

A multi-level scenario-based predictive analytics framework to model community mental health—built environment nexus

Sayanti Mukherjee (✉ [sayantim@buffalo.edu](mailto:syantim@buffalo.edu))

University at Buffalo - The State University of New York

Emmanuel Frimpong Boamah

University at Buffalo - The State University of New York

Prasangsha Ganguly

University at Buffalo - The State University of New York

Nisha Botchwey

Georgia Institute of Technology

Research Article

Keywords: community mental health, built environment nexus, multi-level scenario-based predictive analytics framework (MSPAF)

Posted Date: April 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-465875/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on September 2nd, 2021. See the published version at <https://doi.org/10.1038/s41598-021-96801-x>.

A multi-level scenario-based predictive analytics framework to model community mental health—built environment nexus

Sayanti Mukherjee^{1,*}, Emmanuel Frimpong Boamah², Prasangsha Ganguly¹, and Nisha Botchwey³

¹University at Buffalo - The State University of New York, School of Engineering and Applied Sciences, Department of Industrial and Systems Engineering, Buffalo NY, 14260, U.S.A.

²University at Buffalo - The State University of New York, School of Architecture and Planning, Department of Urban and Regional Planning, Buffalo NY, 14214, U.S.A.

³Georgia Institute of Technology, School of City & Regional Planning, Atlanta GA, 30332, U.S.A.

*sayantim@buffalo.edu

ABSTRACT

Understanding the complex community mental health—built environment nexus remains vital to building healthy and sustainable cities. The recent commission report on mental health and sustainable development goals elevates concerns about understanding these complex relationships for informed-policy actions. However, there are limited analytical and methodological frameworks to unpack these relationships. Here, we develop a multi-level scenario-based predictive analytics framework (MSPAF) to address this limitation. We employ rigorously validated interpretable machine learning algorithms and scenario-based sensitivity analyses to explore the relationship between community mental health, and socio-economic/physical aspects of built environment across the US metropolitan areas. Our results suggest that declining socio-economic conditions (e.g., poverty, low income, unemployment) are significantly associated with increased reported mental health disorders. The results also contribute to the insurance-mental health debate by showing decreased access to public health insurance is associated with increase in reported mental disorders. Finally, adults report increased mental health disorders as travel costs or housing vacancies increase, but this does not hold across all the metropolitan areas, illustrating a mixed effect of built environment's physical aspects on mental health. We conclude by highlighting future opportunities of incorporating other micro-/macro-level data into the MSPAF framework to examine the mental health—built environment nexus further.

1 Introduction

The socio-economic and physical aspects of a built environment significantly influence the community health and well-being of people. Physical aspects of a built environment include human-created infrastructure systems, such as transportation and housing infrastructure, and how they interact with the community to impact people and places [1]. Socio-economic aspects of the built environment, on the other hand, refer to economic, racial and ethnic, and relational conditions that may influence a person's ability to cope with stress. 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 rates [2], educational performance, property values [3, 4], and physical health outcomes such as obesity, heart disease, cancer, stroke, respiratory disease, and diabetes [5, 6]. Understanding and predicting health outcomes as a function of physical and socio-economic aspects 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 these built environment factors impact specific health outcomes, especially the spread of such outcomes among vulnerable and often minority populations.

Mental health is one of the specific health outcomes impacted by the built environment's socio-economic and physical aspects. Mental illness or disorder contributes significantly to the global burden of disease, accounting for 32.4% of years lived with disability (YLDs) and 13.0% of disability-adjusted life-years (DALYs), globally [7]. As of 2016, global estimates revealed that mental (e.g., chronic depression, anxiety) and substance use disorders were among the largest contributors to disability in young adults; depressive and anxiety disorders were high among females, and substance use and autism spectrum disorders were high among males [8]. In the United States, suicide ideation in

adults is increasing, with 10.3 million adults diagnosed with severe thoughts of suicide [9], and over two million youth having severe depression [10]. Depression and hopelessness are found to be key predictors of suicide ideation and attempts in young adults [11]. Based on the commissioning report on mental health and the sustainable development goals (SDGs) [12], 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 people [12].

Although the built environment heavily influences mental health, there are gaps in the literature about the complex 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 disorders [13]. Studies have also indicated a disparity among large and medium/small metropolitan areas’ suicide rates with the latter being higher than the former [14]. Variations in population demographics and socio-economic factors such as unemployment rates, household income, and climate are the key factors associated with such disparities [14]. Other studies have also looked at the link between mental health and low quality of care for mental health disorders and human rights abuses [15]. However, there are gaps in the literature about the link between 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 markers [12].

Some studies attempted to understand the link between mental health and the built environment. For example, one study found that adolescents living in physically deteriorated neighborhoods had higher health problems, including depression and anxiety than those living in ordered neighborhoods [16]. Another research that studied 1,355 residents in New York City found that living in poor quality neighborhoods was associated with a greater likelihood of individuals experiencing chronic depression, after controlling for income, race/ethnicity, age, and neighborhood-level income [17]. A cross-sectional study of adults (sixteen years and above) 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 status [18]. A systematic review of 45 studies reveals that 37 studies reported at least one built environment characteristic associated with depression or depressive symptoms [19].

However, despite some advancements in the domain of understanding the built environment and community mental health associations, some have raised concerns about the need for robust methodological approaches to evaluate the link between the built environment and mental health [20]. 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 conditions (e.g., neighborhood decline, transportation conditions, unemployment, income, race, age, social capital, education). Existing studies examine these factors using linear and static models. However, the complex, non-linear relationship between the built environment and mental health outcomes constrains linear and static models’ explanatory and predictive abilities [21]. Even though these complex interactions are widely acknowledged in the literature, studies are yet to leverage recent advances in big data analytics to explore such complexities.

This study demonstrates a novel approach to studying the complex associations between community mental health and, the physical and socio-economic aspects of the built environment, thus providing guidance on how properly planned neighborhoods may improve the overall community mental health outcomes. Specifically, we employ a multi-level predictive and scenario-based modeling framework that leverages interpretable machine learning algorithms to model the community mental-built environment nexus among adults in the large metropolitan across the US. We develop a scenario-based analysis to predict how the community mental health outcomes in these metropolitan areas change under plausible perturbations of various built environment factors.

2 Results

2.1 Predictive performance of interpretable machine learning models and model selection

In this study, we 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 Section 4.1). In healthcare research, 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. 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 interpretability [1]. In this study, we assessed the predictive performance of eight interpretable machine learning models ranging from parametric to non-parametric—generalized linear model (GLM) [22], ridge regression (RR) [23], lasso regression (LR) [24], generalized additive model (GAM) [25], multivariate adaptive regression splines (MARS) [26], gradient boosting method [27], random forest (RF) [28], and Bayesian additive regression tree (BART) [29]. 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 Section 4.1). Our results indicate that BART outperformed all the other models, and thus the following statistical inferencing is conducted using the the BART model (see Section 4.1 and Supplementary Information for BART algorithm details and model comparisons).

Key factors attributing to socio-economic and physical aspects of built environment

We leverage the variable importance plot (VIP) (see Supplementary Information) and the partial dependence plots (PDPs) to identify the key built environment predictors of community mental health, and evaluate their associated relationships (see Section 4.1 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 in this paper.

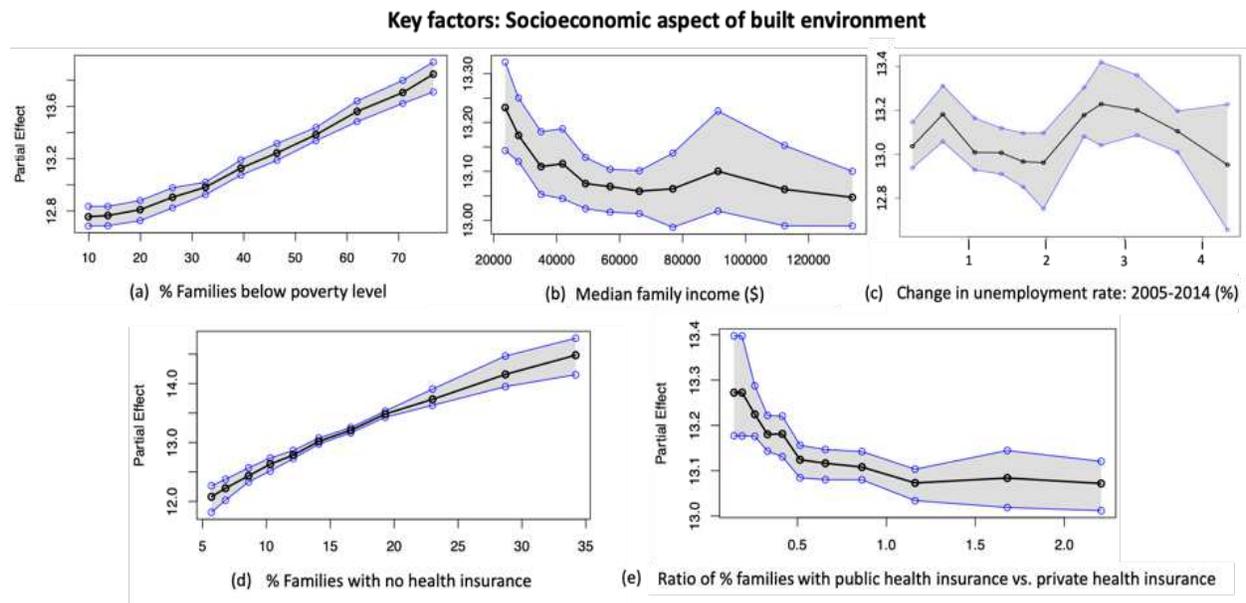


Figure 1. Partial Dependence Plots of the socio-economic aspects of built environment. The black line represents the mean partial effect and the blue lines represent the 95% confidence interval of the partial effect.

