

Simulating the impact of greenspace exposure on metabolic biomarkers in a diverse population living in San Diego, California

A g-computation application

Anaïs Teyton^{a,b,c,*}, Nivedita Nukavarapu^d, Noémie Letellier^{c,e}, Dorothy D. Sears^{f,g,h,i}, Jiue-An Yang^d, Marta M. Jankowska^d, Tarik Benmarhnia^{c,e}

Introduction: Growing evidence exists that greenspace exposure can reduce metabolic syndrome risk, a growing public health concern with well-documented inequities across population subgroups. We capitalize on the use of g-computation to simulate the influence of multiple possible interventions on residential greenspace on nine metabolic biomarkers and metabolic syndrome in adults (N = 555) from the 2014–2017 Community of Mine Study living in San Diego County, California.

Methods: Normalized difference vegetation index (NDVI) exposure from 2017 was averaged across a 400-m buffer around the participants' residential addresses. Participants' fasting plasma glucose, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and triglyceride concentrations, systolic and diastolic blood pressure, hemoglobin A1c (%), waist circumference, and metabolic syndrome were assessed as outcomes of interest. Using parametric g-computation, we calculated risk differences for participants being exposed to each decile of the participant NDVI distribution compared to minimum NDVI. Differential health impacts from NDVI exposure by sex, ethnicity, income, and age were examined.

Results: We found that a hypothetical increase in NDVI exposure led to a decrease in hemoglobin A1c (%), glucose, and high-density lipoprotein cholesterol concentrations, an increase in fasting total cholesterol, low-density lipoprotein cholesterol, and triglyceride concentrations, and minimal changes to systolic and diastolic blood pressure, waist circumference, and metabolic syndrome. The impact of NDVI changes was greater in women, Hispanic individuals, and those under 65 years old.

Conclusions: G-computation helps to simulate the potential health benefits of differential NDVI exposure and identifies which subpopulations can benefit most from targeted interventions aimed at minimizing health disparities.

Keywords: G-formula; Causal inference; Greenspace; Normalized difference vegetation index; Social determinants of health; Intervention simulation; Cardiometabolic disease; Metabolic syndrome

^aHerbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, California; ^bSchool of Public Health, San Diego State University, San Diego, California; ^cScripps Institution of Oceanography, University of California, San Diego, La Jolla, California; ^dPopulation Sciences, Beckman Research Institute, City of Hope, Duarte, California; ^eIrset Institut de Recherche en Santé, Environnement et Travail, UMR-S 1085, Inserm, University of Rennes, EHESP, Rennes, France; ^fCollege of Health Solutions, Arizona State University, Phoenix, Arizona; ^gDepartment of Medicine, University of California, San Diego, La Jolla, California; ^hDepartment of Family Medicine, University of California, San Diego, La Jolla, California; and ⁱMoore's Cancer Center, University of California, San Diego, La Jolla, California

M.M.J. and T.B. are cosenior authors.

Supported by the National Cancer Institute (R01CA228147).

Exposure data available upon request. The GitHub repository that includes the analytical code can be found at the following URL: <https://github.com/hdscalecollab/G-computation>.



Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.enviroepidem.com).

*Corresponding Author. Address: Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: ateyton@ucsd.edu (A. Teyton).

Copyright © 2024 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The Environmental Epidemiology. All rights reserved. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Metabolic syndrome is a prominent public health issue in the United States that affects one in three adults, and it is a growing concern as metabolic health has worsened over time.^{1–3} Defining characteristics of metabolic syndrome include high blood pressure, abdominal obesity, high triglyceride levels, low high-density lipoprotein (HDL) cholesterol, and impaired fasting glucose.^{2,4,5} It is well-established that metabolic

What this study adds:

This study improves understanding of the relationship between greenspace exposure and metabolic risk biomarkers and builds evidence for the potential positive effects of increasing green spaces in urban environments. As metabolic syndrome is a growing public health concern in the United States, identifying population-level strategies to reduce this risk is critical. G-computation, a flexible causal inference method, can be utilized to simulate potential targeted interventions on NDVI exposure and observe its possible health benefits. Additionally, effect modification can be assessed with g-computation to identify which subpopulations can benefit most from greening strategies. This has important implications on policy and interventions aimed at reducing the health burden of metabolic syndrome and promoting health equity.

syndrome increases the risk of cardiovascular disease, including coronary heart disease, carotid artery disease, and ischemic stroke, diabetes, and premature cardiovascular-specific and all-cause mortality.^{1,6} Given these severe consequences to health, it is critical to identify who is most at risk and to prioritize reducing this risk, especially as metabolic syndrome is largely preventable.

Certain individuals are at greater risk of metabolic syndrome. For example, older adults (compared to younger adults), Hispanic and non-Hispanic Black individuals (compared to non-Hispanic White individuals), men (compared to women), and those with lower socioeconomic status (compared to those with higher socioeconomic status) have been shown to be at increased risk of metabolic syndrome.^{7–9} The prevalence of metabolic syndrome has increased over time, especially among these subgroups, leading to an increase in avoidable health inequalities.¹⁰ Thus, it is particularly important to focus on populations with elevated risk and identify potential interventions that may reduce these health inequities.

Assessing the influence of modifiable factors on metabolic syndrome is important to reduce the risk of metabolic syndrome and its health consequences. Certain risk factors for metabolic syndrome have been identified, including environmental exposures such as noise and air pollution.^{11,12} A few studies have also assessed the role of greenspace exposure, often measured as normalized difference vegetation index (NDVI), on metabolic syndrome biomarkers, where a protective relationship has been identified.^{13–18} Yet, this literature has mostly focused on estimating a traditional dose–response and has not provided actionable evidence based on interventions that simulate the influence of modifying NDVI exposure on metabolic biomarkers. Simulating interventions on greenspace can be implemented to understand the potential implications regarding changes in health outcomes resulting from changes in residential greenspace.

One approach to simulate the health impacts of potential interventions on NDVI exposure is g-computation. G-computation is an estimation technique used to calculate the average causal effect through standardization.^{19,20} Though a less frequently used causal inference method, g-computation allows for the effects of an intervention to be evaluated by estimating potential outcomes contrasts.^{21,22} It can estimate outcomes under a range of different exposure scenarios and be used to identify heterogeneity in the effects of these exposure scenarios in subgroups.^{21,23} This can help identify salient policies and interventions as the health benefits of multiple scenarios are quantifiable, and subpopulations that stand the most to gain from such interventions can be prioritized.

In this study, capitalizing on the application of g-computation, we simulated the etiological impact of differential greenspace exposure on nine metabolic biomarkers and metabolic syndrome in a cohort of adults living in San Diego County, California. We also assessed the effect modification of modeled relationships by several demographic characteristics to identify which subgroups would benefit most from increasing exposure to greenspace around their home environments.

Methods

Study population

The cross-sectional Community of Mine study aimed to assess the influence of environmental exposures and lifestyle behaviors on cancer risk in an ethnically diverse population living in San Diego County from 2014 to 2017. Participants were

randomly selected from 764 census block groups from the 1794 total census block groups in San Diego County that varied with regard to the highest and lowest terciles of walkability and food access. The study ensured that no more than 10 participants were recruited from any given census block group. To be eligible for this study, participants had to be living for a minimum of 6 months in their current residence, be able to move without human assistance, travel to a study visit, have a phone, read and write fluently in either English or Spanish, provide informed consent, comply with the protocol, and complete all assessments. A total of 602 individuals aged 35–80 years were included in the study. Participants were scheduled for a clinical visit when the eligibility criteria were confirmed, and consent was received. During this clinical visit, blood pressure and a 12-hour fasted blood draw were collected. Additional measurements, such as waist circumference, were taken. At the visit, participants completed a self-report survey providing their demographic information. Further details on the study protocol and design can be found elsewhere.²⁴ This study was approved by the University of California, San Diego Institutional Review Board (protocol #140510), and all participants provided signed informed consent. In this study, a complete case analysis was utilized, where those with missing information related to the NDVI exposure, metabolic biomarker outcomes, and confounders were removed (see Table S1; <http://links.lww.com/EE/A293> regarding descriptive statistics on missingness), for a final sample size of 555 participants (92.2% retention).

