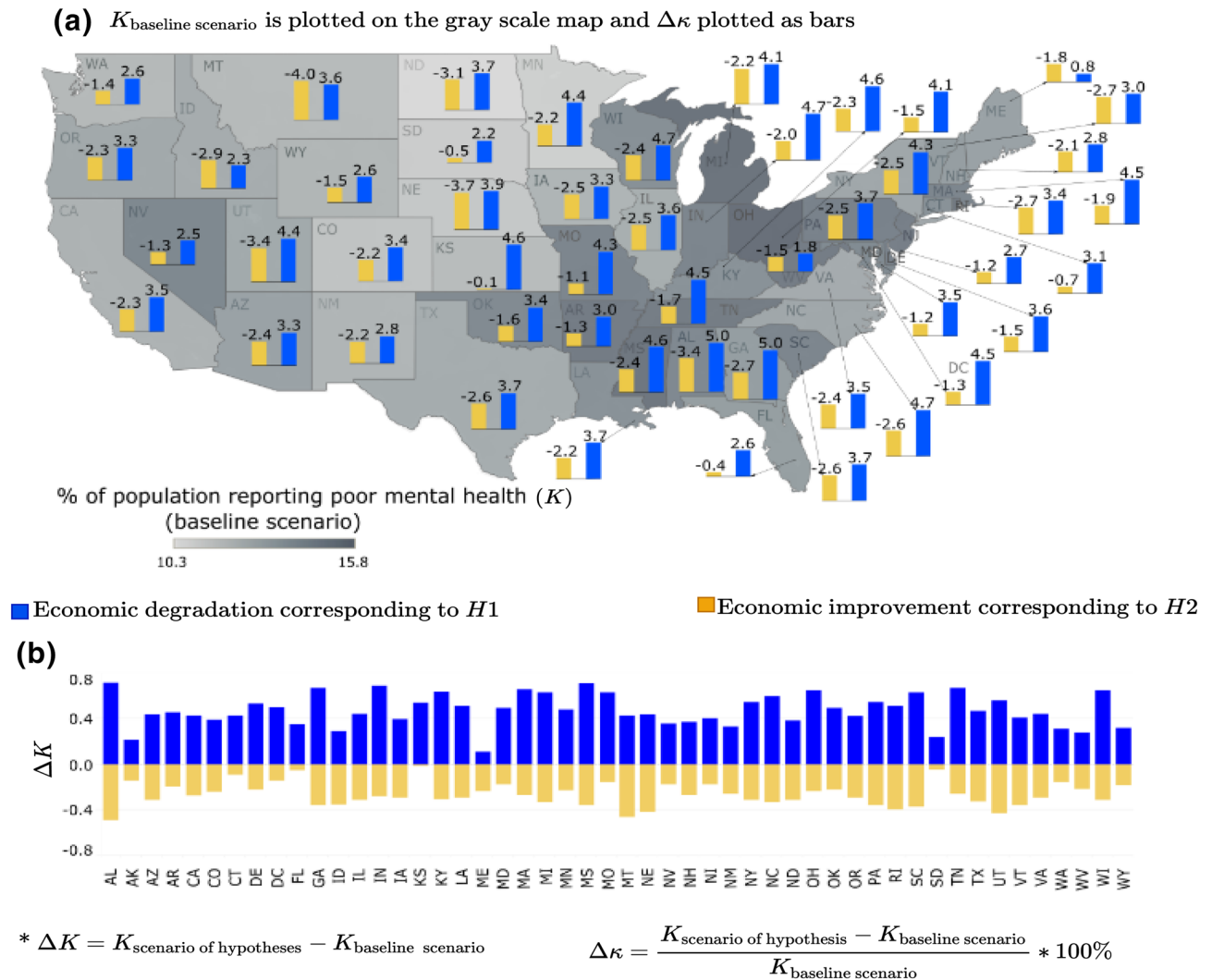


Figure 3. Economic condition scenario for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) ΔK is plotted. The maps are created using Tableau 2021.2 (Tableau 2021.2 New Features) and Adobe Illustrator (Software and services for creative business teams | Adobe Creative Cloud for teams).

The socio-economic aspects of built environment. Overall, our scenario-based sensitivity analysis indicates that the metropolitan areas in the eastern part of the US have poor mental health outcomes. We discuss the observed variations in the community mental health outcomes across the 50 states in the US under the six different scenarios of the socio-economic aspects of built environment—worst- and best-case scenarios of (a) economic condition, (b) lack of health insurance, and (c) access to public health insurance.

Economic condition. The economic conditions capture the interplay of poverty, median household income, and unemployment rate of the population in a metropolitan area. Since economic condition comprises three variables, for simplicity the hypothetical scenario is constructed by perturbing all the three variables simultaneously. It was hypothesized that during declining economic conditions, the expected percentage of people reporting poor mental health (K) would increase (hypothesis: $H1$), and the opposite effect would be observed during an increase in economic growth/boom (hypothesis: $H2$). The scenario-based analysis conducted herein supports these two hypotheses throughout all the states in the US. As depicted in Fig. 3, when economic depression sets in (blue bars), all the states observe a deterioration in community mental health depicted by $\Delta K > 0$. On the other hand, when the community experiences an economic boom (yellow bars), improvement in community mental health is observed depicted by $\Delta K < 0$. The scenario analysis, depicted in Fig. 3, shows that the percentage change (increase or decrease) in reported mental disorders among adults is more pronounced in metropolitan areas within states such as Alabama, Georgia, Indiana, Massachusetts, Kentucky, Michigan, Mississippi, Montana, Ohio, South Carolina, Tennessee, Utah, and Wisconsin. A recent systematic review identifies economic conditions as one of the social determinants of mental health⁴⁸. These conditions are linked to poverty^{16,48}, income⁴⁹, and unemployment⁵⁰. The scenario-based analysis confirms some of these earlier studies, but it also goes a step further to provide a metropolitan-level analysis of how improving or declining economic conditions