Partial dependence plots of the socio-economic aspects of built environment, considered in this study, are depicted in Fig 1. The partial dependence plot of poverty, shown in Fig 1a, indicates a strong positive correlation with community mental health. This shows 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 (shaded grey area) indicates that the estimates are associated with less uncertainty. Other significant factors in this category include economic variables like 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 K to \$20 K, the percentage of adults reporting mental health issues 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 partial dependence plot of 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 community mental health. The partial dependence plot 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 community on average.

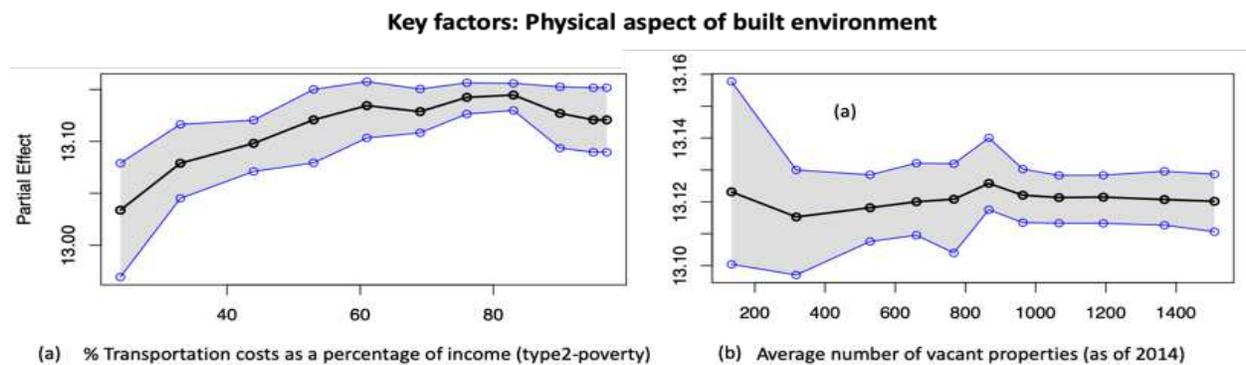


Figure 2. Partial Dependence Plots of Physical Aspects of Built Environment. The black line represents the mean partial effect and the blue lines represent the 95% confidence interval of the partial effect.

Transportation cost (percentage of household income spent on transportation) and the average vacant properties, which constitute the built environment’s physical aspect, were found to be the key predictors of community mental health. The partial dependence plot 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 across the US states. However, the relationship between average vacancy and community mental health is quite 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 it reaches the threshold point around 900 vacant properties in a community on average, the association flattens, i.e., community mental health becomes insensitive to changes in vacant properties.

2.2 Projected community mental health burden under plausible perturbations of built environment scenarios into the future

Having identified the key built environment factors associated with mental health outcomes, we employ scenario-based sensitivity analysis to understand how the mental health burden might change in the future. 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 rates techniques [30]. 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 uncertainties [31]. 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 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**) we verify if our hypothesis holds good nationally or only for certain US states, leveraging our predictive model (see "Data and Methods" section for details on creating scenarios and list of hypotheses summarized in Table 1).

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 community 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) average number of vacant properties (as of 2014). The first three categories of independent variables capture the socio-economic characteristics of built environment and the last two categories represent the physical aspects of built environment. It is assumed that at a given time period in future, two hypothetical scenarios are considered in our study: **1**) the mean of the distributions of socio-economic parameters (i.e., economic conditions, access to health insurance, and type of health insurance) and 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, these statistical perturbations help to provide important insights regarding the trends of community mental health under plausible scenarios. However, it is to be noted that our framework is generalized enough that it can be used to predict how the community mental may change in the future, given the forecasted data on socio-economic aspects and built environment is available. In our study, we present the framework illustrating how future community health might be affected under various future scenarios. Furthermore, to understand whether such shifts results in a favorable outcome (in terms of improvement in the community mental health) or not, we compared the projections with the base case scenario of community mental health outcomes by constructing ten hypotheses (see Table 1 in "Data and Methods" section). Finally, we validated our hypotheses based on our model results and outcomes.

2.2.1 The socio-economic aspects of built environment

In general, it is observed that metropolitan areas of the eastern part of the US suffer severely from poor mental health issues. The variations of community mental health outcomes across the 50 states in the US under the three different scenarios—a) worst- and best-case scenarios of economic condition, b) lack of health insurance, and c) access to public health insurance are discussed in this section.

Economic condition

The economic conditions capture the interplay of poverty, median household income, and unemployment rate experienced within a metropolitan area. Since economic condition comprises three variables, for simplicity the hypothetical scenario is constructed by perturbing all the three variables at the same time. 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 Figure 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 ($\Delta K < 0$). The scenario analysis, depicted in Figure 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 health [32]. These conditions are linked to poverty [32, 12], income [33], and unemployment [34]. 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 increasing or declining economic conditions 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).

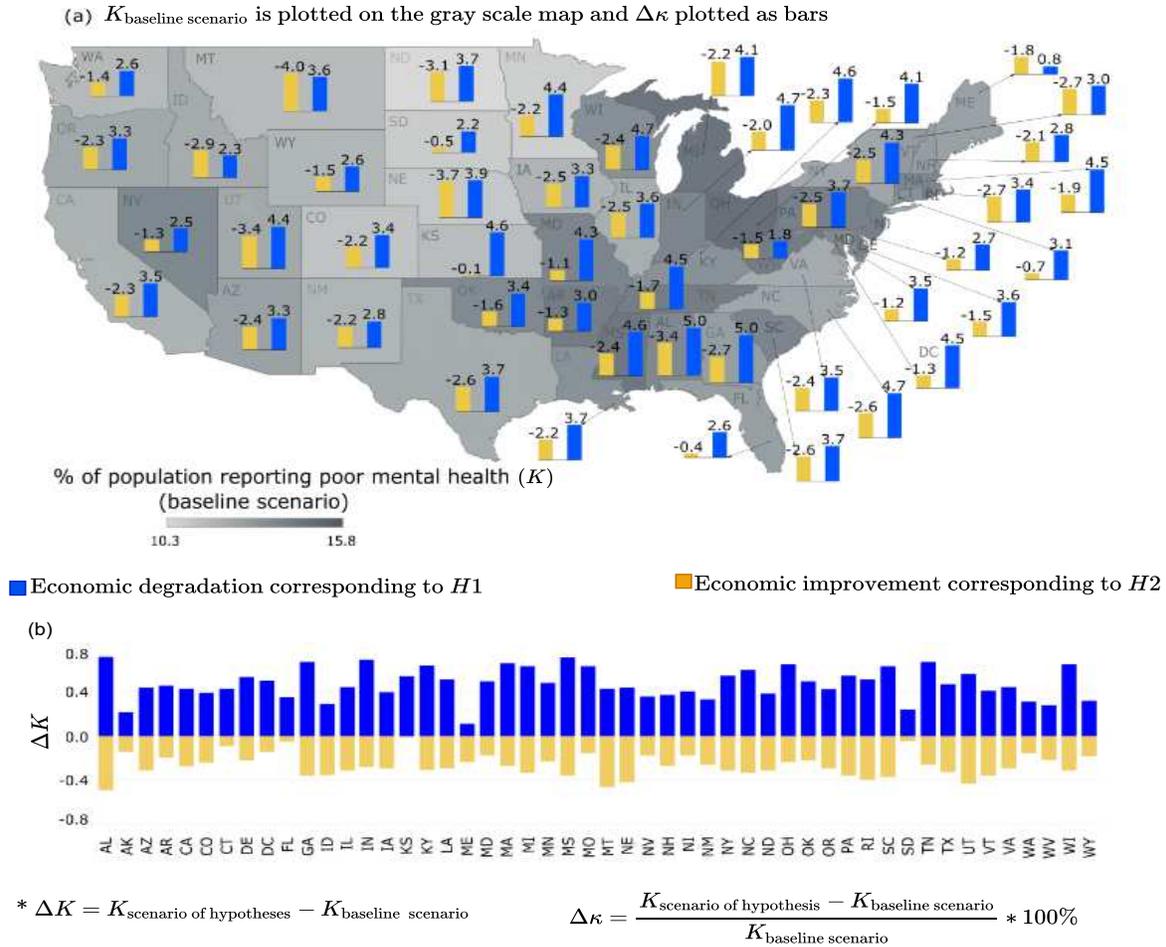


Figure 3. Economic condition scenario for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) ΔK is plotted.

Unavailability of health insurance

In this case, the variable under consideration is the percentage of families with no health insurance or unavailability of health insurance. It was hypothesized that an overall improvement of community mental health would be observed when the unavailability of mental health will decrease or more families will have some health insurance (hypothesis: $H3$). An opposite effect will be expected with increased unavailability of health insurance, i.e., more families being deprived of health insurance which may lead to worsening mental health problems (hypothesis: $H4$). The scenario-based analysis in Figure 4 suggests that these two hypotheses generally hold good for all the metropolitan areas across US considered in this study. In Figure 4, it is observed that when the unavailability of health insurance increases (yellow bars), the number of people 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 less unavailability of 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 states such as Montana, North Dakota, South Dakota, and Vermont. Studies show that states providing access to mental health insurance minimize suicide rates [35], 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 Australians [36]. The underlying logic follows that increasing access to health insurance and mental health insurance specifically will likely increase the likelihood of people accessing mental healthcare, which will ultimately improve 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' stress and mental health outcomes. In other words, not having health insurance as a form of socio-economic deprivation, similar to experiencing low income and poverty, is a pressure point in increasing the percentage of adults reporting mental disorders within metropolitan areas.