Greenspace exposure and metabolic biomarker outcomes

We calculated greenspace exposure by averaging 2017 NDVI for a 400-m buffer around the participants' residences, which is a common and optimal buffer size that has been used in previous studies.^{25,26} NDVI was collected from Google Earth Engine using Landsat 7 Collection 1 Tier 1 data, which has a high spatial resolution of 30 m.²⁷

We assessed nine metabolic biomarkers that are related to the conditions associated with metabolic syndrome. These outcomes were measured and collected during the clinical visit, which included fasting plasma glucose (mg/dl), total cholesterol (mg/dl), low-density lipoprotein (LDL) cholesterol (mg/dl), HDL cholesterol (mg/dl), triglycerides (mg/dl), hemoglobin A1c (%), systolic blood pressure (SBP) (mmHg), diastolic blood pressure (DBP) (mmHg), and waist circumference (cm). We additionally assessed metabolic syndrome (yes; no) as an outcome of interest, where having metabolic syndrome was defined as having at least three of the following criteria: an increased waist circumference (more than 102 cm for men and 88 cm for women), elevated triglycerides (greater or equal to 150 mg/dl), low HDL cholesterol (less than 40 mg/dl in men and 50 mg/dl in women), hypertension (greater than or equal to 130 mmHg SBP or 80 mmHg DBP), and impaired fasting glucose (greater than or equal to 110 mg/dl).

Statistical analysis

G-computation was developed by Robins (1986) and can be used to identify an average causal effect through standardization. Briefly, the mean outcome (Y) in the exposed ($A = 1$) and in the unexposed ($A = 0$) is estimated and standardized across the prevalence of covariates L , as shown in the following equation: $\sum E[Y|A = a, L = l] \times \Pr[L = l]$.

A detailed visualization of the steps of g-computation derived from Hernán and Robins²⁸ is provided in Figure S1; <http://links.lww.com/EE/A293>. There are several key assumptions for g-computation that are typical under the potential outcomes framework, including exchangeability, positivity, and consistency, as well as no measurement error and no model misspecification.

Parametric g-computation was utilized to simulate hypothetical interventions that would increase or decrease NDVI

exposure around the home and observe the impacts of these exposure changes on metabolic biomarkers in participants. Minimum, maximum, and decile values of NDVI were calculated from the exposure distribution in the participants. Ten interventions on NDVI exposure were simulated (e.g., 10th percentile through the maximum, considered the “exposed” groups) and compared to the minimum NDVI exposure (e.g., the “unexposed” group). Generalized linear models were utilized for the metabolic biomarkers (i.e., the nine continuous outcome variables) and a logistic regression model was utilized for metabolic syndrome (i.e., the one binary outcome variable), which adjusted for potential confounders including sex (male; female), income (\$30,000 or less; \$30,000–\$55,000; \$55,000 or more), race (White; Black; Asian; Native Hawaiian, or other Pacific Islander; American Indian/Alaska Native; unknown/other), ethnicity (Hispanic; non-Hispanic), age, and education (less than high school; high school or more). Regression estimates were then used to predict the outcome values if all participants were exposed or unexposed to each of the ten intervention deciles. Risk differences (RDs) between each decile of NDVI exposure starting from 10th percentile and the reference group (minimum NDVI) were calculated from the average predicted outcomes. Bootstrapping was used to obtain the 95% confidence interval (CI) for the RDs using 1000 iterations. A sensitivity analysis was conducted to examine the model misspecification assumption. We utilized a super learner ensemble, which combined and took the weighted average of four machine learning algorithms: generalized linear model with main terms only (SL.glm), random forest (SL.randomForest), elastic net regression (SL.glmnet), and extreme gradient boosting (SL.xgboost) using 10-fold cross-validation.^{29,30} A binomial distribution was specified for the binary metabolic syndrome outcome, and 95% CIs were also obtained using bootstrapping with 1000 iterations. It is important to note that singly robust estimators (such as g-computation) used with machine learning underperform compared to doubly robust estimators with regard to bias, variance, and CI coverage;³¹ however, the purpose of this sensitivity analysis was to compare the model fit of the models used in the main analysis to those of the ensemble models and confirm whether the RD estimates and trends between NDVI exposure and metabolic biomarker outcomes were similar.

We assessed if certain subgroups had more substantial health benefits from NDVI exposure interventions. Effect modification by demographic characteristics including sex, ethnicity, income, and age (categorized as 65 years old or greater; under 65 years old) were examined. For each effect modifier, the stratum of interest (for instance, those 65 years or older) was simulated to be exposed to each of the NDVI exposure levels (e.g., 10th percentile through the maximum), while the other categories (in this case, less than 65 years old) remained exposed to their true residential NDVI exposure value (e.g., the actual average NDVI value within a 400-m buffer around the residence). An interaction term between the given effect modifier and NDVI exposure was added to each model. Cochran Q heterogeneity tests assessed if significant differences existed between the RDs for each effect modifier, where significance was defined with a conservative threshold using a *P* value for heterogeneity at 0.05.

Analysis was conducted using Stata 16.1 (StataCorp LLC, College Station, TX) and R 4.3.1 (R Core Team, Austria, Vienna), and the Super Learner package was used for the sensitivity analysis.²⁹ The analytical code can be found on the following GitHub repository: <https://github.com/hdscalecollab/G-computation>.

Results

Table 1 provides a summary of the demographic characteristics of the study population (*N* = 555). A greater proportion of participants were high school educated or higher (90.5%), had an annual income greater than \$55,000 (49.2%), female (56.0%),

Table 1.

Summary of demographic characteristics of the study population (*N* = 558), distribution of NDVI exposure, and biomarker outcomes

Characteristics	<i>n</i> (%) / mean (SD)
Demographic characteristics	
Age	58.7 (11.0)
Sex	
Female	311 (56.0%)
Male	244 (44.0%)
Income	
\$30,000 or less	154 (27.7%)
\$30,000–\$55,000	128 (23.1%)
\$55,000 or more	273 (49.2%)
Race	
White	393 (70.8%)
Black	18 (3.2%)
Asian	21 (3.8%)
American Indian/Alaska Native	24 (4.3%)
Native Hawaiian or other Pacific Islander	7 (1.3%)
Unknown/other	92 (16.6%)
Ethnicity	
Non-Hispanic	322 (58.0%)
Hispanic	233 (42.0%)
Education	
Less than high school	53 (9.5%)
High school or more	502 (90.5%)
Exposure	
Normalized difference vegetation index (NDVI)	0.17 (0.1)
Fasting plasma biomarkers, waist circumference, and METS	
Glucose (mg/dl)	103.7 (27.5)
Total cholesterol (mg/dl)	188.2 (35.1)
HDL cholesterol (mg/dl)	59.0 (17.4)
LDL cholesterol (mg/dl)	107.8 (30.4)
Triglycerides (mg/dl)	107.1 (51.6)
Systolic blood pressure (mmHg)	126.7 (17.6)
Diastolic blood pressure (mmHg)	72.7 (10.1)
Hemoglobin A1c (%)	5.6 (0.9)
Waist circumference (cm)	91.2 (20.9)
Metabolic syndrome	
Yes	105 (18.9%)
No	450 (81.1%)

Frequencies and percentages are provided for categorical variables, while means and standard deviations are provided for continuous variables.