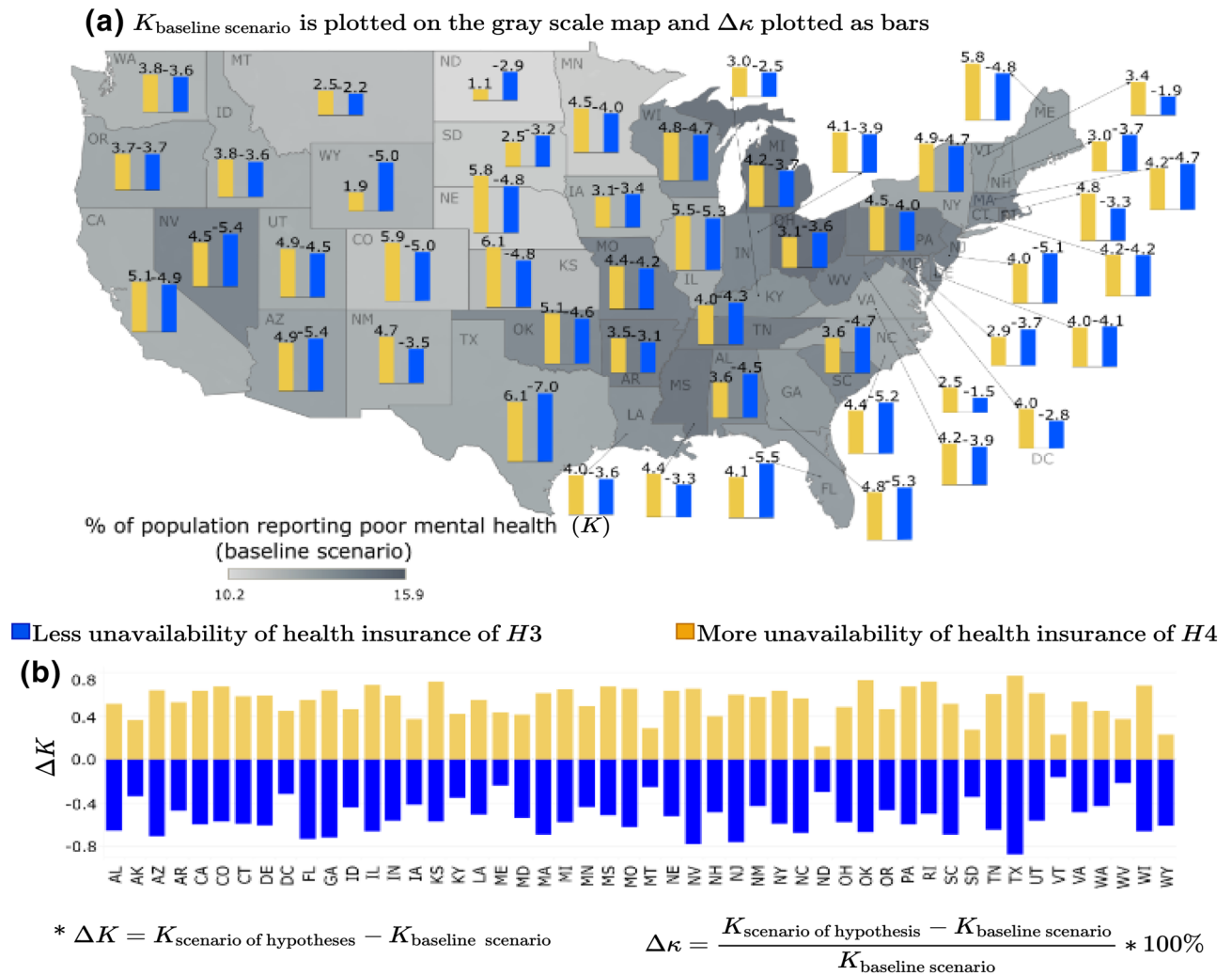


Figure 4. Unavailability of health insurance scenario for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) ΔK is plotted. The maps are created using Tableau 2021.2 (Tableau 2021.2 New Features) and Adobe Illustrator (Software and services for creative business teams | Adobe Creative Cloud for teams).

affect the mental health of adults in specific metropolitan areas in the US. Moreover, we also observe that community mental health is more sensitive to economic depression (longer blue bars for economic degradation) than economic boom (shorter yellow bars representing economic growth).

Unavailability of health insurance. In this case, the variable under consideration is the percentage of families with no health insurance or lack of access (unavailability) to health insurance. It was hypothesized that an overall improvement in community mental health would be observed when the unavailability of mental health will decrease, i.e., more families will have health insurance (hypothesis: $H3$). An opposite effect is expected with increased unavailability of health insurance, or in other words when, more families are being deprived of health insurance, which may lead to worsening mental health problems (hypothesis: $H4$). The scenario-based analysis in Fig. 4 suggests that these two hypotheses generally hold true for all the metropolitan areas across the US, considered in this study. From Fig. 4, we observe that when the percentage of families with no health insurance increases (yellow bars), the number of adults reporting poor mental health in the community (K) increases, compared to the baseline scenario. An opposite effect, i.e., decrease in the number of people reporting poor mental health is observed for the scenario depicting higher percentage of families in a community having health insurance (blue bars). However, a change (increase or decrease) in access to health insurance results in a minimum shift in the percentage of adults reporting mental disorders in the metropolitan areas of the states such as Montana, North Dakota, South Dakota, and Vermont. Studies show that states providing better access to mental health insurance minimize suicide rates⁵¹, but another study found that Australia’s mental health insurance under its “Better Access scheme” has had no significant effect on the mental health of Australians⁵². The underlying logic follows that increasing access to health insurance and, specifically, mental health insurance will likely increase the likelihood of more number of people accessing mental healthcare^{53,54}, which will ultimately improve