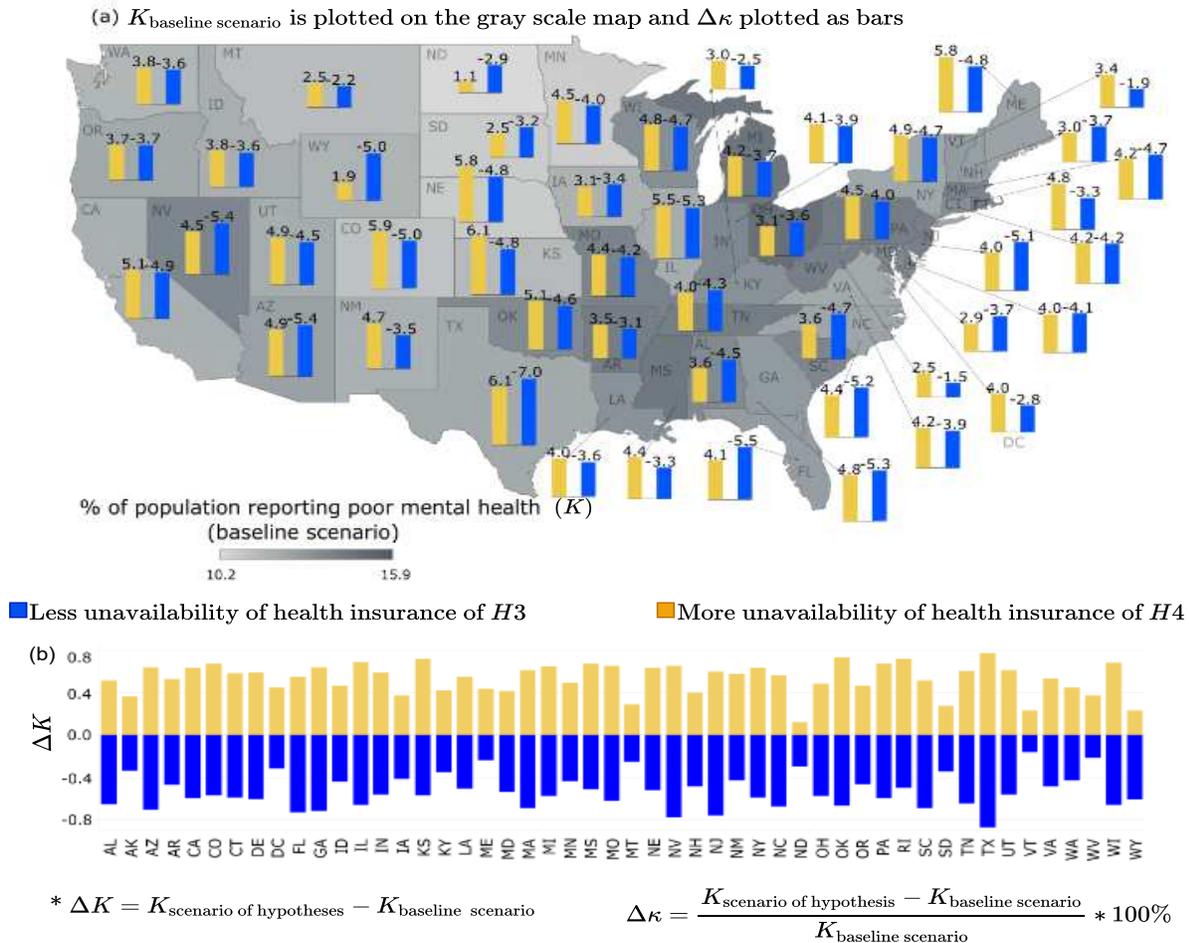


Figure 4. Unavailability of health insurance scenario for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) ΔK is plotted.

Access to public health insurance

Building on the health insurance scenario analysis, it was hypothesized that there are differentials in how access to different types of health insurance, i.e., public vs. private health insurance impact mental health outcomes. Specifically, it was hypothesized that with decreased access to public health insurance (i.e., a less 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 or decreased access to private health insurance within metropolitan areas in the states (hypothesis: $H6$). However, these hypotheses were minimally supported in the scenario results across the states, as depicted in Figure 5. Although the trend of increasing or decreasing mental health outcomes is 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 indicates that the overall mental health outcomes across the US's metropolitan areas are not much sensitive to the type of health insurance that their community has access to. However, the hypothesis of decreasing K with increasing access to public health insurance (hypothesis: $H6$) was overwhelmingly 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 services [37].

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

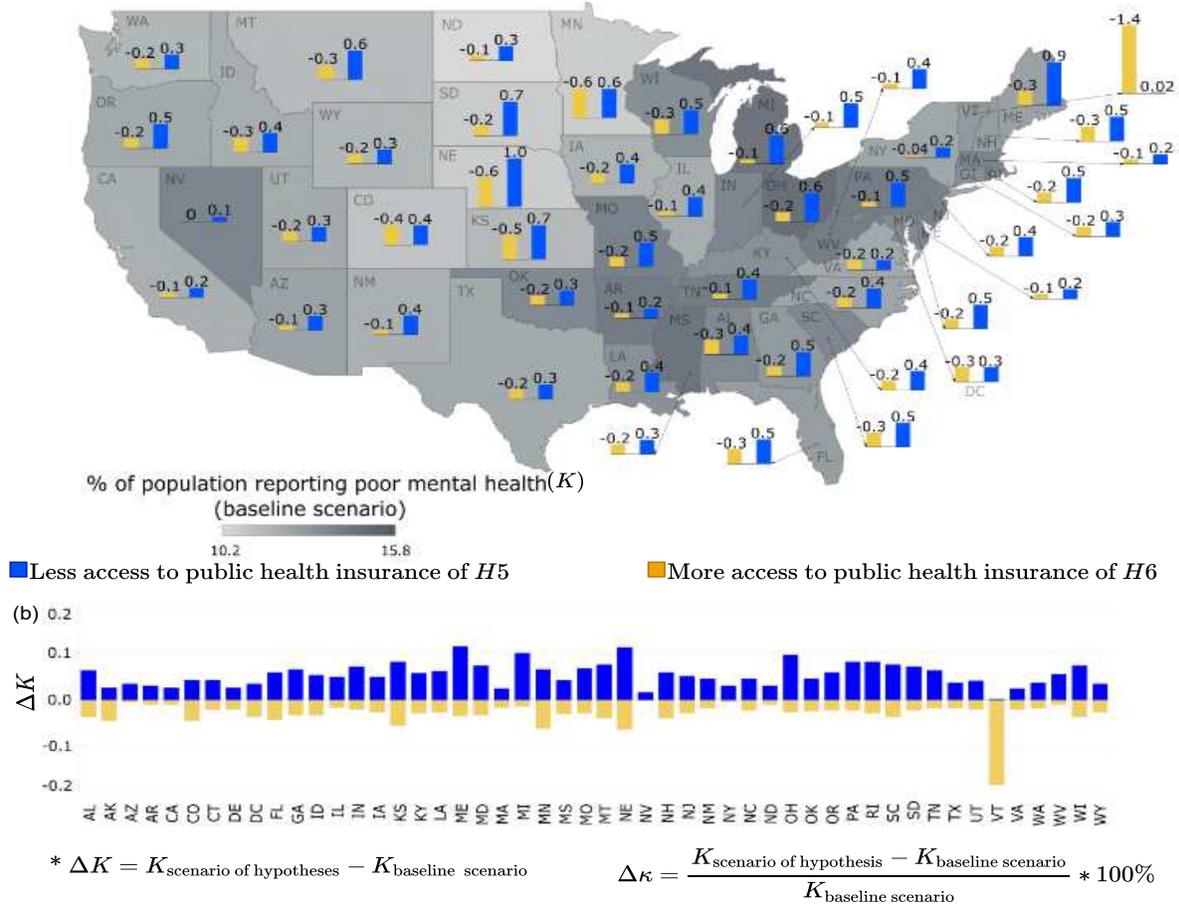


Figure 5. Access to public health insurance scenarios for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) ΔK is plotted.

Physical aspect of the built environment

Travel 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 were that the percentage of adults reporting mental disorders (K) would decrease with decreasing travel cost (hypothesis: $H7$) and that, K would increase with increasing travel cost (hypothesis: $H8$). These two hypotheses do not hold for some metropolitan areas in some states. For instance, increased reported mental health issues (K), as travel cost increases (hypothesis: $H8$) do not hold in metropolitan areas within states such as Colorado, Delaware, Maine, Minnesota, Nebraska, North Dakota, Utah, Vermont, and Washington. In these metropolitan areas, there is a decrease in mental disorders reported by adults as 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., increase in the percentage of adults reporting mental disorders. The mixed results of how travel costs impact mental health outcomes signal the need for an in-depth and granular inquiry into how the built environment’s physical aspect impacts mental health outcomes in the cities.