White (70.8%), non-Hispanic (58.0%), and under the age of 65 (68.7%). About 18.9% of participants met the criteria for metabolic syndrome. Additionally, fasting plasma total cholesterol and glucose concentrations were above the normal range, whereas average LDL cholesterol, triglycerides, hemoglobin A1c, and SBP and DBP were within the normal range. Average waist circumference was also within the normal range for both male and female participants (average female waist circumference: 86.8 cm; average male waist circumference: 96.9 cm).

Figures S2a–S2k; <http://links.lww.com/EE/A293> provide maps depicting average NDVI exposure and health outcomes at the ZIP code level. NDVI exposure tended to be higher in northern San Diego County and lower in the southern region of the county. Higher fasting glucose concentrations seemed to be more concentrated in southern San Diego, compared to other regions, while triglyceride concentrations, waist circumference, and metabolic syndrome were higher in northern San Diego. HDL cholesterol concentrations were highest in both northern and southern San Diego, particularly inland. In contrast, greater spatial heterogeneity can be seen for LDL cholesterol, total cholesterol, SBP and DBP, and hemoglobin A1c, where patterns are more difficult to identify.

The RDs for the relationship between NDVI exposure and metabolic health outcomes from the g-computation results for the entire population are provided in Figure 1, and Figures 2–4

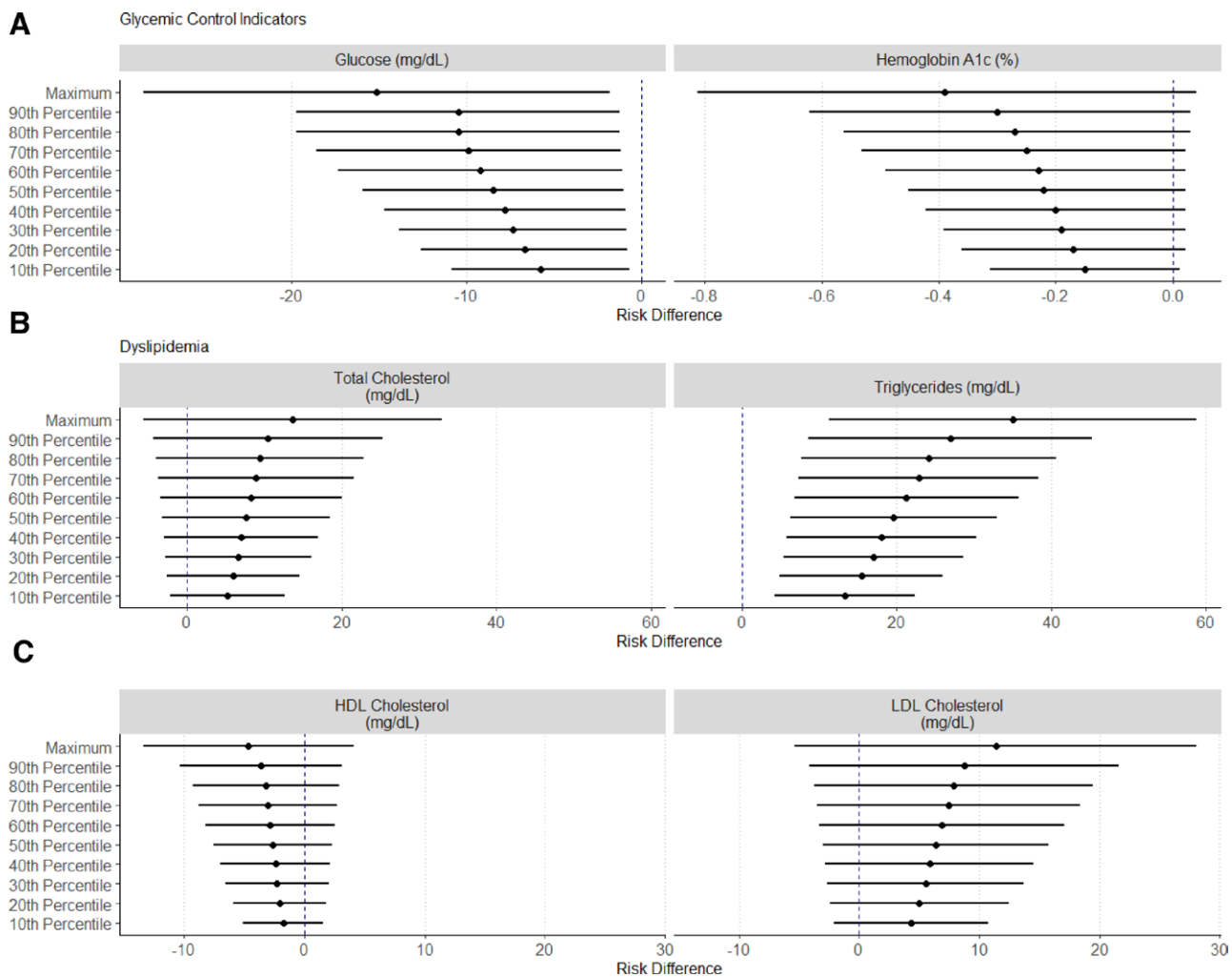


Figure 1. Risk differences (RDs) of average biomarker changes with simulated changes to NDVI exposure for the total population. RDs are shown as points and 95% CIs as lines on the x-axis and simulated NDVI exposure on the y-axis as deciles compared to minimum NDVI exposure for the glycemic control indicators (A: fasting glucose levels and hemoglobin A1c), dyslipidemia (B: total cholesterol and triglycerides; C: HDL cholesterol and LDL cholesterol), blood pressure (D: diastolic blood pressure and systolic blood pressure), and waist circumference and METS (E) outcomes.

provide the significant effect modification results (tabulated results are provided in Tables S2–S5; <http://links.lww.com/EE/A293>, and all effect modification results are displayed in Figures S3–S6; <http://links.lww.com/EE/A293>). The following NDVI exposure groups were used: the minimum (−0.082; used as the reference group), the 10th percentile (0.087), the 20th percentile (0.114), the 30th percentile (0.134), the 40th percentile (0.147), the 50th percentile (0.166), the 60th percentile (0.188), the 70th percentile (0.208), the 80th percentile (0.225), the 90th percentile (0.260), and the maximum (0.362). For the entire population, a simulated increase in NDVI exposure led to a decrease in glucose concentrations (maximum NDVI RD: −15.16 mg/dl; 95% CI: −28.44, −1.87), hemoglobin A1c (maximum NDVI RD: −0.39%; 95% CI: −0.81, 0.04), and HDL cholesterol (maximum NDVI RD: −4.68 mg/dl; 95% CI: −13.35, 4.00). In contrast, an increase in NDVI exposure increased total cholesterol (maximum NDVI RD: 13.61 mg/dl; 95% CI: −5.59, 32.80), LDL cholesterol (maximum NDVI RD: 11.37 mg/dl; 95% CI: −5.30, 28.04), and triglyceride concentrations (maximum NDVI RD: 35.07 mg/dl; 95% CI: 11.37, 58.77). With increasing NDVI exposure, SBP (maximum NDVI RD: −2.36 mmHg; 95% CI: −11.75, 7.04), DBP (maximum NDVI RD: 0.34 mmHg; 95% CI: −4.95, 5.62), waist circumference (maximum NDVI RD: 0.46 cm; 95% CI: −11.28, 12.19), and metabolic syndrome (maximum NDVI RD: 0.25; 95% CI: −1.24,

1.74) changed minimally and remained close to the null. The sensitivity analysis using the super learner ensemble is depicted in Figure S7; <http://links.lww.com/EE/A293> (tabulated results are provided in Table S6; <http://links.lww.com/EE/A293>). Overall, a slight attenuation in the strength of the RDs can be noted for most metabolic biomarkers, including LDL cholesterol, HDL cholesterol, triglycerides, total cholesterol, glucose concentrations, hemoglobin A1c, and metabolic syndrome. The positive relationship between NDVI exposure and waist circumference and the negative relationship for SBP were stronger in the super learner model. The RDs for DBP went from slightly positive to slightly negative for the main model compared to the super learner model. With some slight but insignificant non-linearities noted in the super learner models, most of the metabolic biomarker trends with increasing NDVI exposure tended to remain the same as the main models.