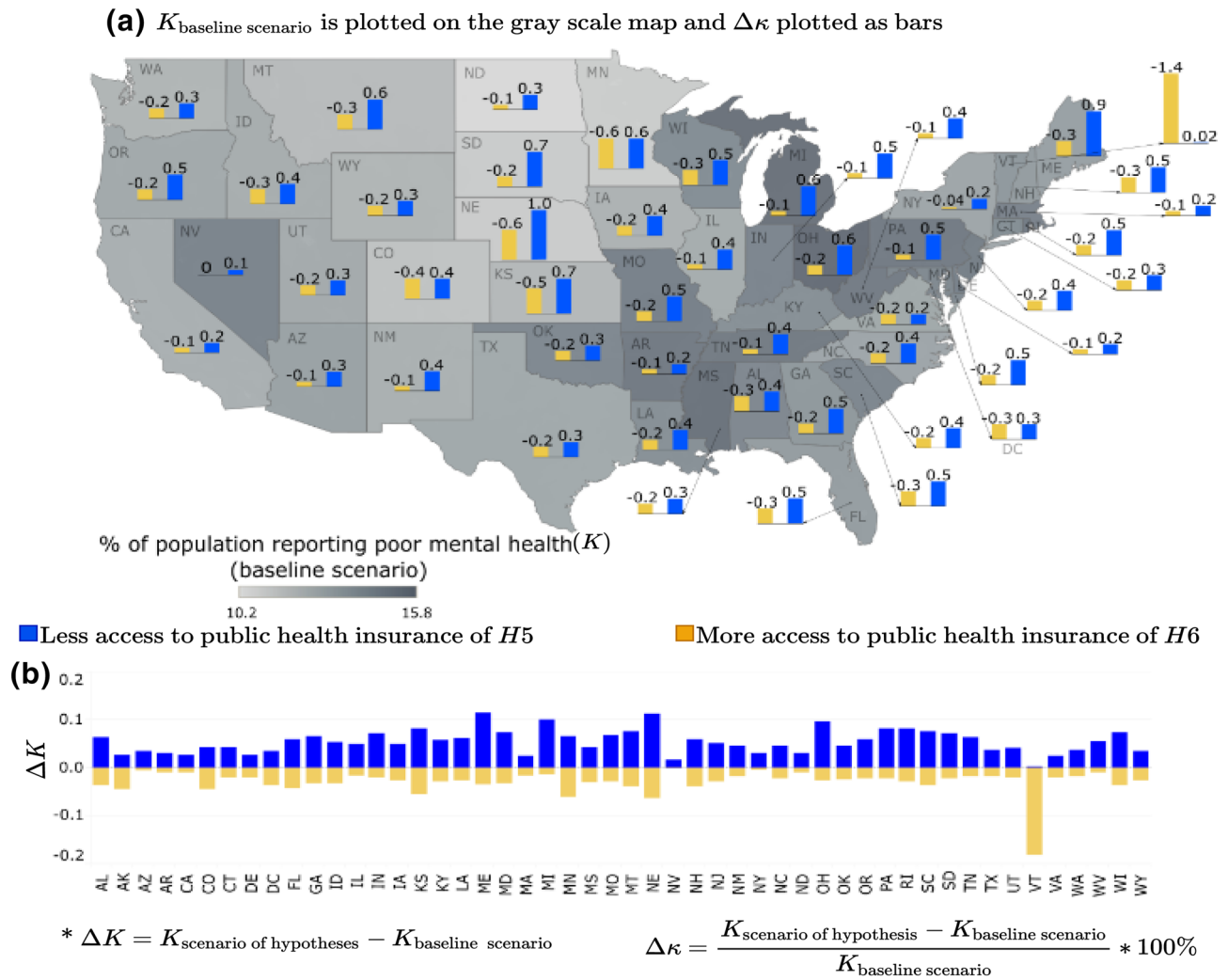


Figure 5. Access to public health insurance scenarios for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) ΔK is plotted. The maps are created using Tableau 2021.2 (Tableau 2021.2 New Features) and Adobe Illustrator (Software and services for creative business teams | Adobe Creative Cloud for teams).

overall mental health outcomes. The scenario-based analysis results contribute to this debate by explicitly looking at how the lack of access to health insurance in general, not only mental health insurance, may contribute to adults’ increasing stress and poor mental health outcomes.

Access to public health insurance. Building on the health insurance scenario analysis, we hypothesized prevalence of differential impacts of the two different types of health insurance, i.e., public vs. private health insurance on the community mental health outcomes. Specifically, we hypothesized that with decreased access to public health insurance (i.e., a lower proportion of people with access to public health insurance), the overall mental health of the community would worsen, leading to an increase in K (hypothesis: $H5$). The opposite effect of improving mental health would be observed with increased access to public health insurance (hypothesis: $H6$). However, these hypotheses were minimally supported in the scenario results across the states, as depicted in Fig. 5. Although the trend of increasing or decreasing mental health outcomes was found to be consistent across all the states having $\Delta K < 0$ for all the states when access to public health increases (yellow bars) and $\Delta K > 0$ for all the states when access to public health decreases (blue bars), the magnitude of such deviations significantly varies, ranging between -0.5% to $+1.0\%$. This varying range indicates that the overall mental health outcomes across the US’s metropolitan areas are not very sensitive to the type of health insurance. However, the hypothesis of decreasing K with increasing access to public health insurance (hypothesis: $H6$) was significantly supported for the metropolitan areas in Vermont. For context, Vermont was the first state in the US to adopt legislation for universal health care for its residents in 2011, making health insurance and healthcare publicly available to many residents, including free preventative services such as mental health and substance-based disorder services⁵⁵.

The physical aspect of the built environment. Travel/commuting cost. The scenario-based sensitivity analysis for travel cost—measured by the “% of transportation cost spent as a % of household income”—illustrates the extent to which the commuting cost within sprawling metropolitan areas can impact community mental health outcomes. The hypotheses explored here are—(1) the percentage of adults reporting mental disorders (K) would decrease with decreasing travel cost (hypothesis: $H7$) and, (2) the percentage of adults reporting mental disorders (K) would increase with increasing travel cost (hypothesis: $H8$). However, our analysis shows that these two hypotheses do not hold for some metropolitan areas in some of the states. For instance, the hypothesis ($H8$) that increased mental health disorders (K) are reported as the travel cost increases do not hold in metropolitan areas within the states such as Colorado, Delaware, Maine, Minnesota, Nebraska, North Dakota, Utah, Vermont, and Washington. There is a decrease in mental health disorders reported by adults in these metropolitan areas as the travel costs increase. For metropolitan areas in Washington DC, Maryland, and New Hampshire, an increase or decrease in travel costs has the same effect, i.e., an increase in the percentage of adults reporting mental health disorders. The decrease in mental health disorders as travel cost increases is generally consistent with findings in the literature. An increase in travel cost is often associated with sprawling areas, i.e., travel cost increases with sprawl^{56,57}. Some studies found that increasing sprawl (or commuting cost) either had no association with mental health disorders^{58,59} or was positively associated with better mental health, by allowing those living in low-density sprawl areas to enjoy proximity to nature^{60,61}. Some studies also found that shorter distances (decreased travel cost) to the city center positively influence subjective wellbeing^{62,63}. On the other hand, some studies found that increase sprawl or travel cost negatively impacts mental health, especially for residents living in auto-dependent sprawling neighborhoods with no access to personal vehicles⁶⁴. The mixed results of how travel costs impact mental health outcomes in our analysis resonates with the existing literature, and it signals the need for an in-depth and granular inquiry into how the built environment’s physical aspect impacts mental health outcomes in cities.

Housing vacancy. In the housing vacancy scenario, the hypotheses explored in this study investigated the extent to which neighborhood decline impacts mental health outcomes in the metropolitan areas across the 50 states in the US. Specifically, we hypothesized that a decrease in housing vacancy would lead to a decline in adults reporting poor mental health in metropolitan areas or K (hypothesis: $H9$). On the other hand, an increase in vacant properties or a decline in neighborhood size was expected to increase the percentage of adults reporting mental disorders (hypothesis: $H10$). However, overall, these hypotheses were not supported in our study. As depicted in Fig. 7, when the housing vacancy increased (yellow bars), most metropolitan areas across the US states experience an improvement or deterioration in the community mental health (K). On the other hand, when a community is expanding, attributed by decreased vacancy (blue bars), most of the metropolitan areas see an increase in the percentage of adults reporting poor mental health (K). This result may be an outcome of the “Behavioral Sink” phenomenon^{65,66}. However, for some states the reverse phenomenon has been observed. For instance, when the vacancy is decreasing, metropolitan areas of some states such as Alabama, Florida, Montana, New Mexico, North Carolina, Ohio, Oregon, Washington, and Wyoming see an improvement in mental health depicted by $\Delta K < 0$. On the other hand, when the vacancy is increasing, some states’ metropolitan areas (Arizona, Colorado, Nevada, and New Jersey) see a deterioration of mental health with $\Delta K > 0$. The mixed results from this scenario analysis support our earlier observations related to the transportation cost, emphasizing that there is more to the story when parsing the impacts of the built environment on community mental health outcomes. More granular-level analysis complemented by macro-level analyses might better help unpack how the built environment’s physical conditions at the household, neighborhood, city, and county levels may impact an individual’s mental health.

As discussed, the results depicting community mental health (K) sensitivity to housing vacancy are highly varied across the US states. Hence, it is difficult to classify whether a particular scenario of housing vacancy perturbation leads to the *best case scenario*, representing improvement in community mental health unanimously across the US states; or if the perturbation leads to the *worst-case scenario*, where the community mental health unanimously deteriorates across the nation. To address this, we aggregate the individual state-wide results into the mean value of the response variable K (for detailed results, see Supplementary Information). If the mean value of $\Delta K = K_{\text{scenario under consideration}} - K_{\text{base case scenario}}$ is (+)ve, then the perturbation scenario under consideration is depicted as the worst-case scenario. Similarly, if the mean ΔK is found to be (–)ve, then there is a decline in the percentage of the adults reporting mental health issues, so the scenario is termed as a best-case scenario.