Housing vacancy

In the housing vacancy scenario, the hypotheses explored in this study looked at the extent to which neighborhood

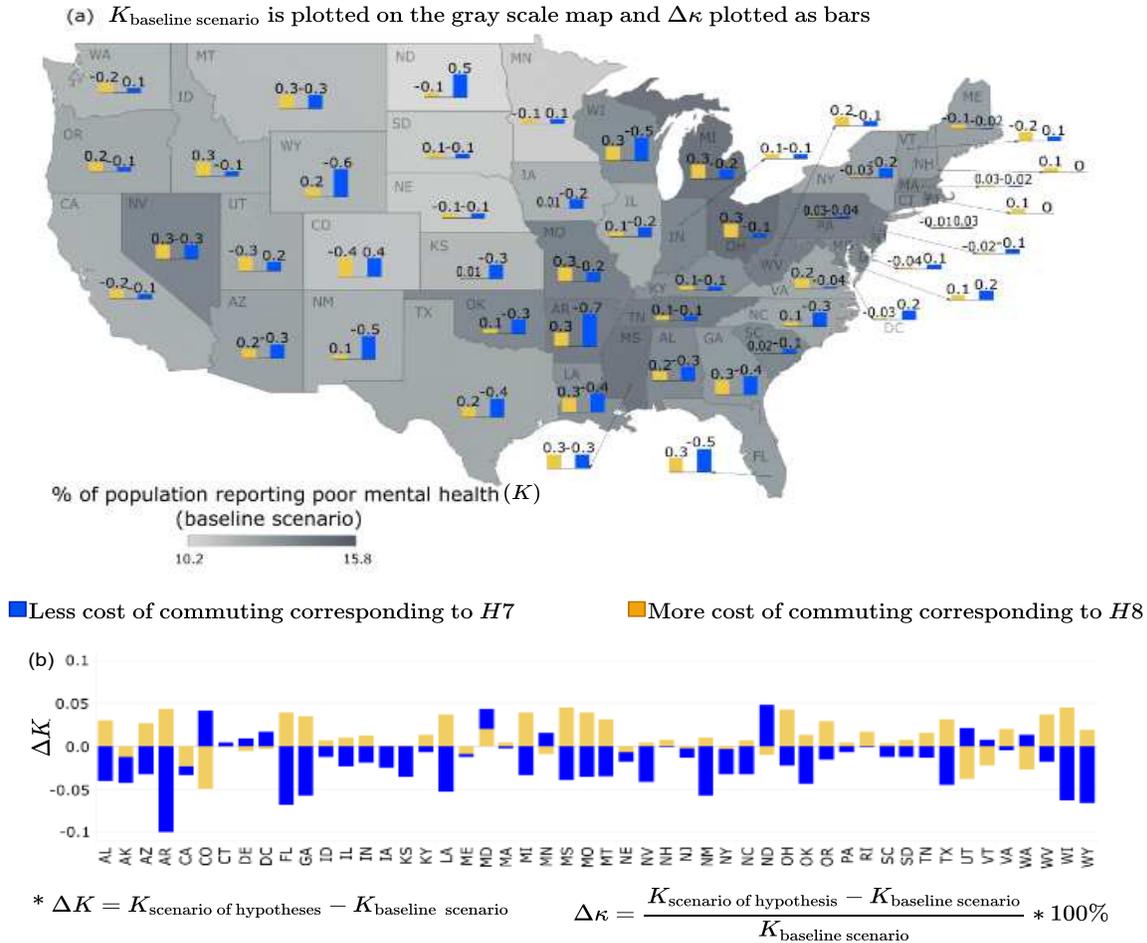


Figure 6. Travel cost scenarios for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) ΔK is plotted.

decline impacts mental health outcomes in the metropolitan areas across the 50 states in the US. Specifically, it was 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 is expected to increase the percentage of adults reporting mental disorders (hypothesis: $H10$). Overall, it has been identified that the intuition-based hypotheses did not hold good in our study. As depicted in Figure 7, when the vacancy is increasing (yellow bars), most states experience an improvement or deterioration in community mental health (K). On the other hand, when a community is expanding, attributed by decreased vacancy (blue bars), most of the states see an increase in K . This result may be an outcome of the "Behavioral Sink" phenomenon [38, 39]. However, for some states the reverse phenomenon has been observed. When the vacancy is decreasing, metropolitan areas of some states (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 support the earlier observation that there is more to the story when parsing the impacts of the built environment on mental health outcomes. More granular level analysis complemented by macro-level levels 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 sensitivity of community mental health (K) to housing vacancy is highly varied across the US states. Hence, it is difficult to classify whether a particular scenario of housing vacancy perturbation leads

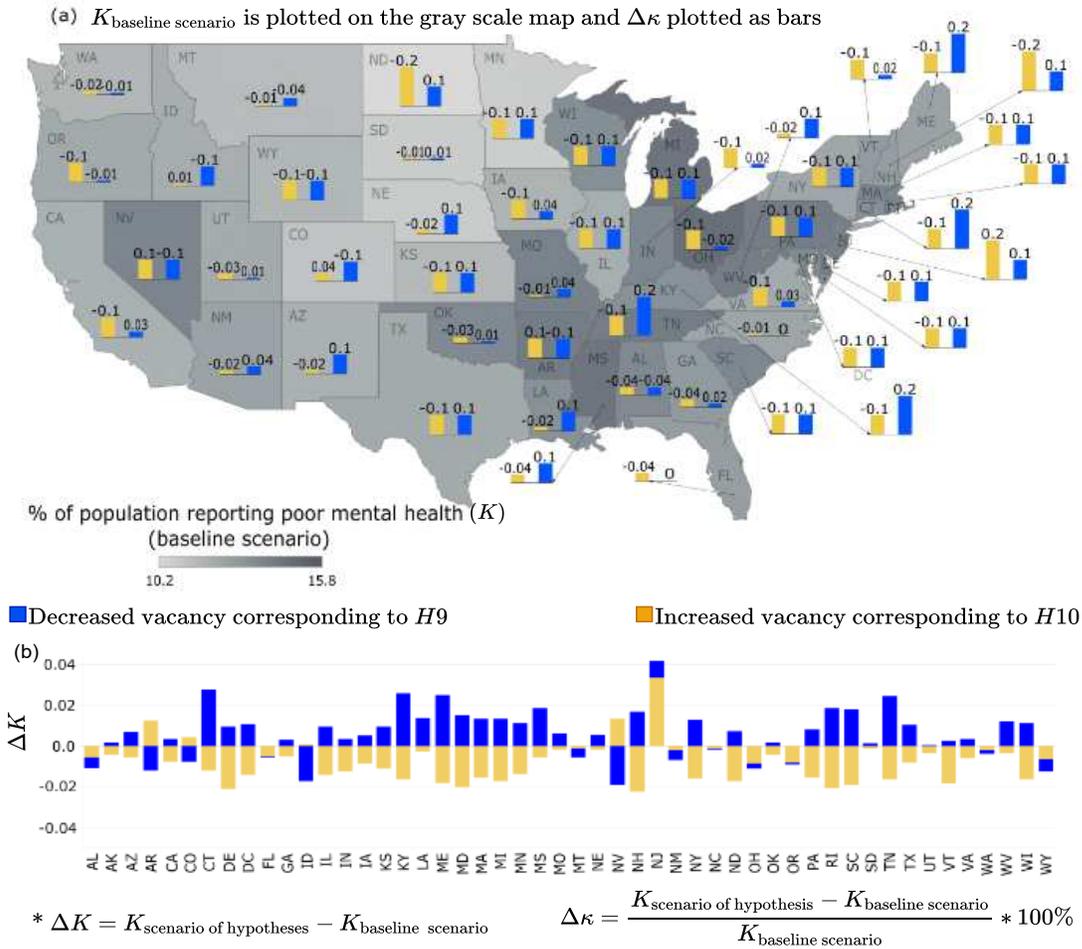


Figure 7. Housing vacancy scenario for % of adults aged > 18 years reporting poor mental health for > 14 days (K): (a) $\Delta\kappa$ is plotted as the bars and K for base line scenario is plotted as gray scale intensity on the US map; and for (b) $\Delta\kappa$ is plotted.

to the **best case scenario** indicating that the community level mental health improves unanimously across the US; or if the perturbation leads to a **worst-case scenario** where the community mental health 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 $\Delta\kappa = K_{\text{scenario under consideration}} - K_{\text{base case scenario}}$ is (+)ve, then the perturbation scenario under consideration is depicted as the worst-case scenario. Similarly, if the mean $\Delta\kappa$ is found to be (-)ve, then there is a decline in the percentage of the population reporting mental health issues, so the scenario is termed as a best-case scenario.

3 Discussion

This paper employs a library of supervised interpretable machine learning models and scenario-based sensitivity analysis to explore the relationship between adults' mental health, and the socio-economic and physical aspects of the built environment in the US's largest metropolitan areas. The interpretable machine learning models and scenario-based analysis 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 [40, 12, 34, 33]. This paper provides evidence to support 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 literature debate 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 [35, 36, 41], the partial dependence plots of the health insurance variables show that there is a strong increasing trend between lack of health insurance and adults reporting mental health disorders in metropolitan areas across 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 disorders reported increased as housing vacancy increase in some metropolitan areas, but this also does not hold in all metropolitan areas. The commissioning about the sustainable development goals (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 health [12]. This paper adds to some of the existing studies [42, 43], supporting concerns raised in this commissioning report. More importantly, it also adds to the literature on how urbanization (e.g., increasing sprawl and associated commuting costs) impact mood disorders [44]. The mixed results call for caution when discussing how the built environment's physical aspects impact community mental health. More importantly, in the future our proposed multi-level scenario-based predictive analytics framework (MSPAF), leveraging data-driven interpretable machine learning algorithms, can consider other multiple micro- and macro-level physical features of the built environment to examine how these different physical aspects impact mental health outcomes in metropolitan areas. The national scale of this paper's analysis and limitations in obtaining data for all metropolitan areas at this scale, made it impossible to include other variables. This paper provides a robust data-driven methodological framework and evidentiary basis to examine the community mental health—built environment nexus. Future research can build on this framework and evidence by incorporating other datasets on the socio-economic and physical aspects of the built environment and their impacts on mental health, not only in metropolitan areas, but also in rural areas.