We identified significant heterogeneous effects by sex, ethnicity, and age across certain health outcomes. Differences by sex were noted for fasting plasma glucose, hemoglobin A1c, and waist circumference. For these three metabolic health outcomes, a positive, increasing relationship was identified for men with increasing NDVI exposure (glucose maximum NDVI RD: −1.40 mg/dl; 95% CI: −9.39, 6.59; hemoglobin A1c maximum NDVI RD: 0.01%; 95% CI: −0.25, 0.27; waist circumference maximum NDVI RD: 5.00; 95% CI: −1.79, 11.80), whereas

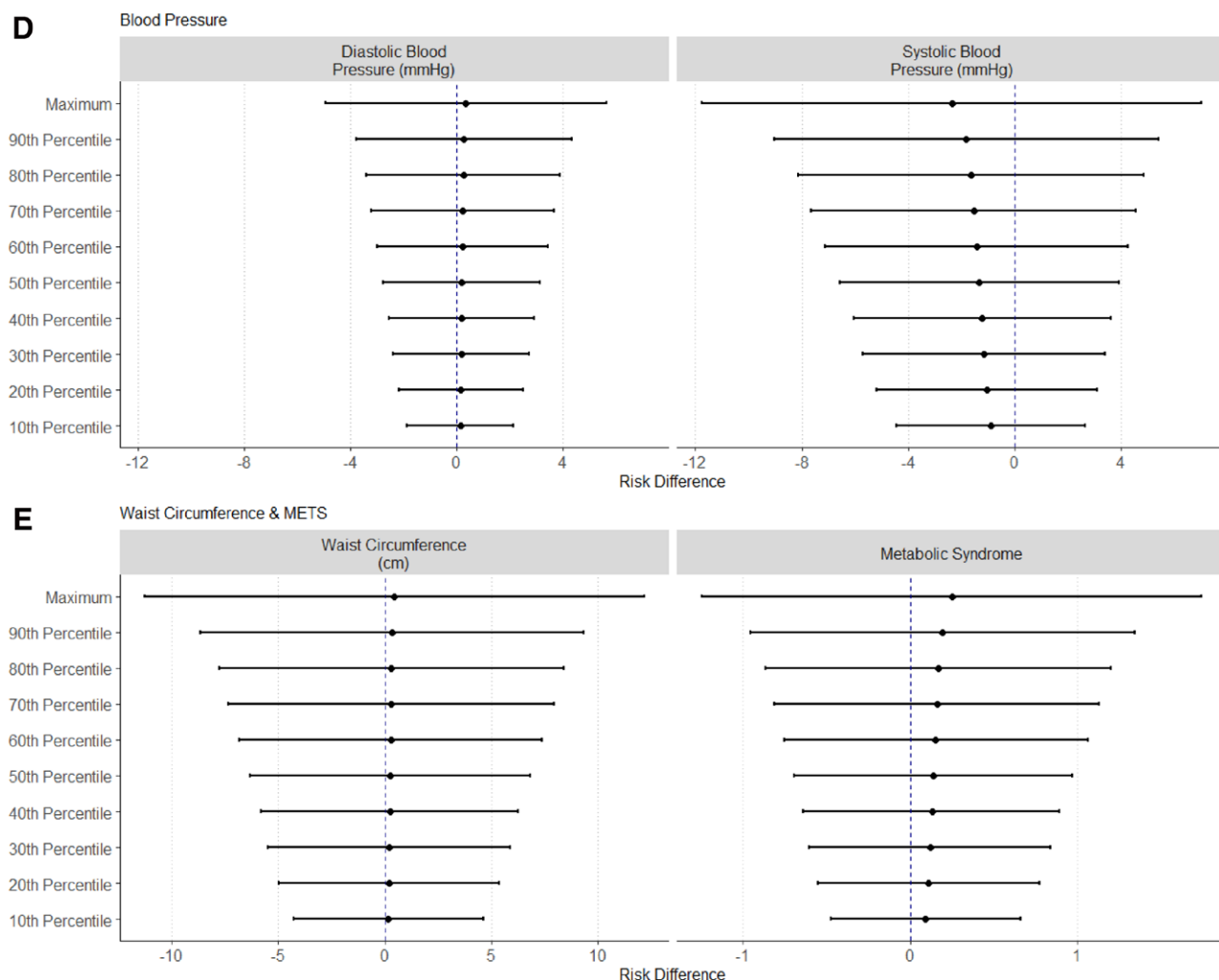


Figure 1. Continued

a negative, decreasing relationship was identified for women (glucose maximum NDVI RD: -15.63 mg/dl; 95% CI: $-26/23$, -5.04 ; hemoglobin A1c maximum NDVI RD: -0.46% ; 95% CI: -0.84 , -0.08 ; waist circumference maximum NDVI RD: -6.26 ; 95% CI: -14.95 , 2.44). Hispanic and non-Hispanic participants differed in their relationship between NDVI exposure and hemoglobin A1c, where a negative and decreasing relationship was observed for Hispanic participants (maximum NDVI RD: -0.72% ; 95% CI: -1.15 , -0.30), and a positive and slightly increasing relationship was found for non-Hispanic participants (maximum NDVI RD: 0.11% ; 95% CI: -0.12 , 0.34). Moreover, HDL cholesterol was on the verge of being statistically significant (P values ranging from 0.0500 to 0.0505), where HDL concentrations increased with increasing NDVI for Hispanic individuals (maximum NDVI RD: 3.04 mg/dl; 95% CI: -2.96 , 9.04), and HDL concentrations decreased with increasing NDVI for non-Hispanic individuals (maximum NDVI RD: -5.68 mg/dl; 95% CI: -12.02 , 0.65). Finally, differences by age were noted by total cholesterol and triglyceride concentrations. Increasing NDVI exposure led to a negative and slightly decreasing cholesterol as well as positive and increasing triglyceride concentrations for those 65 and older (total cholesterol maximum NDVI RD: -0.91 mg/dl; 95% CI: -8.83 , 7.02 ; triglyceride maximum NDVI RD: 6.14 mg/dl; 95% CI: -2.18 , 14.46), while a positive and increasing relationship was identified for both biomarkers for those under 65 years old (total cholesterol maximum NDVI RD: 20.74 mg/dl; 95% CI: 3.27 , 38.22 ; triglyceride maximum NDVI RD: 34.77 mg/dl; 95% CI: 11.18 , 58.36). In comparison,

no significant differences by income were observed for any of the health outcomes.

Discussion

We utilized g-computation to simulate increasing NDVI exposure levels and modeled their effects on nine metabolic biomarkers and metabolic syndrome in a diverse cohort. We found that increasing NDVI exposure led to a reduction in fasting plasma glucose, HDL cholesterol, and hemoglobin A1c and an increase in total cholesterol, LDL cholesterol, and triglyceride concentrations across all participants. In contrast, increasing NDVI exposure minimally impacted SBP, DBP, waist circumference, and metabolic syndrome. Differences in the influence of NDVI exposure on certain biomarkers were identified by sex, ethnicity, and age.