Discussion

This study employs a library of supervised interpretable machine learning models and scenario-based sensitivity analyses to explore the relationship between adults’ mental health, and the socio-economic and physical aspects of the built environment in the US largest metropolitan areas. The interpretable machine learning models and scenario-based analyses elicit three essential issues for discussion and serve as crucial conversation points for policy discourses and future research.

First, the built environment’s socio-economic aspects are vital to understanding the social determinants of adults’ mental health in metropolitan communities across the US. The interpretable machine learning models suggest that increasing poverty and unemployment levels are associated with a significant increase in adults reporting mental health disorders. The scenario-based analysis supports this finding by showing that declining economic conditions within metropolitan areas are expected to increase the number of adults reporting mental disorders, and this is pronounced in metropolitan areas within states such as Georgia, Massachusetts, Kentucky, Michigan, Ohio, and Wisconsin. A number of studies have long observed the impact of poor economic conditions, manifesting in issues such as poverty, low-income, and unemployment, on mental health^{16,49,50,67}. This

paper provides evidence to support such existing findings across multiple metropolitan areas, and it allows for both within and across the states comparisons for policy conversations around how to center discussions on community mental health within economic policies at local, state, and national levels.

Second, the results from both the interpretable machine learning models and scenario-based analysis provide an opening to conversations around health insurance and mental health. The debate in the literature focuses on whether or not access to mental health insurance schemes improves the likelihood of a person accessing mental health services, which leads to improved mental health outcomes. While the evidence seems inconclusive based on contradictory studies across countries^{51,52,68}, the partial dependence analysis of the health insurance variables in our study show that there is a strong increasing trend between lack of health insurance and adults reporting mental health disorders in metropolitan areas across the states in the US. This analysis goes a step further to show that decreased access to public health insurance is linked to increased mental disorders reported. The scenario-based analysis showed Vermont, the first state to adopt universal healthcare, as an outlier case. Increased access to public health insurance was linked to a significant decrease in mental health disorders reported within Vermont's largest metropolitan area. This finding does not necessarily suggest the need for universal healthcare. At the very least, it calls for an in-depth research inquiry and policy discourses around how the lack of health insurance, a critical socio-economic need, can impact a person's mental health.

Finally, the physical aspects of the built environment are found to have mixed impacts on community mental health. Adults report increased mental health disorders as travel costs increase in some metropolitan areas, but this does not hold across all the metropolitan areas in our data sample. Similarly, mental health disorders reported increased as housing vacancy increased in some metropolitan areas, but this also does not hold in all metropolitan areas. The commissioning about the SDGs and mental health rightly observes the need to understand "how neighborhood domain" impact the community mental health. Specifically, it indicates that, besides biological markers, the decline in neighborhood conditions should also be considered as one of the important social determinants of mental health¹⁶. In conclusion, this article adds to existing studies on how the built environment impacts mental health outcomes⁶⁹, supporting concerns raised in the commissioning report. More importantly, it also adds to the literature on how urbanization (e.g., increasing sprawl and associated commuting costs) impacts mood disorders⁷⁰. The mixed results call for caution when discussing how the built environment's physical aspects impact community mental health. Future studies may incorporate other physical properties of the built environment such as street density, street connectivity, and land use mix into our proposed multi-level scenario-based predictive analytics framework (MSPAF), to further examine the relationship between the built environment and the community mental health outcomes. Although this article focuses on the large metropolitan areas at a national scale, micro-level data can be collected to explore at a more granular-level, such as intra- and inter-urban and rural dynamics in terms of built environment conditions and mental health outcomes. Future studies may focus more on studying the dynamics within and across urban and rural areas, which remains vital to developing context-specific urban planning, public health and public policy interventions to improve built environment and mental health outcomes.

Data and methods

Data collection and pre-processing. In this study, we conducted a nation-level study for all the metropolitan regions in 50 states across the US. We obtained and aggregated data for public health characteristics, built environment features, and socio-economic conditions from multiple sources. From the US Centers for Disease Control and Prevention (CDC) Behavioral Risk Factor Surveillance System (BRFSS), information about the health-related variables like, mental health conditions, pre-clinical conditions and behavioral factors for the adults aged 18 or above are collected at a census tract level for the year 2014⁷¹. The housing vacancy data for the year 2014 is obtained from US Housing and Urban Development (HUD) at a census tract level⁷². Finally, the socio-economic characteristics like, race, income, unemployment rate, marital status, education level, and access to health insurance information are obtained for the census tract and metropolitan levels from the American Community Survey (ACS) for the years 2011 to 2015⁷³. The travel cost data is obtained from the US Department of Housing and Urban Development Low Transportation Cost Index (LAI)⁷², which uses data on housing costs from the American Community Survey (ACS) and estimates transportation costs based on land use mix, commute patterns, and socio-economic information. The data from the multiple sources are matched and aggregated to create the final data set. In our analysis, *the percentage of participants who were adults aged 18 years or more and reported that they were suffering from mental health issues for more than 14 days in the last month* is considered as the response variable. The other variables on health characteristics, built environment features and socio-economic characteristics are considered as the predictors or independent variables. Out of all the categories of the predictor variables, the pre-clinical health condition related variables are found to be highly correlated. To consider the effect of all the pre-clinical variables while having a bound on the number of dimensions, we performed principal component analysis (PCA) (see Supplementary Information). PCA is an unsupervised learning method that uses orthogonal transformations to convert a multidimensional data set of observations of possibly correlated variables into a new multidimensional data set of values of linearly uncorrelated variables⁷⁴. PCA is useful for dimension reduction purpose, because a fewer orthogonal components of the transformed data can capture most of the variance of the original data. In this research, we considered three principal components as they were able to express 92% variability of the observations of the original 12 pre-clinical health related variables taken into consideration.

Overview of statistical learning. Given a dataset with a response variable Y and a set of p predictor variables $X = X_1, X_2, \dots, X_p$, interpretable machine learning algorithms try to identify the function f that relates the predictors with the response variable as, $Y = f(X) + \epsilon$ ²⁶. Here, ϵ is the irreducible error term that arises

from unobserved heterogeneity from the data and is normally distributed $N(\mu, \sigma^2)$ where, μ = mean and σ^2 = variance²⁵. Using the *training data* which is a known set of data points, a model is trained to estimate f and using an unknown set of data points known as *test data*, the performance of the model is evaluated. In this study, we implemented a suite of interpretable machine learning models, which can be crudely classified into three categories, viz. (i) parametric models, (ii) semi-parametric models and (iii) non-parametric models. In parametric models, the problem of estimating the unknown function f gets reduced to estimating a set of parameters through which the model is represented. On the other hand, the non parametric models make no assumption about the unknown function. A semi-parametric model is a hybrid of parametric and non-parametric models. More specifically, we implemented the following algorithms—

1. *Parametric models* Generalized Linear model³⁸, Ridge regression³⁹ and Lasso regression⁴⁰
2. *Semi-parametric models* Generalized additive model⁴¹, multi adaptive regression splines⁴²,
3. *Non-parametric models* Random forest⁴⁴ and gradient boosting method⁴³ Bayesian additive regression trees⁴⁵

To achieve optimal generalization performance for an interpretable machine learning model, it's complexity should be controlled using the bias-variance trade off technique. Cross validation is the most widely used technique for balancing models' bias and variance. In this study, the best model was selected using an 80–20 randomized percentage holdout cross validation technique, where the models were trained on randomly selected 80% of the data set and the remaining 20% of the data set were used as holdout set to assess the out-of-sample predictive performance of the models. This technique is repeated 30 times to ensure each data point of the original data set is used at least once for training the models. The metrics used to compare the performances of the models are R^2 , $RMSE$ (root mean squared error) and MAE (mean absolute error). This method of model selection is a well-established method and has been used in various previous studies^{7,27–30,32–34}. In the following section, we described the Bayesian additive regression trees, which is the best model found in our analysis, and leave the discussion on other methods in the Supplementary Information.