4 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), 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 2014. The housing vacancy data for the year 2014 is obtained from US Housing and Urban Development (HUD) at a census tract level. 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 2015. The travel cost data is obtained from the US Department of Housing and Urban Development Location Affordability Index (LAI), 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 variables [45]. 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.

4.1 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, $Y = f(X) + \varepsilon$ [46]. 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 = variance [47]. 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 Model [22], Ridge Regression [23] and Lasso Regression [24]
2. *Semi-parametric Models*: Generalized Additive Model [25], Multi Adaptive Regression Splines [26],
3. *Non-parametric Models*: Random Forest [28] and Gradient Boosting Method [27] Bayesian Additive Regression trees [29]

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 [14, 48, 49, 50, 51, 52, 53, 54]. 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.

4.2 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 [29, 55]. Mathematically, BART can be expressed as,

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

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.

4.3 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) [46, 29, 55]. 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 affect 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 .

4.4 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 estimation [56]. After the best distribution(s) of the predictors(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.

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 1).

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.

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

In Figures 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 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.

References

- [1] P James, P Troped, and F Laden. “The Impact of the Built Environment on Health”. In: *Women and Health*. Ed. by M Goldman, R Troisi, and K Rexrode. Second. Academic Press, 2013. Chap. 50, pp. 753–763.
- [2] Prasangsha Ganguly and Sayanti Mukherjee. “A multifaceted risk assessment approach using statistical learning to evaluate socio-environmental factors associated with regional felony and misdemeanor rates”. In: *Physica A: Statistical Mechanics and its Applications* (2021), p. 125984.
- [3] Dan Immergluck and Geoff Smith. “The External Costs of Foreclosure: The Impact of Single-Family Mortgage Foreclosures on Property Values”. In: *Housing Policy Debate* 17 (Jan. 2006), pp. 57–79.
- [4] Erica Raleigh and George Galster. “Neighborhood Disinvestment, Abandonment, And Crime Dynamics”. In: *Journal of Urban Affairs* 37 (Mar. 2014), pp. 367–396.
- [5] Jason N Houle. “Mental health in the foreclosure crisis”. In: *Social science & medicine* 118 (2014), pp. 1–8.
- [6] Mark R Lindblad and Sarah F Riley. “Loan modifications and foreclosure sales during the financial crisis: Consequences for health and stress”. In: *Housing Studies* 30.7 (2015), pp. 1092–1115.
- [7] Daniel Vigo, Graham Thornicroft, and Rifat Atun. “Estimating the true global burden of mental illness”. In: *The Lancet Psychiatry* 3.2 (2016), pp. 171–178.
- [8] Theo Vos, Amanuel Alemu Abajobir, and Kalkidan Hassen Abate 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”. In: *The Lancet* 390.10100 (2017), pp. 1211–1259.
- [9] T. Nguyen, M. Hellebuyck, and M. et al. Halpern. “The state of mental health in America 2018”. In: *Mental Health America* (2017).
- [10] G. Enos. “Annual report raises red flags on youth access to care”. In: *Mental Health Weekly* 29 (2019), pp. 3–4.
- [11] Zhiyuan Wei and Sayanti Mukherjee. “Health-Behaviors Associated With the Growing Risk of Adolescent Suicide Attempts: A Data-Driven Cross-Sectional Study”. In: *American Journal of Health Promotion* 0.0 (0). PMID: 33297721, p. 0890117120977378. DOI: [10.1177/0890117120977378](https://doi.org/10.1177/0890117120977378). eprint: <https://doi.org/10.1177/0890117120977378>. URL: <https://doi.org/10.1177/0890117120977378>.
- [12] V Patel, S Saxena, and C et al. Lund. “The Lancet Commission on global mental health and sustainable development”. In: *The Lancet* 392 (2018), pp. 1553–1598.
- [13] Crick Lund et al. “Poverty and common mental disorders in low and middle income countries: A systematic review”. In: *Social Science Medicine* 71.3 (2010), pp. 517–528.
- [14] Sayanti Mukherjee and Zhiyuan Wei. “Suicide disparities across urban and suburban areas in the U.S.: A comparative assessment of socio-environmental factors using a data-driven predictive approach”. In: *arXiv Preprint - arXiv:2011.08171* (2020).
- [15] V Patel, A Kleinman, and B Saraceno. “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, and Courage*. Ed. by M Dudley, D Silove, and F Gale. Oxford University Press, 2012, pp. 362–375.
- [16] Carol Aneshensel and Clea Mcneely. “The Neighborhood Context of Adolescent Mental Health”. In: *Journal of health and social behavior* 37 (Jan. 1997), pp. 293–310. DOI: [10.2307/2137258](https://doi.org/10.2307/2137258).
- [17] Sandro Galea et al. “Urban built environment and depression: a multilevel analysis”. In: *Journal of Epidemiology & Community Health* 59.10 (2005), pp. 822–827.
- [18] Scott Weich et al. “Mental health and the built environment: Cross-sectional survey of individual and contextual risk factors for depression”. In: *The British journal of psychiatry : the journal of mental science* 180 (June 2002), pp. 428–33. DOI: [10.1192/bjp.180.5.428](https://doi.org/10.1192/bjp.180.5.428).
- [19] James Sallis et al. “Neighborhood built environment and income: Examining multiple health outcomes”. In: *Social science medicine* 68 (Mar. 2009), pp. 1285–93.
- [20] D. Halpern. “Mental Health and the Built Environment More than Bricks and Mortar?” In: *London and New York: Routledge* (2014).
- [21] Marco Helbich et al. “More green space is related to less antidepressant prescription rates in the Netherlands: A Bayesian geospatial regression approach”. In: *Environmental Research* 166 (2018), pp. 290–297.

- [22] JA Nelder and RWM Wedderburn. “Generalized Linear Models”. In: *Journal of Statistical Society* 135.3 (1972), pp. 370–384.
- [23] A. N. Tikhonov. “Solution of incorrectly formulated problems and the regularization method”. In: *Soviet Mathematics* 4 (1963), pp. 1035–1038.
- [24] R. Tibshirani. “Regression Shrinkage and Selection via the Lasso”. In: *Journal of the Royal Statistical Society* 58.1 (1996), pp. 267–288.
- [25] T Hastie and R. Tibshirani. “Generalized Additive Models”. In: *Statistical Science* 1 (1986), pp. 297–310.
- [26] JH Friedman. “Multivariate Adaptive Regression Splines”. In: *Annals of Statistic* 19 (1991), pp. 1–67.
- [27] L. Breiman. “Arcing the Edge”. In: *Technical Report Statistics Department, University of California, Berkeley* 486 (1997).
- [28] L. Breiman. “Random Forests”. In: *Machine Learning* 45 (2001), pp. 5–32.
- [29] A. Chipman Hugh, Edward I. George, and Robert E. McCulloch. “BART: Bayesian Additive Regression Trees”. In: *The Annals of Applied Statistics* 4.1 (2010), pp. 266–298.
- [30] Colin D Mathers et al. “Sensitivity and uncertainty analyses for burden of disease and risk factor estimates”. In: *Global burden of disease and risk factors* (2006).
- [31] Millett Granger Morgan, Max Henrion, and Mitchell Small. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge university press, 1990.
- [32] Crick Lund, Emily Baron, and Erica Breuer. “Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews”. In: *The Lancet Psychiatry* 5 (Mar. 2018), pp. 357–369.
- [33] Mckenzie Roddy, Galena Rhoades, and Brian Doss. “Effects of ePREP and OurRelationship on Low-Income Couples’ Mental Health and Health Behaviors: a Randomized Controlled Trial”. In: *Prevention Science* 21 (Feb. 2020). DOI: [10.1007/s11121-020-01100-y](https://doi.org/10.1007/s11121-020-01100-y).
- [34] Karsten Paul and Klaus Moser. “Unemployment Impairs Mental Health: Meta-Analyses”. In: *Journal of Vocational Behavior* 74 (June 2009), pp. 264–282. DOI: [10.1016/j.jvb.2009.01.001](https://doi.org/10.1016/j.jvb.2009.01.001).
- [35] Matthew Lang. “The impact of mental health insurance laws on state suicide rates”. In: *Health Economics* 22.1 (2013), pp. 73–88. DOI: <https://doi.org/10.1002/hec.1816>.
- [36] Anthony Jorm. “Australia’s ‘Better Access’ scheme: Has it had an impact on population mental health?”. In: *Australian New Zealand Journal of Psychiatry* 52 (Oct. 2018), p. 000486741880406. DOI: [10.1177/0004867418804066](https://doi.org/10.1177/0004867418804066).
- [37] Gillian MacNaughton et al. “The Impact of Human Rights on Universalizing Health Care in Vermont, USA”. In: *Health and human rights* 17 (Dec. 2015), E83–E95.
- [38] Edward Hall. “The Hidden Dimension: An Anthropologist Examines Humans’ Use of Space in Public and in Private”. In: *Anchor Books* (1966).
- [39] J. Calhoun. “Population density and social pathology.” In: *California medicine* (1970).
- [40] R. Catalano et al. “The health effects of economic decline”. In: *Annual review of public health* 32 (2011), pp. 431–450.
- [41] Graham Meadows et al. “Resolving the paradox of increased mental health expenditure and stable prevalence”. In: *Australian New Zealand Journal of Psychiatry* 53 (June 2019), p. 000486741985782. DOI: [10.1177/0004867419857821](https://doi.org/10.1177/0004867419857821).
- [42] Sandro Galea et al. “Urban built environment and depression: A multilevel analysis”. In: *Journal of epidemiology and community health* 59 (Nov. 2005), pp. 822–7. DOI: [10.1136/jech.2005.033084](https://doi.org/10.1136/jech.2005.033084).
- [43] Kyungsoon Wang and Dan Immergluck. “The Geography of Vacant Housing and Neighborhood Health Disparities After the U.S. Foreclosure Crisis”. In: 20 (July 2018), pp. 139–164.
- [44] Erin Hoare, Felice Jacka, and Michael Berk. “The impact of urbanization on mood disorders: an update of recent evidence”. In: *Current Opinion in Psychiatry* 32 (May 2019), p. 1. DOI: [10.1097/YCO.0000000000000487](https://doi.org/10.1097/YCO.0000000000000487).
- [45] Karl Pearson. “On lines and planes of closest fit to systems of points in space”. In: *Philosophical Magazine* (1901). DOI: [10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720). URL: <https://doi.org/10.1080/14786440109462720>.