We observed that indicators of glycemic control (e.g., fasting glucose and hemoglobin A1c levels) decreased with increasing NDVI exposure. Several studies support this finding, as they identified a decrease in hemoglobin A1c^{32–34} and a decrease in fasting glucose^{33,35,36} with an increase in NDVI exposure. In contrast, contrary to our hypotheses, biomarkers related to dyslipidemia (e.g., high total cholesterol, LDL cholesterol, and triglyceride levels as well as low HDL cholesterol) increased with increasing NDVI exposure. Mixed findings have been previously identified. Yang et al³⁷ and Mei et al³⁸ identified a negative relationship between NDVI exposure and these blood lipids. In contrast, a study found no association between NDVI

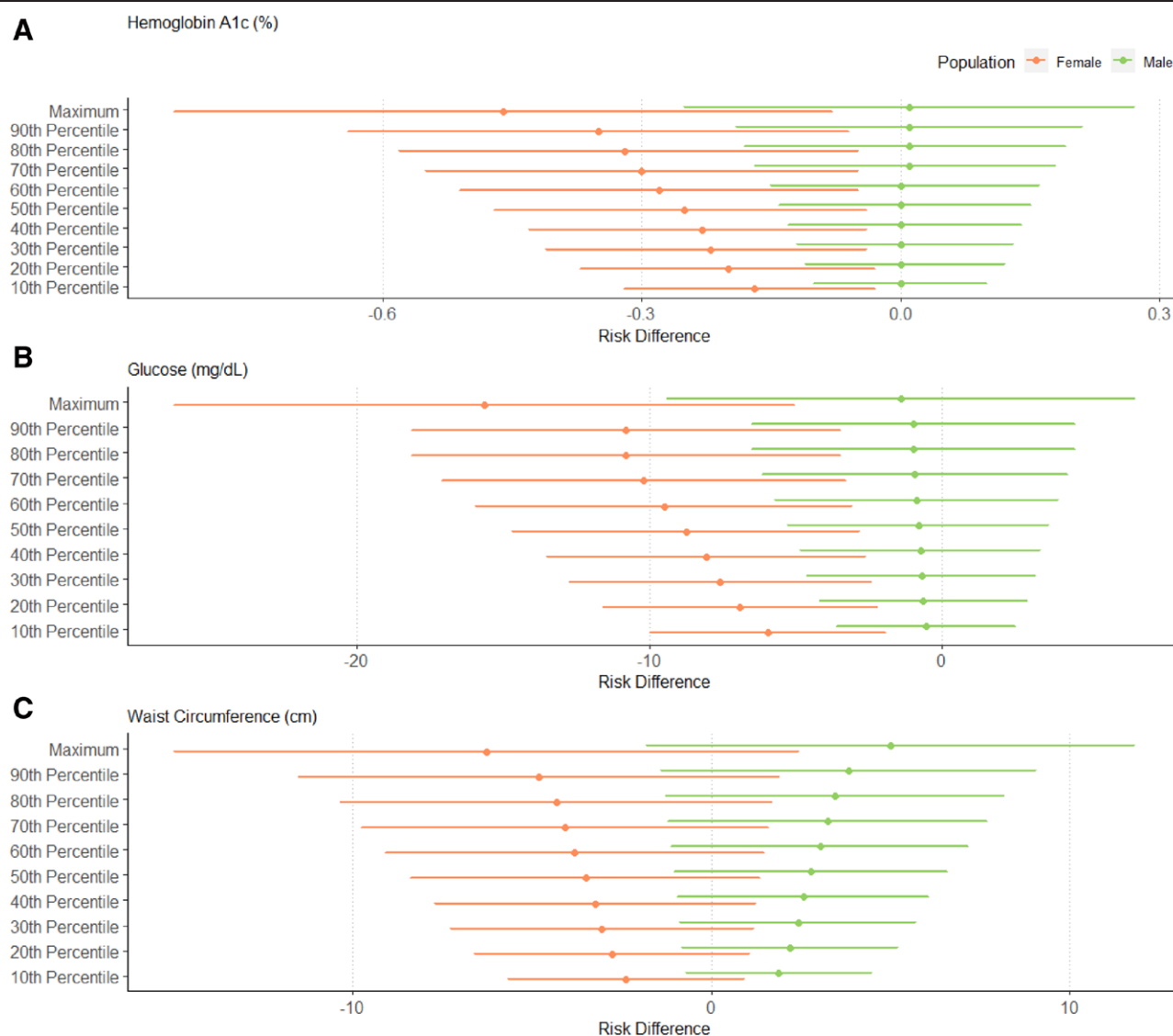


Figure 2. Risk differences (RDs) of average biomarker changes with simulated changes to NDVI exposure, stratified by sex. RDs are shown as points and 95% CIs as lines on the x-axis and simulated NDVI exposure on the y-axis as deciles compared to minimum NDVI exposure for hemoglobin A1c (A), glucose (B), and waist circumference (C). Effect modification by sex is depicted, where males are shown in green and females are shown in orange.

exposure and total and LDL cholesterol in children.³⁹ A separate study also found close to null effects of NDVI exposure on total cholesterol and triglycerides, specifically when they assessed NDVI exposure around a 500-m buffer.⁴⁰ Another study found a decrease in total cholesterol and an increase in LDL cholesterol with increasing NDVI exposure.⁴¹ Given these mixed findings, additional studies are needed to better understand the relationship between NDVI exposure and biomarkers related to dyslipidemia. Moreover, we found that SBP decreased with increasing NDVI exposure, while DBP was not strongly influenced by NDVI exposure. A systematic review identified a small protective influence of NDVI exposure on blood pressure, where SBP decreased by -0.77 mmHg and DBP decreased by -0.32 mmHg with a 0.1 unit increase in NDVI exposure.⁴² Several other studies found similar results as well.^{43–45} One study identified that greenspace exposure was protective of both high and low blood pressure.⁴⁶ Moreover, contrary to our findings, a systematic review assessing the relationship between greenspace exposure and metabolic syndrome identified 18 papers examining this association, and they observed an odds ratio of 0.90 (95% CI: 0.87, 0.93) for the influence of 500-m NDVI on the risk of metabolic syndrome.¹⁷ Therefore, while our findings tend to be consistent with previous studies, such as for biomarkers

related to glycemic control, more studies are needed as mixed results have been identified. Patwary et al emphasized the fact that few studies have examined the overall association between greenspace exposure and metabolic syndrome risk, making this a critical relationship to examine. Few studies have investigated the underlying mechanisms through which a greener environment influences metabolic biomarkers. One such study found that physical activity mediated the relationship between NDVI and fasting glucose,³³ and two other studies identified that both air pollution and body mass index mediated the relationship between NDVI and biomarkers related to dyslipidemia.^{37,41} Another study identified physical activity, body mass index, and air pollution including $PM_{2.5}$, PM_{10} , nitrogen dioxide, and sulfur dioxide, as mediators in the relationship between exposure to greenness and metabolic syndrome risk.⁴⁷ Additionally, it may be possible that increased green space improves social capital and reduces stress, which may in turn reduce the risk of metabolic syndrome.^{13,48–51} However, given the limited focus of previous literature on this topic, future work should improve understanding of the underlying mechanisms that result in greenspace impacts on these biomarkers and metabolic syndrome.