Bayesian additive regression trees. Bayesian additive regression tree (BART) is a sum-of-trees model where the outputs from m 'small' decision trees are aggregated with an underlying Bayesian probability model to generate the response function^{45,75}. Mathematically, BART can be expressed as,

$$Y = \left[\sum_{j=1}^m g(X; T_j, M_j) \right] + \epsilon \quad \epsilon \sim N(0, \sigma^2). \quad (1)$$

There are m distinct regression trees T_j with their terminal node parameters M_j . The function $g(X; T_j, M_j)$ assigns the leaf node parameters M of tree T to the independent variables X for all m trees. The main difference of BART compared to other tree ensemble methods is that, BART develops on an underlying Bayesian probability model and consists of a prior, likelihood and posterior probability space. The prior terms are responsible for the tree structure, model complexity, regularization and incorporating expert knowledge in the model. Generally, the Metropolis–Hastings algorithm is used to generate draws from the posterior probability space.

Model inference. Although the non-parametric models outperform parametric models in terms of predictive performance, the improved predictability comes at the cost of reduced interpretability. However, statistical inferencing can be conducted for the non parametric models using the variable importance ranking and partial dependence plots (PDPs)^{26,45,75}. The importance of the variables are depicted by the inclusion proportion of the variables which denote the number of times a particular variable has been selected to develop the model. To understand how a particular predictor variable affects the response variable, the PDPs are used. The PDP is estimated as follows:

$$p_j(x_j) = \frac{1}{n} \sum_{i=1}^n p_j(x_j, x_{-j}, i). \quad (2)$$

Here, p is the statistical response surface; n denotes the number of observations, x_{-j} represents all the independent variables except x_j .

Scenario-based sensitivity analysis. The scenario-based sensitivity analysis implemented in this study involves a systematic approach of statistical simulation. First, the independent variable or the set of independent variables for which the scenario is to be created are selected. For each state, the best parametric distribution that fits the sample data of independent variables (predictors) is identified using the Chi-squared goodness of fit and method of moments for parameter estimation⁷⁶. After the best distribution(s) of the predictor(s) is identified, for each state random sampling is implemented to obtain the base case values (BV). Then, according to the hypothesized scenario, the mean of the historical parametric distribution of the variable of interest is perturbed. Then, using random sampling, new values are obtained from the new distribution with the shifted mean, which corresponds to the hypothesized scenario. The original values of the variable are then substituted by the new values corresponding to the scenario while keeping all the other variables same as original. Following this, using the selected statistical learning model, the percentage of population reporting poor mental health are predicted for the new data set. Finally, we identify whether any significant nation-level and/or state-level increase or decrease in the response (compared to the original response variable) is observed or not.

(a) $K_{\text{baseline scenario}}$ is plotted on the gray scale map and $\Delta\kappa$ plotted as bars

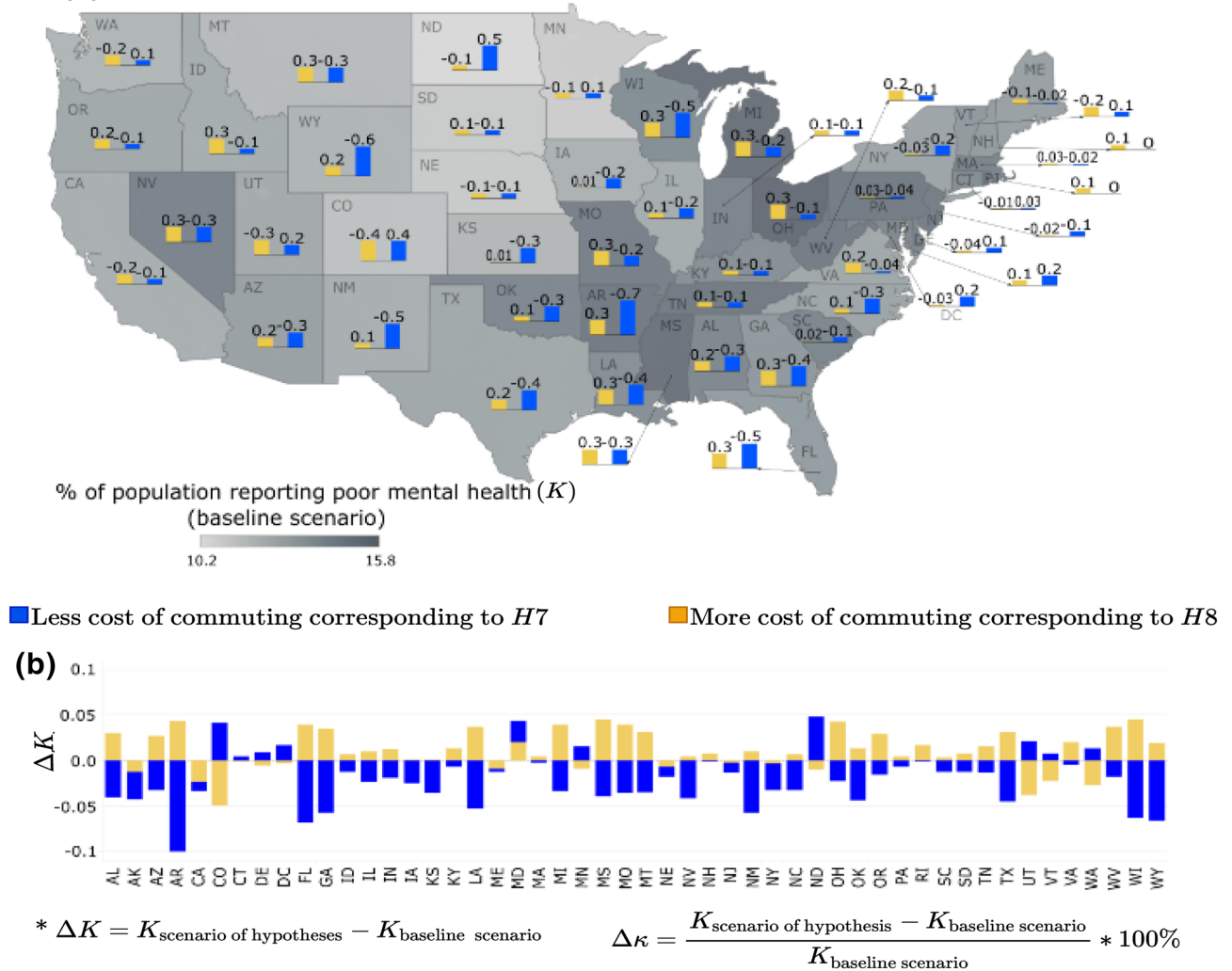


Figure 6. Travel cost scenarios for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) ΔK is plotted. The maps are created using Tableau 2021.2 (Tableau 2021.2 New Features) and Adobe Illustrator (Software and services for creative business teams | Adobe Creative Cloud for teams).