- [46] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.
- [47] Gareth James et al. “An Introduction to Statistical Learning with Applications in R”. In: Springer, New York, NY, 2015.
- [48] Sayanti Mukherjee and Roshanak Nateghi. “Climate sensitivity of end-use electricity consumption in the built environment: an application to the state of Florida, United States”. In: *Energy* 128 (2017), pp. 688–700.
- [49] Sayanti Mukherjee and Roshanak Nateghi. “A Data-Driven Approach to Assessing Supply Inadequacy Risks Due to Climate-Induced Shifts in Electricity Demand”. In: *Risk Analysis* (2018).
- [50] Sayanti Mukherjee, Roshanak Nateghi, and Makarand Hastak. “A multi-hazard approach to assess severe weather-induced major power outage risks in the us”. In: *Reliability Engineering & System Safety* 175 (2018), pp. 283–305.
- [51] Sayanti Mukherjee, Vineeth C.R., and Roshanak Nateghi. “Evaluating regional climate-electricity demand nexus: A composite Bayesian predictive framework”. In: *Applied Energy* 235.2019 (2019), pp. 1561–1582.
- [52] Roshanak Nateghi and Sayanti Mukherjee. “A multi-paradigm framework to assess the impacts of climate change on end-use energy demand”. In: *PloS One* 12.11 (2017), e0188033.
- [53] Panteha Alipour, Sayanti Mukherjee, and Roshanak Nateghi. “Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the Texas region”. In: *Energy* 185 (2019), pp. 1143–1153.
- [54] Renee Obringer, Sayanti Mukherjee, and Roshanak Nateghi. “Evaluating the climate sensitivity of coupled electricity-natural gas demand using a multivariate framework”. In: *Applied Energy* 262.114419 (2020).
- [55] Adam Kapelner and Justin Bleich. “BartMachine: Machine Learning with Bayesian Additive Regression Trees”. In: *Journal of Statistical Software, Articles* 70.4 (2016), pp. 1–40.
- [56] Karl Pearson. “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”. In: *Philosophical Magazine and Journal of Science* 50.302 (1900), pp. 157–175.

Author contributions statement

S.M. and E.F.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.

Table 1. Summary of the hypotheses

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

Figures

Key factors: Socioeconomic aspect of built environment

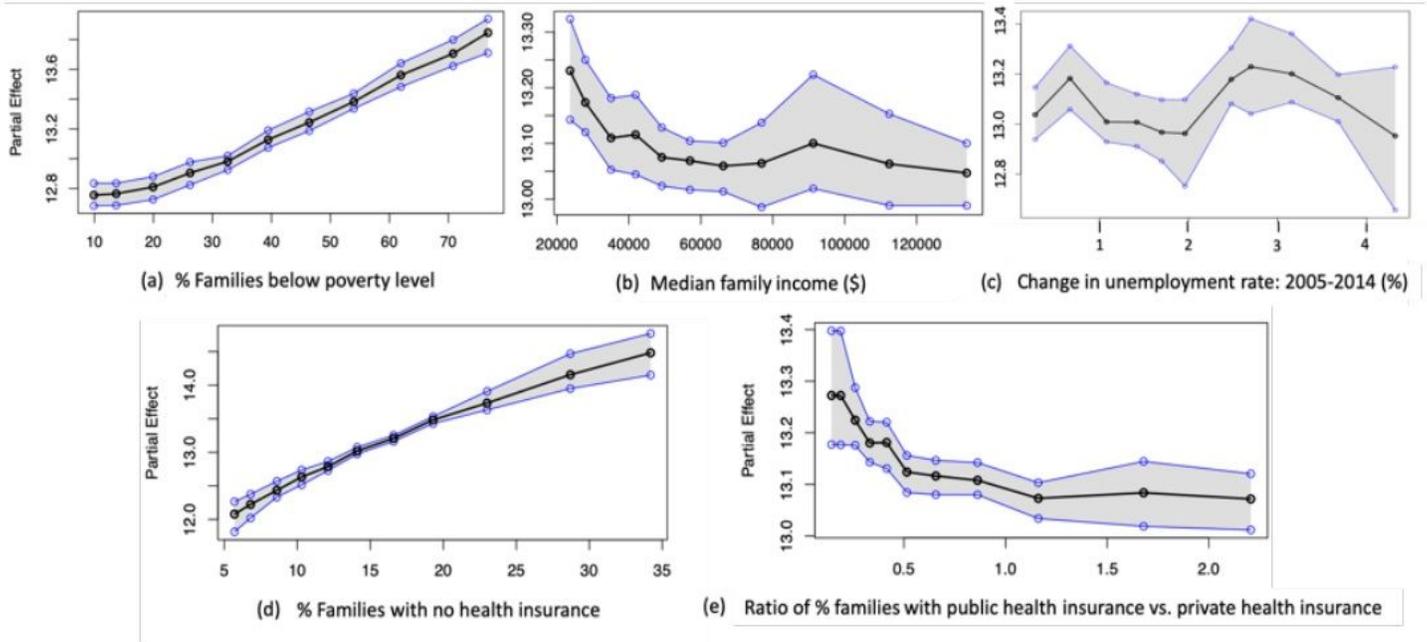


Figure 1

Partial Dependence Plots of the socio-economic aspects of built environment. The black line represents the mean partial effect and the blue lines represent the 95% confidence interval of the partial effect.

Key factors: Physical aspect of built environment

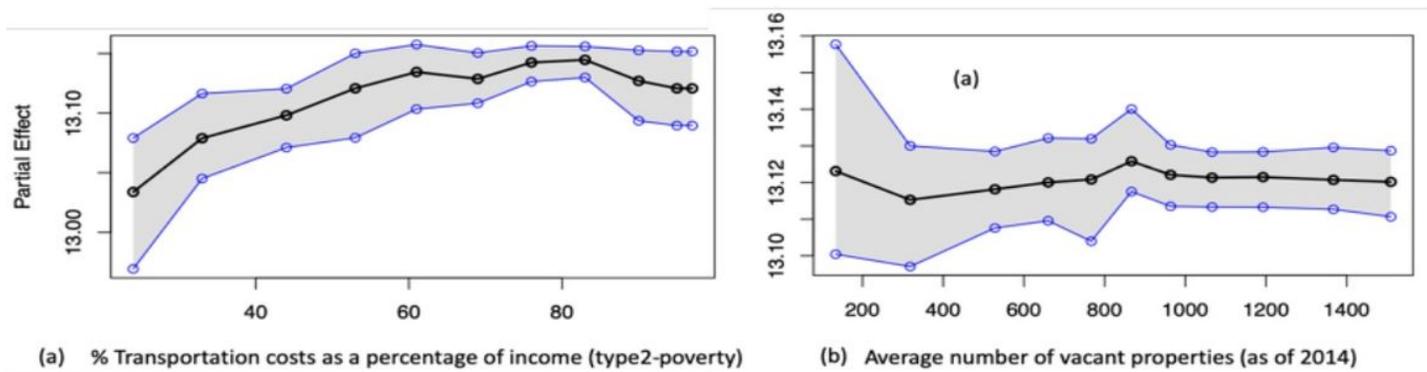
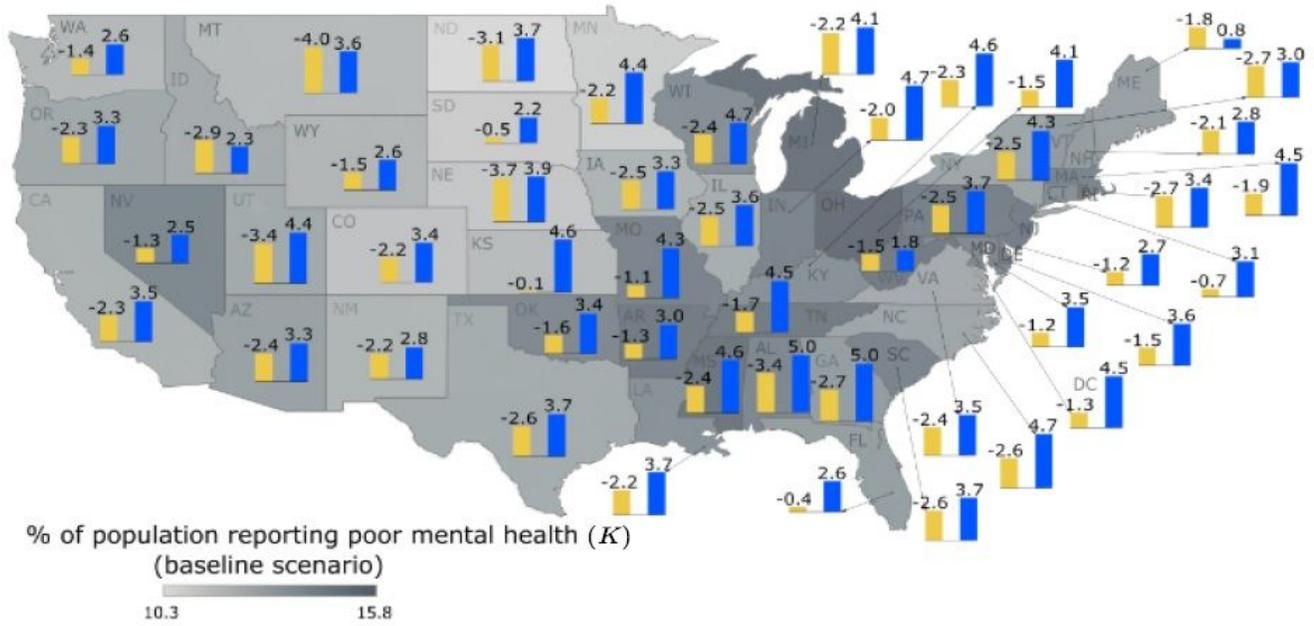


Figure 2

Partial Dependence Plots of Physical Aspects of Built Environment. The black line represents the mean partial effect and the blue lines represent the 95% confidence interval of the partial effect.