We also identified effect modification by sex, ethnicity, and age for the relationship between NDVI exposure and certain

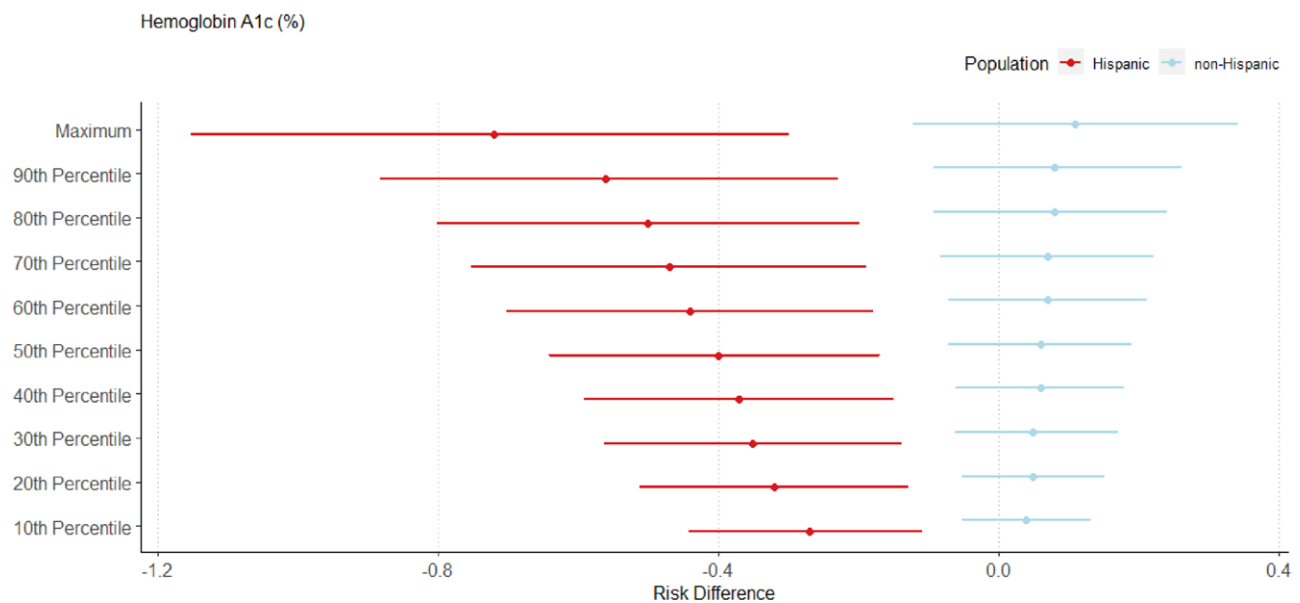


Figure 3. Risk differences (RDs) of average biomarker changes with simulated changes to NDVI exposure, stratified by ethnicity. RDs are shown as points and 95% CIs as lines on the x-axis and simulated NDVI exposure on the y-axis as deciles compared to minimum NDVI exposure for hemoglobin A1c. Effect modification by ethnicity is depicted, where Hispanic participants are shown in red and non-Hispanic participants are shown in blue.

biomarkers. We found beneficial impacts of increasing NDVI exposure for women on their hemoglobin A1c, fasting glucose concentrations, and waist circumference as well as for Hispanic individuals on their hemoglobin A1c. Additionally, counter to our hypotheses, increasing NDVI exposure contributed to higher total cholesterol and triglycerides for those less than 65 years old. A few studies have assessed the effect of modification on this relationship; however, mixed results have been identified. In contrast with our study, Li et al³⁵ identified a stronger negative relationship in men and older adults between NDVI exposure and fasting glucose. While we did not identify significant effect modification by sex on the relationship between NDVI and dyslipidemia-related biomarkers, inconsistent results exist, where one study identified a stronger negative relationship for men,⁴¹ while another identified a stronger negative relationship for women.⁴⁰ A stronger negative relationship was also found for those with a higher education level and older adults.^{37,40} We also did not identify effect modification in the relationship between NDVI exposure and blood pressure; however, studies found higher negative relationships for men, smokers, drinkers, those exposed to lower PM_{2.5}, younger individuals, and those who are overweight or obese.^{43–46} Mixed findings have been previously identified regarding NDVI exposure and dyslipidemia-related biomarkers by age. One study similarly identified an increase in NDVI led to lower HDL cholesterol and higher triglyceride levels in individuals living in London, United Kingdom, aged 45–69 years.¹³ In contrast, another study found that for every one standard deviation increase in NDVI, HDL cholesterol increased and triglycerides decreased in older adults living in Shenzhen, China.⁵² Given that few studies have assessed this relationship and mixed results have been identified, additional studies are needed to understand the impact of NDVI exposure on dyslipidemia-related biomarkers and the possible underlying mechanisms by which age plays a role in this relationship. Moreover, no studies to date have investigated the modification of this relationship between NDVI exposure and metabolic biomarkers by ethnicity. Improved understanding regarding the factors that contribute to differential associations by subgroups is necessary. It is possible that social and economic disadvantage and discrimination may contribute to certain subgroups having a greater risk of metabolic syndrome or green space

deprivation,⁵³ making simulated increased green space more beneficial in these subgroups compared to their counterparts; however, more research is needed to better understand these differential subgroup associations. Additionally, further studies are needed to assess this relationship in other regions and to identify whether certain subpopulations benefit more greatly from increasing greenspace exposure. This has implications for potential targeted interventions, as policies may be able to prioritize subgroups that benefit most from greening strategies.

Our study illustrates the advantages of utilizing g-computation, as it flexibly estimates the difference in the outcomes if all participants were exposed to the intervention compared to all participants being unexposed to the intervention. We assessed a diverse set of hypothetical changes in NDVI exposure on metabolic biomarkers, which can be useful for potential interventions and stakeholders. We selected these specific scenarios to have a wide range of simulated NDVI exposures to understand this etiological relationship; however, it should be noted that this approach can be used to reflect any type of intervention and can be applied to other potential policy interventions. Moreover, g-computation can be utilized to simulate the differential impacts of an intervention on health outcomes in various subpopulations. A g-computation approach that considers effect modification can aid in identifying potential inequities, where a targeted intervention can be simulated on a specific subpopulation. This has implications for tailored policies, as the most effective greening interventions can be identified, and the subpopulations that benefit most from these interventions can be prioritized. Several strengths exist for this study. This study was conducted in the ethnically diverse Community of Mine sample that includes a high proportion of Hispanic individuals with several collected metabolic risk biomarkers. Furthermore, the sampling frame was specifically designed to ensure high environmental variability across the San Diego region. However, this study also has limitations. It is possible that residual or unmeasured confounding exists, which can potentially violate the exchangeability and positivity assumptions needed to causally interpret the g-computation findings. One such example of residual confounding is structural racism and discrimination targeted toward marginalized and underrepresented populations, which are not controlled for by solely adjusting on race and ethnicity.^{54–56} Moreover, we