As described before, in this paper, we considered five categories of variables representing socio-economic and physical aspects of a built environment: (i) the economic status of a community characterized by incidence of poverty, unemployment rate and household income, (ii) % of families in a community with no health insurance, (iii) access to public health insurance, (iv) transport cost expressed as a % of income spent towards transportation, and (v) housing vacancy. The mean of each variable’s historical distribution is perturbed 1σ (standard deviation) of the variable. Corresponding to these sets of variables, ten hypotheses are created (see Table 2).

For each category of the independent variables, we validate our hypotheses by predicting $K_{\text{scenario of hypothesis}}$ denoting the “% adults aged > 18 suffering from poor mental health for > 14 days” under the specific scenario of independent variable perturbation (e.g., economic depression) considered for a particular hypothesis (e.g., H1). The change in the response corresponding to this perturbed condition is captured by,

$$\Delta K = K_{\text{scenario of hypothesis}} - K_{\text{base case scenario}}$$

To normalize the effect of the base line response value, we consider $\Delta\kappa$ which captures the projected change in % of adults aged > 18 years reporting poor mental health for > 14 days and expressed as a percentage of the baseline estimates.

$$\Delta\kappa = \frac{K_{\text{scenario of hypothesis}} - K_{\text{base case scenario}}}{K_{\text{base case scenario}}} \times 100\%$$

In Figs. 3, 4, 5, 6 and 7, the output of the sensitivity analysis has been depicted. The ΔK is plotted in part (b) of each figure, representing the exact projected change in K . For each figure, in part (a), the $\Delta\kappa$ is plotted as the

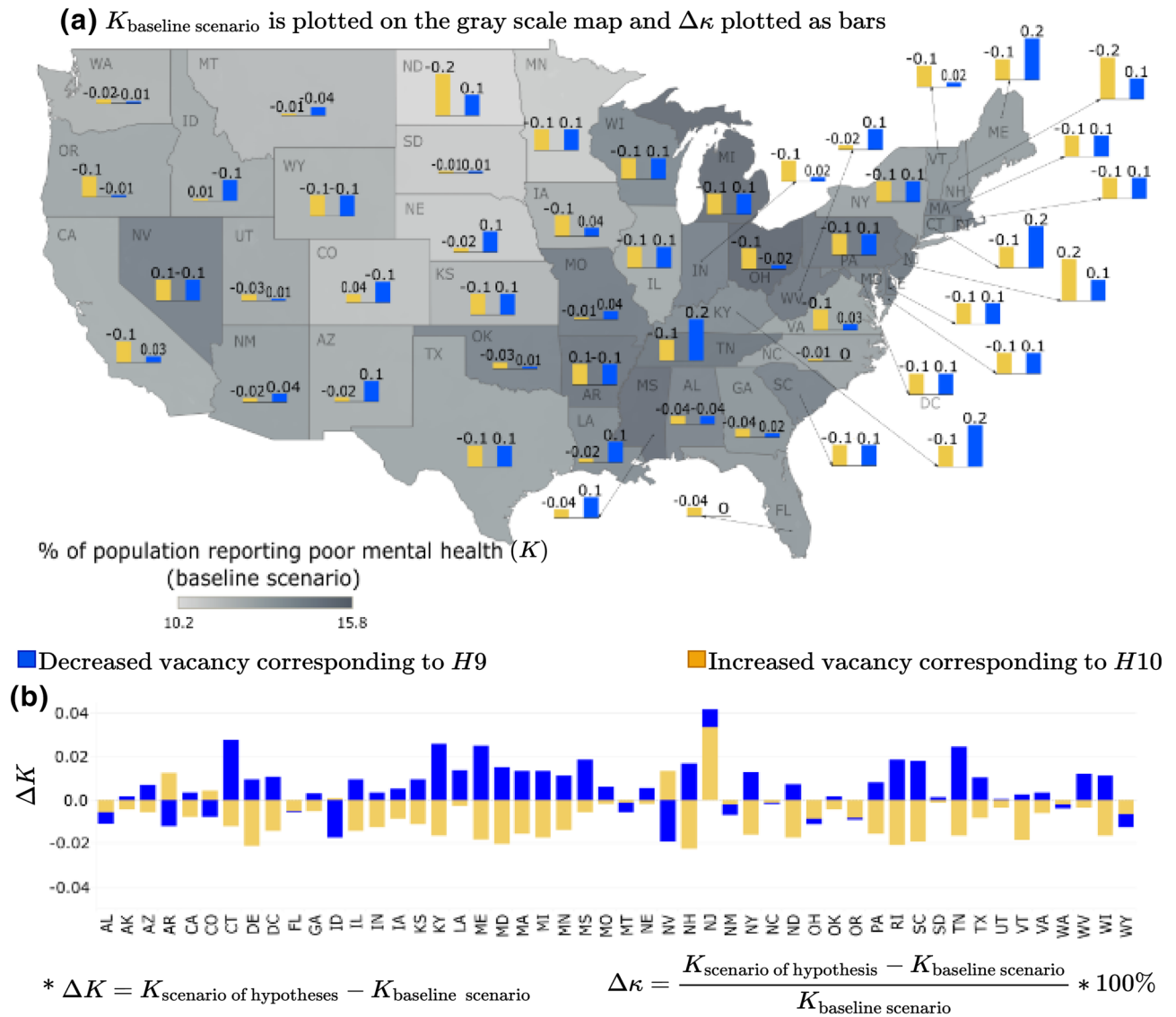


Figure 7. Housing vacancy scenario for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) ΔK is plotted. The maps are created using Tableau 2021.2 (Tableau 2021.2 New Features) and Adobe Illustrator (Software and services for creative business teams | Adobe Creative Cloud for teams).

bars representing the projected change expressed as a percentage of the baseline estimate with the underlying $K_{\text{base case scenario}}$ depicted in the map as gray scale intensities. In the subsequent sections, we discuss the result of the sensitivity of K to different categories of independent variables.