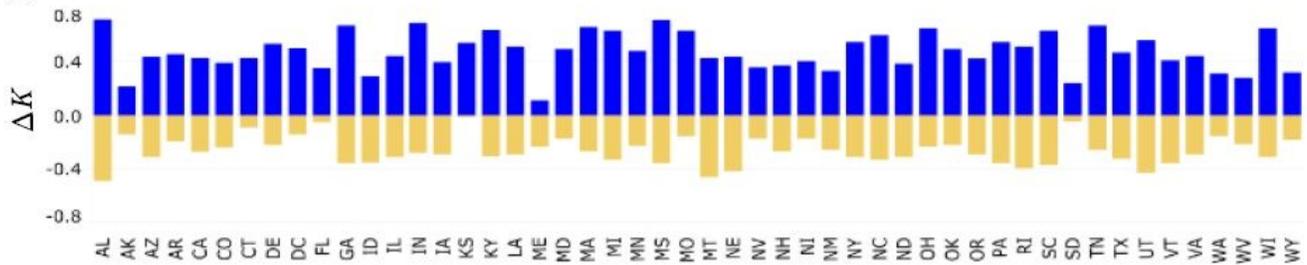
(a) $K_{\text{baseline scenario}}$ is plotted on the gray scale map and $\Delta\kappa$ plotted as bars



■ Economic degradation corresponding to $H1$

■ Economic improvement corresponding to $H2$

(b)



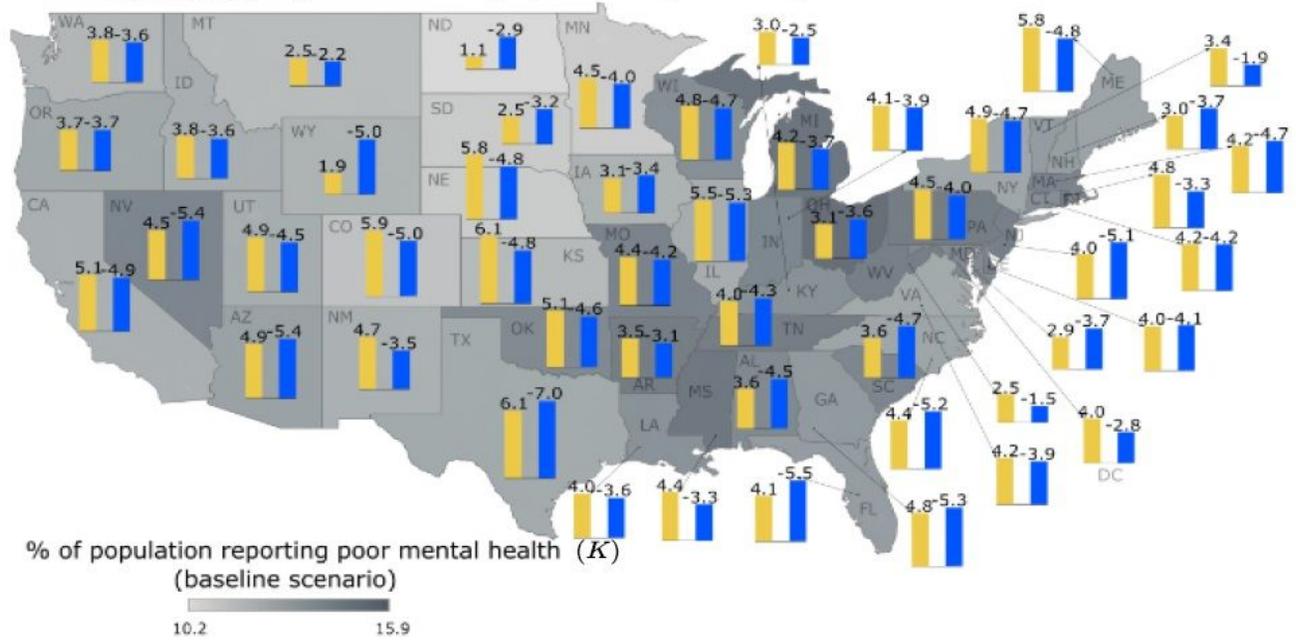
* $\Delta K = K_{\text{scenario of hypotheses}} - K_{\text{baseline scenario}}$

$$\Delta\kappa = \frac{K_{\text{scenario of hypotheses}} - K_{\text{baseline scenario}}}{K_{\text{baseline scenario}}} * 100\%$$

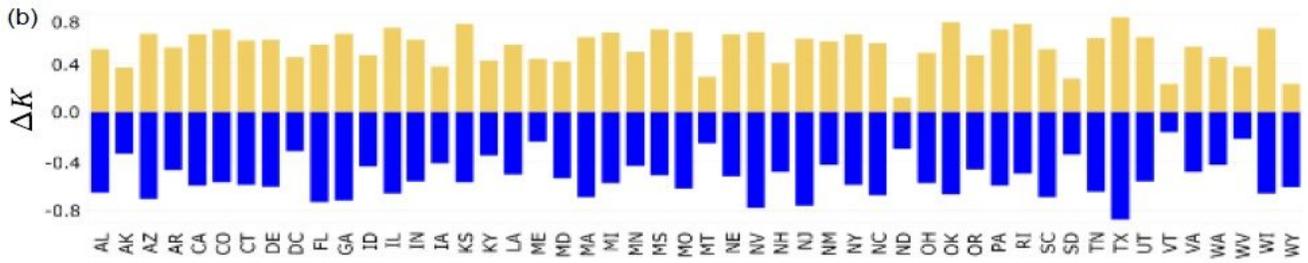
Figure 3

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.

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



■ Less unavailability of health insurance of $H3$ ■ More unavailability of health insurance of $H4$



* $\Delta K = K_{\text{scenario of hypotheses}} - K_{\text{baseline scenario}}$ $\Delta\kappa = \frac{K_{\text{scenario of hypotheses}} - K_{\text{baseline scenario}}}{K_{\text{baseline scenario}}} * 100\%$

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.