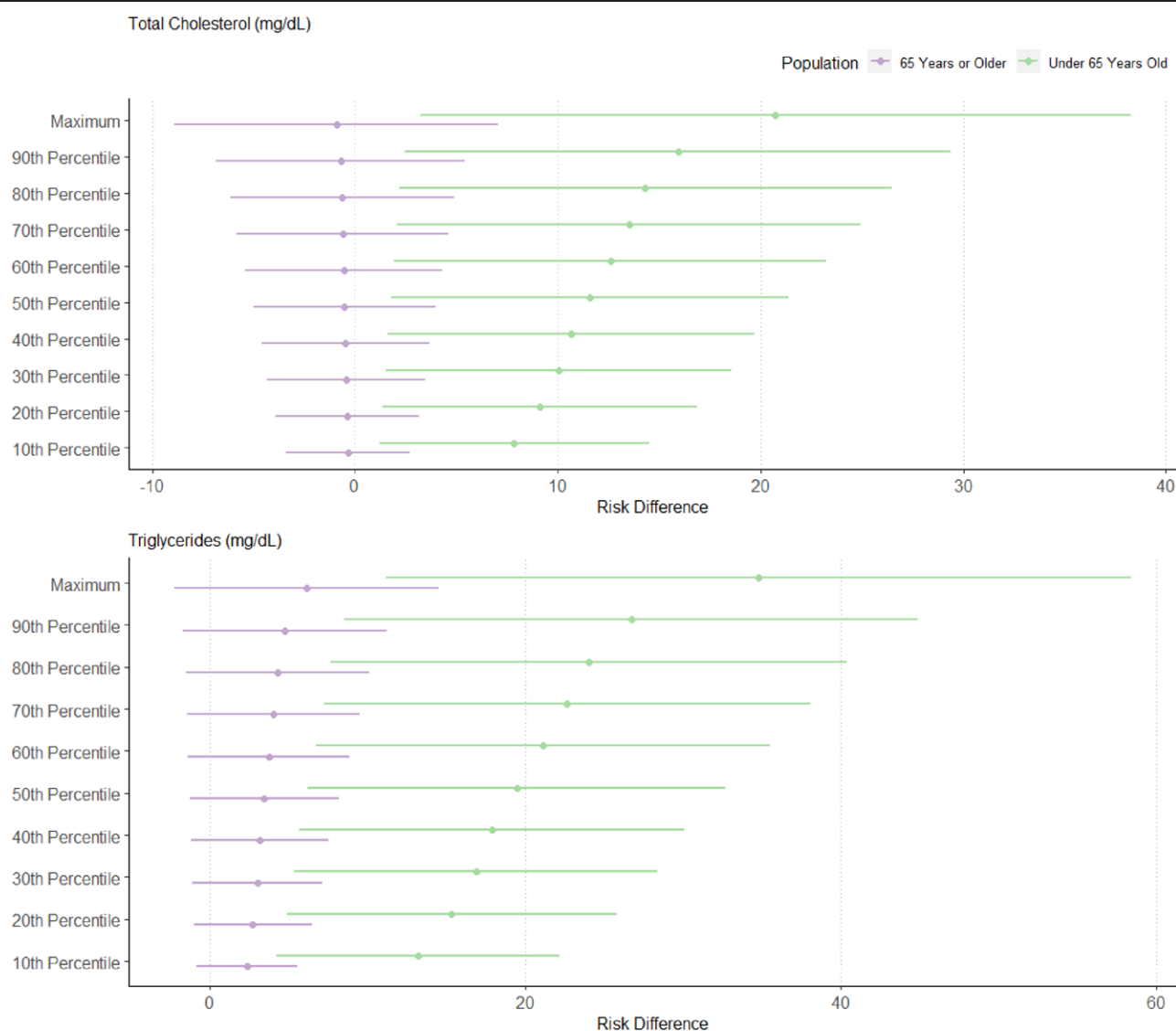


Figure 4. Risk differences (RDs) of average biomarker changes with simulated changes to NDVI exposure, stratified by age. RDs are shown as points and 95% CIs as lines on the x-axis and simulated NDVI exposure on the y-axis as deciles compared to minimum NDVI exposure for total cholesterol and triglycerides. Effect modification by age is depicted, where those 65 years or older are shown in purple and those under 65 years old are shown in green.

used static NDVI exposure, where NDVI was averaged using a 400-m buffer around the residence. First, a range of buffer sizes have been used in the literature, such as those within a 500-m buffer. It may be worthwhile to explore whether differences in the associations exist based on the utilized buffer size. Second, while this static metric using such a buffer has previously been shown to be important to examine, the risk of misclassification and measurement error (another assumption needed for g-computation) is reduced when considering dynamic NDVI exposure that incorporates participants' mobility and daily trajectories. It is important to note, however, that violations to the other g-computation assumptions, including consistency, measurement error (with regards to the outcomes), and model misspecification, are minimal, as a result of our use of well-defined NDVI exposures, the outcomes being measured and collected during a clinical visit, and our sensitivity analysis confirming similar results between our main models and our super learner ensemble models. This study was also performed in a relatively small and geographically specific sample, with limited diversity with regard to race and education, which may impact the generalizability of these findings. This sample of participants was also older, more female, more educated, more ethnically

diverse, had a different racial composition (e.g., a greater proportion of American Indian/Alaska Native individuals and a lower proportion of non-Hispanic Black and non-Hispanic Asian individuals), and had a similar income composition compared to San Diego County, which may also have consequences on the generalizability of these findings.³⁷ Thus, future studies should continue to evaluate the public health implications of NDVI exposure, simulate the impact of dynamic NDVI exposure on metabolic biomarkers using g-computation, and assess this relationship in other populations or regions, particularly in racially diverse groups.

Conclusions

In this study, we simulated the impact of greenspace interventions on several metabolic biomarkers and metabolic syndrome in a population living in San Diego County. Increasing NDVI exposure was found to lower hemoglobin A1c, fasting plasma glucose, and HDL cholesterol concentrations and increase total cholesterol, LDL cholesterol, HDL cholesterol, and triglyceride concentrations across all participants. Moreover, modification of this relationship by sex, ethnicity,

and age was identified. Specifically, increasing NDVI exposure was found to reduce hemoglobin A1c and fasting plasma glucose for women but not men, hemoglobin A1c for Hispanic but not non-Hispanic participants, and increasing NDVI exposure increased total cholesterol and triglycerides for participants under 65 but not participants 65 years and older. By simulating these targeted interventions, the greening strategies with the most health benefits can be identified and the subpopulations that benefit most from these interventions can be prioritized.

Conflicts of interest statement

The authors declare that they have no conflicts of interest with regard to the content of this report.

Acknowledgments

We thank Anna Dimitrova for her guidance on data visualization.