Received: 27 April 2021; Accepted: 30 July 2021
 Published online: 02 September 2021

References

- James, P., Troped, P. & Laden, F. The impact of the built environment on health. In *Women and Health* 2nd edn (eds Goldman, M. et al.) 753–763 (Academic Press, 2013).
- Ganguly, P. & Mukherjee, S. A multifaceted risk assessment approach using statistical learning to evaluate socio-environmental factors associated with regional felony and misdemeanor rates. *Physica A* 574, 125984. <https://doi.org/10.1016/j.physa.2021.125984> (2021).
- Raleigh, E. & Galster, G. Neighborhood disinvestment, abandonment, and crime dynamics. *J. Urban Aff.* 37, 367–396 (2014).
- Houle, J. N. Mental health in the foreclosure crisis. *Soc. Sci. Med.* 118, 1–8 (2014).
- Lindblad, M. R. & Riley, S. F. Loan modifications and foreclosure sales during the financial crisis: Consequences for health and stress. *Hous. Stud.* 30(7), 1092–1115 (2015).
- Wei, Z. & Mukherjee, S. Health-behaviors associated with the growing risk of adolescent suicide attempts: A data-driven cross-sectional study. *Am. J. Health Promot.* <https://doi.org/10.1177/0890117120977378> (2020).

7. Mukherjee, S. & Wei, Z. Suicide disparities across urban and suburban areas in the U.S.: A comparative assessment of socio-environmental factors using a data-driven predictive approach. Preprint at <http://arXiv.org/2011.08171> (2020).
8. Kan, Z. *et al.* Identifying the space-time patterns of COVID-19 risk and their associations with different built environment features in Hong Kong. *Sci. Total Environ.* **772**, 145379. <https://doi.org/10.1016/j.scitotenv.2021.145379> (2021).
9. Li, S., Ma, S. & Zhang, J. Association of built environment attributes with the spread of COVID-19 at its initial stage in China. *Sustain. Cities Soc.* **67**, 102752 (2021).
10. Carraturo, F. *et al.* Persistence of SARS-CoV-2 in the environment and COVID-19 transmission risk from environmental matrices and surfaces. *Environ. Pollut.* **265**, 115010. <https://doi.org/10.1016/j.envpol.2020.115010> (2020).
11. Khan, N., Khan, A. & Ahmed, S. COVID-19 transmission, vulnerability, persistence and nanotherapy: A review. *Environ. Chem. Lett.* **1**, 27–35. <https://doi.org/10.1007/s10311-021-01229-4> (2021).
12. Vigo, D., Thornicroft, G. & Atun, R. Estimating the true global burden of mental illness. *Lancet Psychiatry* **3**(2), 171–178 (2016).
13. Vos, T. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **390**, 1211–1259 (2017).
14. Nguyen, T. *et al.* *The State of Mental Health in America 2018* (Mental Health America, 2017).
15. Enos, G. Annual report raises red flags on youth access to care. *Ment. Health Wkly.* **29**, 3–4 (2019).
16. Patel, V. *et al.* The Lancet Commission on global mental health and sustainable development. *The Lancet* **392**, 1553–1598 (2018).
17. Lund, C. *et al.* Poverty and common mental disorders in low and middle income countries: A systematic review. *Soc. Sci. Med.* **71**(3), 517–528 (2010).
18. Patel, V., Kleinman, A. & Saraceno, B. Protecting the human rights of people with mental illnesses: A call to action for global mental health. In *Mental Health and Human Rights: Vision, Praxis* (eds Courage, M. *et al.*) 362–375 (Oxford University Press, 2012).
19. Aneshensel, C. & Mcneely, C. The neighborhood context of adolescent mental health. *J. Health Soc. Behav.* **37**, 293–310. <https://doi.org/10.2307/2137258> (1997).
20. Galea, S. *et al.* Urban built environment and depression: A multilevel analysis. *J. Epidemiol. Community Health* **59**(10), 822–827 (2005).
21. Weich, S. *et al.* Mental health and the built environment: Cross-sectional survey of individual and contextual risk factors for depression. *Br. J. Psychiatry J. Ment. Sci.* **180**, 428–433. <https://doi.org/10.1192/bjp.180.5.428> (2002).
22. Sallis, J. *et al.* Neighborhood built environment and income: Examining multiple health outcomes. *Soc. Sci. Med.* **68**, 1285–93 (2009).
23. Halpern, D. *Mental Health and the Built Environment More than Bricks and Mortar?* (Routledge, 2014).
24. Helbich, M. *et al.* More green space is related to less antidepressant prescription rates in the Netherlands: A Bayesian geoadaptive quantile regression approach. *Environ. Res.* **166**, 290–297 (2018).
25. James, G. *et al.* *An Introduction to Statistical Learning with Applications in R* (Springer, 2015).
26. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009).
27. Mukherjee, S. & Nateghi, R. Climate sensitivity of end-use electricity consumption in the built environment: An application to the state of Florida, United States. *Energy* **128**, 688–700 (2017).
28. Mukherjee, S., Nateghi, R. & Hastak, M. A multi-hazard approach to assess severe weather-induced major power outage risks in the us. *Reliab. Eng. Syst. Saf.* **175**, 283–305 (2018).
29. Alipour, P., Mukherjee, S. & Nateghi, R. Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the Texas region. *Energy* **185**, 1143–1153 (2019).
30. Mukherjee, S. & Nateghi, R. A data-driven approach to assessing supply inadequacy risks due to climate-induced shifts in electricity demand. *Risk Anal.* **39**(3), 673–694 (2019).
31. Mukherjee, S. & Nateghi, R. Estimating climate-demand nexus to support long-term adequacy planning in the energy sector. In *Power & Energy Society General Meeting, 2017 IEEE, IEEE*, 10–1109 (2018).
32. Mukherjee, S., Vineeth, C. R. & Nateghi, R. Evaluating regional climate-electricity demand nexus: A composite Bayesian predictive framework. *Appl. Energy* **235**(2019), 1561–1582 (2019).
33. Nateghi, R. & Mukherjee, S. A multi-paradigm framework to assess the impacts of climate change on end-use energy demand. *PLoS ONE* **12**(11), e0188033 (2017).
34. Obringer, R., Mukherjee, S. & Nateghi, R. Evaluating the climate sensitivity of coupled electricity-natural gas demand using a multivariate framework. *Appl. Energy* **262**, 114419 (2020).
35. Fontecha, J. E. *et al.* A two-stage data-driven spatiotemporal analysis to predict failure-risk of urban sewer systems leveraging machine learning algorithms. *Risk Anal.* <https://doi.org/10.1111/risa.13742> (2021).
36. Masoudvaziri, N. *et al.* Integrated risk-informed decision framework to minimize wildfire-induced power outage risks: A county-level spatiotemporal analysis. In *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference, Venice Italy* (2020). <https://www.rpsonline.com.sg/proceedings/esrel2020/html/4243.xml>.
37. Masoudvaziri, N. *et al.* Impact of geophysical and anthropogenic factors on wildfire size: A spatiotemporal data-driven risk assessment approach using statistical learning. *Stoch. Environ. Res. Risk Assess.* <https://doi.org/10.21203/rs.3.rs-539684/v1> (2021).
38. Nelder, J. A. & Wedderburn, R. W. M. Generalized linear models. *J. Stat. Soc.* **135**(3), 370–384 (1972).
39. Tikhonov, A. N. Solution of incorrectly formulated problems and the regularization method. *Sov. Math.* **4**, 1035–1038 (1963).
40. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc.* **58**(1), 267–288 (1996).
41. Hastie, T. & Tibshirani, R. Generalized additive models. *Stat. Sci.* **1**, 297–310 (1986).
42. Friedman, J. H. Multivariate adaptive regression splines. *Ann. Stat.* **19**, 1–67 (1991).
43. Breiman, L. Arcing the edge. In *Technical Report Statistics Department*, Vol. 486 (University of California, 1997).
44. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
45. Hugh, A. C., George, E. I. & McCulloch, R. E. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4**(1), 266–298 (2010).
46. Mathers, C. D. *et al.* Sensitivity and uncertainty analyses for burden of disease and risk factor estimates. In *Global Burden of Disease and Risk Factors* (The International Bank for Reconstruction and Development/The World Bank, Washington, DC, 2006).
47. Morgan, M. G., Henrion, M. & Small, M. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (Cambridge University Press, 1990).
48. Lund, C., Baron, E. & Breuer, E. Social determinants of mental disorders and the sustainable development goals: A systematic review of reviews. *Lancet Psychiatry* **5**, 357–369 (2018).
49. Roddy, M., Rhoades, G. & Doss, B. Effects of ePREP and our relationship on low-income couples' mental health and health behaviors: A randomized controlled trial. *Prev. Sci.* <https://doi.org/10.1007/s1121-020-01100-y> (2020).
50. Paul, K. & Moser, K. Unemployment impairs mental health: Meta-analyses. *J. Vocat. Behav.* **74**, 264–282. <https://doi.org/10.1016/j.jvb.2009.01.001> (2009).
51. Lang, M. The impact of mental health insurance laws on state suicide rates. *Health Econ.* **22**(1), 73–88. <https://doi.org/10.1002/hec.1816> (2013).
52. Jorm, A. Australia's 'Better Access' scheme: Has it had an impact on population mental health? *Austral. N. Z. J. Psychiatry* **52**, 000486741880406. <https://doi.org/10.1177/0004867418804066> (2018).