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

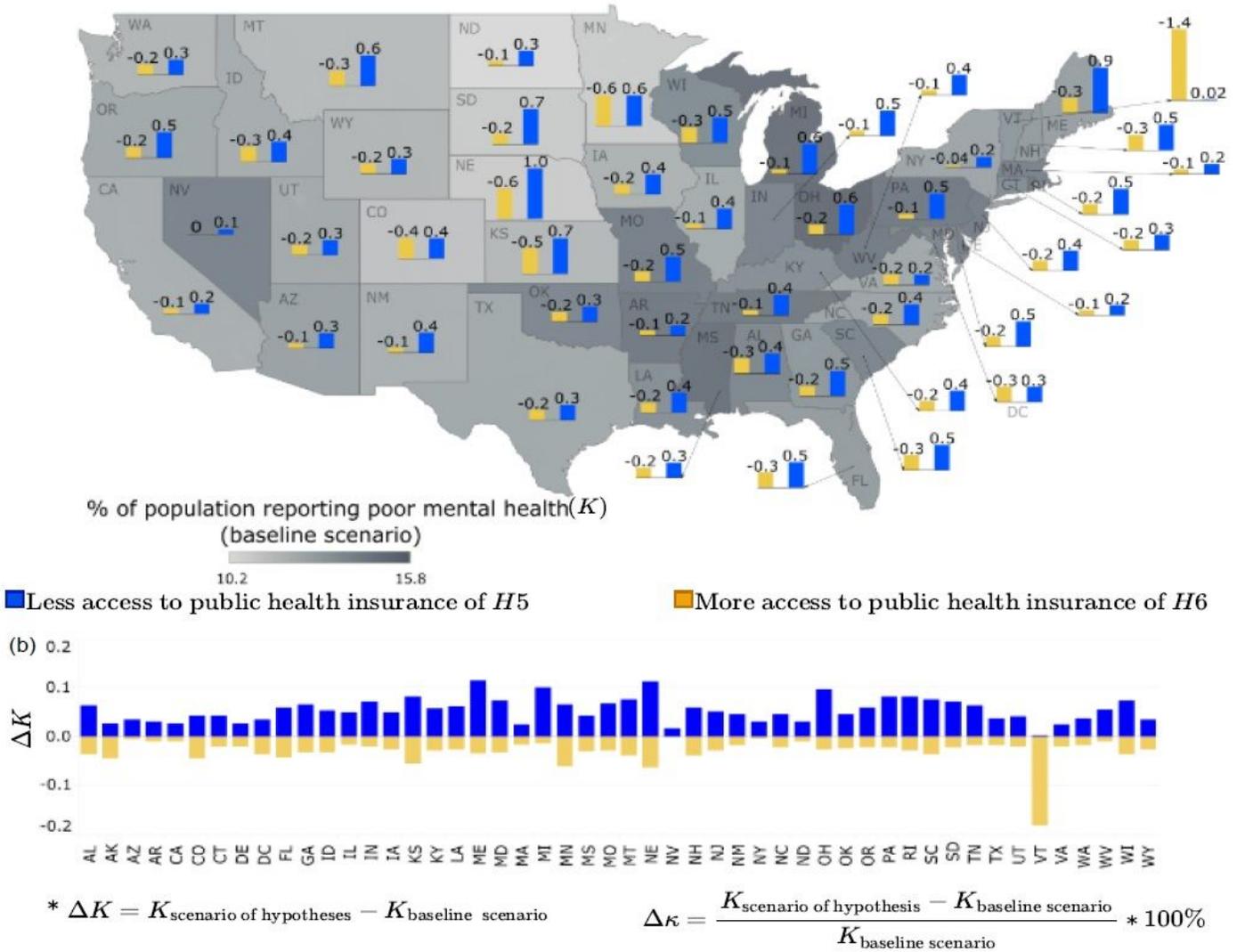
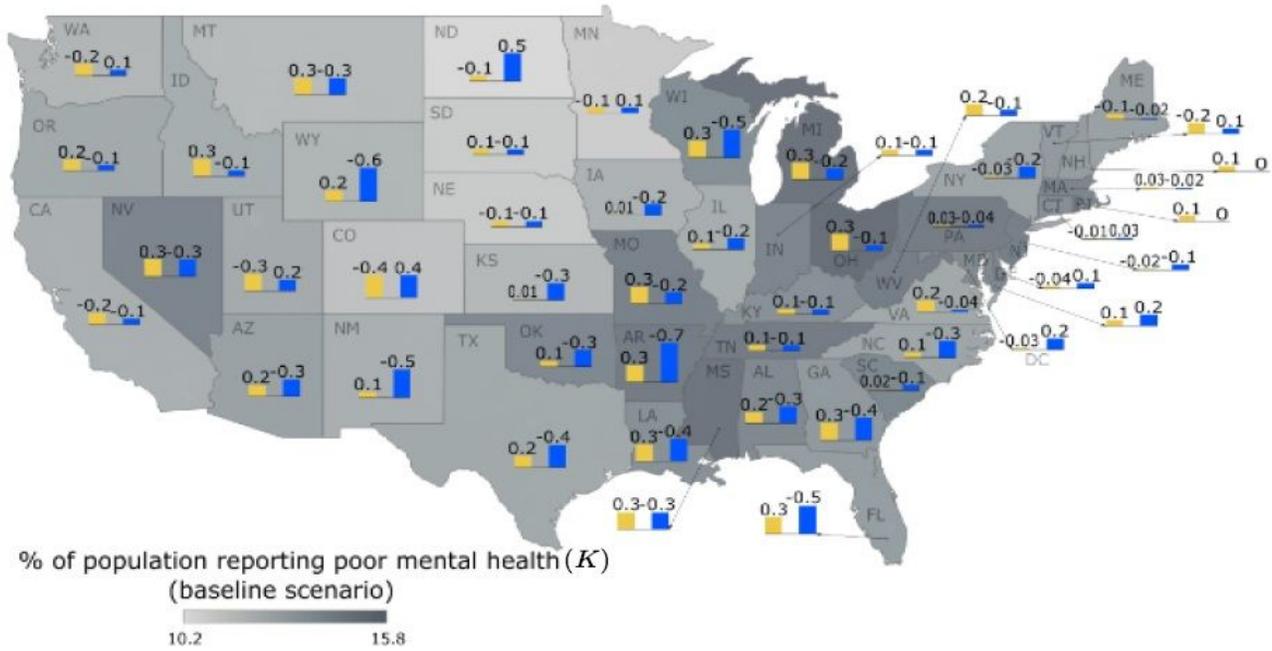


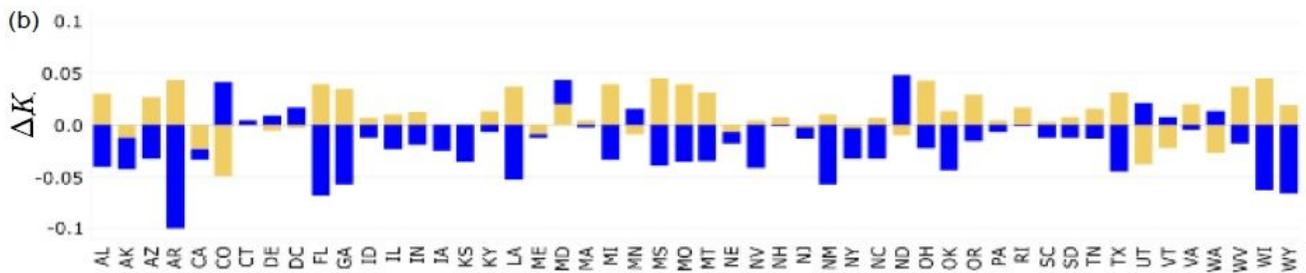
Figure 5

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.

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



■ Less cost of commuting corresponding to $H7$ ■ More cost of commuting corresponding to $H8$



* $\Delta K = K_{\text{scenario of hypotheses}} - K_{\text{baseline scenario}}$

$$\Delta\kappa = \frac{K_{\text{scenario of hypothesis}} - K_{\text{baseline scenario}}}{K_{\text{baseline scenario}}} * 100\%$$

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.

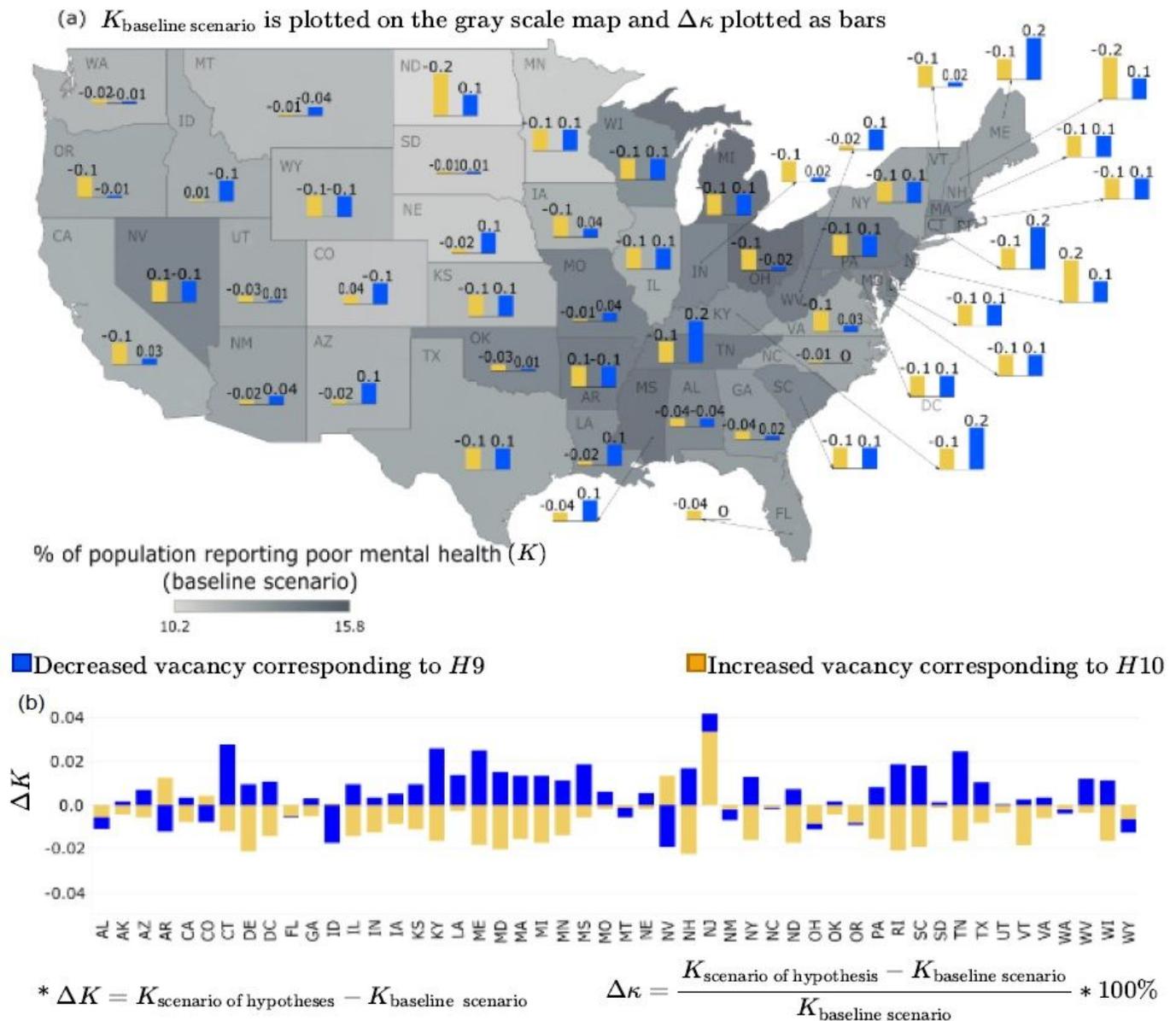


Figure 7

Housing vacancy scenario for % of adults aged > 18 years reporting poor mental health for > 14 days (K):
(a) ΔK 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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SuuplementaryInformation.pdf](#)