References

- Cheng X, Ma T, Ouyang F, Zhang G, Bai Y. Trends in the prevalence of cardiometabolic multimorbidity in the United States, 1999–2018. *Int J Environ Res Public Health*. 2022;19:4726.
- National Heart, Lung and BL. Metabolic syndrome: what is metabolic syndrome? 2022. Available at: <https://www.nhlbi.nih.gov/health/metabolic-syndrome#:~:text=Metabolic%20syndrome%20is%20common%20in,3%20adults%20have%20metabolic%20syndrome>. Accessed 28 November 2023.
- Saklayen MG. The global epidemic of the metabolic syndrome. *Curr Hypertens Rep*. 2018;20:12.
- Kirk EP, Klein S. Pathogenesis and pathophysiology of the cardiometabolic syndrome. *J Clin Hypertens (Greenwich)*. 2009;11:761–765.
- Williams VF, Oh G-T, Stahlman S. Incidence and prevalence of the metabolic syndrome using ICD-9 and ICD-10 diagnostic codes, active component, U.S. armed forces, 2002–2017. *MSMR*. 2018;25:20–25.
- Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults findings from the third national health and nutrition examination survey. *JAMA*. 2002;287:356–359.
- Shin D, Kongpakpaisarn K, Bohra C. Trends in the prevalence of metabolic syndrome and its components in the United States 2007–2014. *Int J Cardiol*. 2018;259:216–219.
- Gerdts E, Regitz-Zagrosek V. Sex differences in cardiometabolic disorders. *Nat Med*. 2019;25:1657–1666.
- Shi S, Huang H, Huang Y, Zhong VW, Feng N. Lifestyle behaviors and cardiometabolic diseases by race and ethnicity and social risk factors among US young adults, 2011 to 2018. *J Am Heart Assoc*. 2023;12:e028926.
- O'Hearn M, Lauren BN, Wong JB, Kim DD, Mozaffarian D. Trends and disparities in cardiometabolic health among U.S. adults, 1999–2018. *J Am Coll Cardiol*. 2022;80:138–151.
- Münzel T, Sørensen M, Gori T, et al. Environmental stressors and cardio-metabolic disease: part II-mechanistic insights. *Eur Heart J*. 2017;38:557–564.
- Ahmed CMS, Jiang H, Chen JY, Lin Y-H. Traffic-related particulate matter and cardiometabolic syndrome: a review. *Atmosphere*. 2018;9:336.
- de Keijzer C, Basagaña X, Tonne C, et al. Long-term exposure to green-space and metabolic syndrome: a Whitehall II study. *Environ Pollut*. 2019;255:113231.
- Campero MN, Scavuzzo CM, Andreo V, et al. A geospatial analysis of cardiometabolic diseases and their risk factors considering environmental features in a mid-sized city in Argentina. *Geospat Health*. 2023;18. doi:10.4081/gh.2023.1212.
- Yang B-Y, Hu L-W, Jalaludin B, et al. Association between residential greenness, cardiometabolic disorders, and cardiovascular disease among adults in China. *JAMA Netw Open*. 2020;3:e2017507–e2017507.
- Yang B-Y, Liu K-K, Markevych I, et al. Association between residential greenness and metabolic syndrome in Chinese adults. *Environ Int*. 2020;135:105388.
- Patwary MM, Sakhvidi MJZ, Ashraf S, et al. Impact of green space and built environment on metabolic syndrome: a systematic review with meta-analysis. *Sci Total Environ*. 2024;923:170977.
- Bikomeye JC, Beyer AM, Kwarteng JL, Beyer KMM. Greenspace, inflammation, cardiovascular health, and cancer: a review and conceptual framework for greenspace in cardio-oncology research. *Int J Environ Res Public Health*. 2022;19:2426.
- Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7:1393–1512.
- Wang A, Nianogo RA, Arah OA. G-computation of average treatment effects on the treated and the untreated. *BMC Med Res Methodol*. 2017;17:3.
- Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol*. 2017;46:756–762.
- Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol*. 2011;173:731–738.
- Breskin A, Edmonds A, Cole SR, et al. G-computation for policy-relevant effects of interventions on time-to-event outcomes. *Int J Epidemiol*. 2020;49:2021–2029.
- Jankowska MM, Sears DD, Natarajan L, et al. Protocol for a cross sectional study of cancer risk, environmental exposures and lifestyle behaviors in a diverse community sample: the community of mine study. *BMC Public Health*. 2019;19:186.
- Hou J, Tu R, Dong Y, et al. Associations of residing greenness and long-term exposure to air pollution with glucose homeostasis markers. *Sci Total Environ*. 2021;776:145834.
- Labib SM, Lindley S, Huck JJ. Scale effects in remotely sensed greenspace metrics and how to mitigate them for environmental health exposure assessment. *Comput Environ Urban Syst*. 2020;82:101501.
- Chander G, Markham BL, Helder DL. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens Environ*. 2009;113:893–903.
- Hernán MA, Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC; 2020.
- Polley E, van der Laan MJ. SuperLearner: super learner prediction, 2013. 2023:0–2. Available at: <https://cran.r-project.org/web/packages/SuperLearner/SuperLearner.pdf>. Accessed 29 May 2024.
- Phillips RV, Van Der Laan MJ, Lee H, Gruber S. Practical considerations for specifying a super learner. *Int J Epidemiol*. 2023;52:1276–1285.
- Zivich PN, Breskin A. Machine learning for causal inference: on the use of cross-fit estimators. *Epidemiology*. 2021;32:393–401.
- Ma J, Li D, Xie J, Tian Y. Effects of residential greenness and genetic predisposition on hemoglobin A1c and type 2 diabetes: gene-environment interaction analysis from a nationwide study. *Environ Res*. 2023;228:115830.
- Lin B-C, Yen Y-T, Lao XQ, Chen Y-H, Chan T-C. Association between neighborhood greenspace and fasting plasma glucose from a large cohort study in Taiwan. *Urban For Urban Green*. 2019;44:126439.
- Lin L, Dong G, Tang N, Chen Y, Cai L. Association between residential greenness and glycosylated hemoglobin in pregnant women: findings from the baseline data of Yuexiu birth cohort. *Int J Hyg Environ Health*. 2021;234:113721.
- Li R, Chen G, Jiao A, et al. Residential green and blue spaces and type 2 diabetes mellitus: a population-based health study in China. *Toxics*. 2021;9:11.
- Liao J, Chen X, Xu S, et al. Effect of residential exposure to green space on maternal blood glucose levels, impaired glucose tolerance, and gestational diabetes mellitus. *Environ Res*. 2019;176:108526.
- Yang B-Y, Markevych I, Heinrich J, et al. Residential greenness and blood lipids in urban-dwelling adults: the 33 communities Chinese health study. *Environ Pollut*. 2019;250:14–22.
- Mei Y, Li A, Zhao J, et al. Association of long-term exposure to air pollution and residential greenness with lipid profile: mediating role of inflammation. *Ecotoxicol Environ Saf*. 2023;257:114920.
- Markevych I, Standl M, Sugiri D, et al. Residential greenness and blood lipids in children: a longitudinal analysis in GINIplus and LISAPLUS. *Environ Res*. 2016;151:168–173.
- Fan S, Yang B-Y, Xue Z, et al. Associations between residential greenness and blood lipids in Chinese Uyghur adults. *Environ Int*. 2020;142:105903.
- Jiang J, Mao S, Xie Y, et al. Is residential greenness associated with dyslipidemia and lipid levels in Chinese rural-dwelling adults? The Henan rural cohort study. *Environ Sci Pollut Res Int*. 2022;29:5852–5862.
- Zhao Y, Bao W-W, Yang B-Y, et al. Association between greenspace and blood pressure: a systematic review and meta-analysis. *Sci Total Environ*. 2022;817:152513.
- Jiang J, Chen G, Li B, et al. Associations of residential greenness with hypertension and blood pressure in a Chinese rural population: a cross-sectional study. *Environ Sci Pollut Res Int*. 2021;28:51693–51701.

44. Jiang J, Chen G, Li B, et al. Does long-term green space exposure improve hypertension and blood pressure? The Henan rural cohort study. *SSRN Electr J*. 2020. doi:10.2139/ssrn.3667846.
45. Yang T, Wang J, Xu Z, et al. Associations between greenness and blood pressure and hypertension in Chinese middle-aged and elderly population: a longitudinal study. *Environ Res*. 2022;212:113558.
46. Dzhambov AM, Markevych I, Lercher P. Greenspace seems protective of both high and low blood pressure among residents of an alpine valley. *Environ Int*. 2018;121:443–452.
47. Li X, Wang Q, Feng C, et al. Associations and pathways between residential greenness and metabolic syndromes in Fujian Province. *Front Public Health*. 2022;10:1014380.
48. Yang YC, Li T, Ji Y. Impact of social integration on metabolic functions: evidence from a nationally representative longitudinal study of US older adults. *BMC Public Health*. 2013;13:1210.
49. Gong Y, Palmer S, Gallacher J, Marsden T, Fone D. A systematic review of the relationship between objective measurements of the urban environment and psychological distress. *Environ Int*. 2016;96:48–57.
50. Bergmann N, Gyntelberg F, Faber J. The appraisal of chronic stress and the development of the metabolic syndrome: a systematic review of prospective cohort studies. *Endocr Connect*. 2014;3:R55–R80.
51. Hong A, Sallis JF, King AC, et al. Linking green space to neighborhood social capital in older adults: the role of perceived safety. *Soc Sci Med*. 2018;207:38–45.
52. Xu J, Yuan X, Ni W, et al. Associations between residential greenness and blood lipids in Chinese elderly population. *J Endocrinol Invest*. 2022;45:2329–2339.
53. Karlsen S, Nazroo JY. Measuring and analyzing “race,” racism, and racial discrimination. *Methods Soc Epidemiol*. 2006;1:86–111.
54. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. *Lancet*. 2017;389:1453–1463.
55. Adkins-Jackson PB, Chantarat T, Bailey ZD, Ponce NA. Measuring structural racism: a guide for epidemiologists and other health researchers. *Am J Epidemiol*. 2022;191:539–547.
56. Dean LT, Thorpe RJ Jr. What structural racism is (or is not) and how to measure it: clarity for public health and medical researchers. *Am J Epidemiol*. 2022;191:1521–1526.
57. 2017 City Demographic Profiles: San Diego County. County of San Diego, Health & Human Services Agency, Public Health Services, Community Health Statistic Unit. 2020. Available at: <https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/phs/CHS/demographics/2017%20City%20Demographic%20Profiles.pdf>. Accessed 29 May 2024.