53. McMorro, S. *et al.* Medicaid expansions from 1997 to 2009 increased coverage and improved access and mental health outcomes for low-income parents. *Health Serv. Res.* **51**(4), 1347–1367. <https://doi.org/10.1111/1475-6773.12432> (1997).
54. Lang, M. The impact of mental health insurance laws on state suicide rates. *Health Econom.* **22**, 73–88. <https://doi.org/10.1002/hec.1816> (2013).
55. MacNaughton, G. *et al.* The impact of human rights on universalizing health care in Vermont, USA. *Health Hum. Rights* **17**, E83–E95 (2015).
56. Sultana, S. & Weber, J. Journey-to-work patterns in the age of sprawl: Evidence from two midsize southern metropolitan areas. *Prof. Geogr.* **59**(2), 193–208. <https://doi.org/10.1111/j.1467-9272.2007.00607.x> (2007).
57. Glaeser, E. L. & Kahn, M. E. Chapter 56—Sprawl and urban growth. In *Cities and Geography. Handbook of Regional and Urban Economics* Vol. 4 (eds Vernon Henderson, J. & Thisse, J.-F.) 2481–2527 (Elsevier, 2004). [https://doi.org/10.1016/S1574-0080\(04\)80013-0](https://doi.org/10.1016/S1574-0080(04)80013-0).
58. Sturm, R. & Cohen, D. A. Suburban sprawl and physical and mental health. *Public Health* **118**, 488–496. <https://doi.org/10.1016/j.puhe.2004.02.007> (2004).
59. Garrido-Cumbrera, M. *et al.* 2054—The effects of urban sprawl on mental health: A study of a municipality in the seville metropolitan area (Spain). *J. Transp. Health* **5**, S77–S78. <https://doi.org/10.1016/j.jth.2017.05.228> (2017).
60. Garrido-Cumbrera, M. *et al.* Exploring the association between urban sprawl and mental health. *J. Transp. Health* **10**, 381–390. <https://doi.org/10.1016/j.jth.2018.06.006> (2018).
61. Freeman, L. The effects of sprawl on neighborhood social ties: An explanatory analysis. *J. Am. Plann. Assoc.* **67**, 69–77. <https://doi.org/10.1080/01944360108976356> (2001).
62. Wang, F. & Wang, D. Geography of urban life satisfaction: An empirical study of Beijing. *Travel Behav. Soc.* **5**, 14–22. <https://doi.org/10.1016/j.tbs.2015.10.001> (2016).
63. Mouratidis, K. Compact city, urban sprawl, and subjective well-being. *Cities* **92**, 261–272. <https://doi.org/10.1016/j.cities.2019.04.013> (2019).
64. Dong, H. & Qin, B. Exploring the link between neighborhood environment and mental wellbeing: A case study in Beijing, China. *Landsc. Urban Plann.* **164**, 71–80. <https://doi.org/10.1016/j.landurbplan.2017.04.005> (2017).
65. Hall, E. *The Hidden Dimension: An Anthropologist Examines Humans, Use of Space in Public and in Private* (Anchor Books, 1966).
66. Calhoun, J. Population density and social pathology. *Calif. Med.* **113**, 54 (1970).
67. Catalano, R. *et al.* The health effects of economic decline. *Annu. Rev. Public Health* **32**, 431–450 (2011).
68. Meadows, G. *et al.* Resolving the paradox of increased mental health expenditure and stable prevalence. *Austral. N. Z. J. Psychiatry* **53**, 000486741985782. <https://doi.org/10.1177/0004867419857821> (2019).
69. Galea, S. *et al.* Urban built environment and depression: A multilevel analysis. *J. Epidemiol. Community Health* **59**, 822–7. <https://doi.org/10.1136/jech.2005.033084> (2005).
70. Hoare, E., Jacka, F. & Berk, M. The impact of urbanization on mood disorders: An update of recent evidence. *Curr. Opin. Psychiatry* **32**, 1. <https://doi.org/10.1097/YCO.0000000000000487> (2019).
71. Centers for Disease Control and Prevention (CDC). *500 Cities Project: Local Data for Better Health, 2014* (2016). <https://cdcargis.maps.arcgis.com/home/item.html?id=ea8b721cf9034814bce067ddefd21ecc>.
72. U.S. Department of Housing and Urban Development. *HUD Aggregated USPS Administrative Data on Address Vacancies* (2016). <https://www.huduser.gov/portal/datasets/usps.html>.
73. U.S. Department of Housing and Urban Development. *HUD Aggregated USPS Administrative Data on Address Vacancies* (2016). <https://www.census.gov/programs-surveys/acs/>. Accessed 11 Mar 2020.
74. Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* <https://doi.org/10.1080/14786440109462720> (1901).
75. Kapelner, A. & Bleich, J. BartMachine: Machine learning with bayesian additive regression trees. *J. Stat. Softw.* **70**(4), 1–40 (2016).
76. Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag. J. Sci.* **50**(302), 157–175 (1900).

Author contributions

S.M., E.F.B. and N.D.B. co-conceptualized the idea. S.M. conceived the methods, conducted the analytical experiment, analyzed data, implemented the models, and obtained funding. E.F.B. and N.D.B. collected and curated data. P.G. assisted in preparing the plots and helped in analysis. S.M., E.F.B. and P.G. conducted statistical inferencing, wrote and edited the manuscript draft. S.M. provided overall guidance. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-96801-x>.

Correspondence and requests for materials should be addressed to S.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021