

## Contextual Correlates of Physical Activity among Older Adults: A Neighborhood Environment-Wide Association Study (NE-WAS)

Stephen J. Mooney<sup>1</sup>, Spruha Joshi<sup>2</sup>, Magdalena Cerda<sup>3</sup>, Gary J. Kennedy<sup>4</sup>, John R. Beard<sup>5</sup>, and Andrew G. Rundle<sup>6</sup>



### Abstract

**Background:** Few older adults achieve recommended physical activity levels. We conducted a "neighborhood environment-wide association study (NE-WAS)" of neighborhood influences on physical activity among older adults, analogous, in a genetic context, to a genome-wide association study.

**Methods:** Physical Activity Scale for the Elderly (PASE) and sociodemographic data were collected via telephone survey of 3,497 residents of New York City aged 65 to 75 years. Using Geographic Information Systems, we created 337 variables describing each participant's residential neighborhood's built, social, and economic context. We used survey-weighted regression models adjusting for individual-level covariates to test for associations between each neighborhood variable and (i) total PASE score, (ii) gardening activity, (iii) walking, and (iv) housework (as a negative control). We also applied two "Big Data" analytic techniques, LASSO regression, and Random Forests, to algorithmically select neighborhood variables predictive of these four physical activity measures.

**Results:** Of all 337 measures, proportion of residents living in extreme poverty was most strongly associated with total physical activity [−0.85; (95% confidence interval, −1.14 to −0.56) PASE units per 1% increase in proportion of residents living with household incomes less than half the federal poverty line]. Only neighborhood socioeconomic status and disorder measures were associated with total activity and gardening, whereas a broader range of measures was associated with walking. As expected, no neighborhood measures were associated with housework after accounting for multiple comparisons.

**Conclusions:** This systematic approach revealed patterns in the domains of neighborhood measures associated with physical activity.

**Impact:** The NE-WAS approach appears to be a promising exploratory technique. *Cancer Epidemiol Biomarkers Prev*; 26(4): 495–504. ©2017 AACR.

See all the articles in this *CEBP Focus* section, "Geospatial Approaches to Cancer Control and Population Sciences."

### Introduction

Physical activity, bodily movement produced by skeletal muscles, prevents colon and breast cancer, even after accounting for differences in body size (1–3). Physical activity also prevents obesity, which itself causes cancer at 13 organ sites including breast, colon, kidney, pancreas, and esophagus (4). There is also evidence that physical activity can reduce fatigue, improve quality of life, and improve survival among cancer patients (5–8). Among cancer survivors, particularly colorectal and breast cancer survivors, obesity increased more rapidly from 1997 to 2014 than in

the general population (9). Roughly 20% of older adults (individuals above the age of 65) living in the United States are cancer survivors (10); promoting physical activity among older adults may thus substantially improve cancer outcomes in the population. The Centers for Disease Control and Prevention recommends that older adults engage in at least five hours a week of moderate aerobic activity (or equivalent amounts of vigorous activity), in part to prevent cancer (11). Yet in practice, few older adults meet physical activity recommendations (12, 13).

One thread of physical activity promotion, including among older adults, has sought to identify and remove neighborhood contextual barriers to physical activity as part of everyday life (14–16). For example, safe and attractive sidewalks may make older adults choose to walk rather than drive for certain short trips, thereby increasing their activity levels (17, 18). However, despite considerable qualitative evidence supporting the concept that supportive neighborhood environments encourage physical activity in older adults (19, 20) and the development of numerous theoretical frameworks (14, 21–23) exploring the conceptual links between neighborhood environments and physical activity, quantitative evidence confirming that specific neighborhood features support specific activities has been inconsistent (24).

One reason for inconsistency may be the difficulty of objectively operationalizing the neighborhood constructs described qualitatively. For example, in interviews, older adults frequently indicate that they do not like to walk in their neighborhoods if

<sup>1</sup>Harborview Injury Prevention & Research Center, University of Washington, Seattle, Washington. <sup>2</sup>Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota. <sup>3</sup>Department of Emergency Medicine, University of California, Davis, Davis, California. <sup>4</sup>Albert Einstein College of Medicine, New York, New York. <sup>5</sup>Department of Ageing and Life Course, World Health Organization, Geneva, Switzerland. <sup>6</sup>Department of Epidemiology, Mailman School of Public Health, New York, New York.

**Note:** Supplementary data for this article are available at *Cancer Epidemiology, Biomarkers & Prevention* Online (<http://cebp.aacrjournals.org/>).

**Corresponding Author:** Stephen J. Mooney, Harborview Injury Prevention & Research Center, 401 Broadway, 4th Floor, Seattle, WA 98122. Phone: 206-799-3977; Fax: 206-744-9962; E-mail: [sjm2186@u.washington.edu](mailto:sjm2186@u.washington.edu)

doi: 10.1158/1055-9965.EPI-16-0827

©2017 American Association for Cancer Research.

they feel that they may be targeted for crime (19). However, measuring neighborhood crime risk is extremely difficult (25). Whereas one study may operationalize neighborhood crime as reported neighborhood crime in an administrative area such as a county or zip code (26), another may ask subjects to report their perceptions of neighborhood safety (27), and a third might use geostatistical techniques such as ordinary kriging to estimate prevalence of crime within a buffer around subject homes (28). If the third measure most accurately reflected the true impediment to activity, both the zip code-based and self-report-based studies would likely underestimate true effects due to nondifferential measurement error (29).

A complementary reason for quantitative inconsistency is the sheer number of ways researchers characterize neighborhoods using "Big Data," geographic information systems (GIS) tools and spatial analysis techniques (30–32). There is no perfect definition of neighborhood; the physical space connoted by "neighborhood" is subjective and may vary as older adults lose or recover functional capacity (33–35). As a result, researchers define neighborhood in many different ways, including radial buffers (the area within a given distance "as the crow flies" from the subject's home address), network buffers (the area reachable via walking a given distance walk on city streets), and activity spaces [the area a subject was observed to travel through over some period during which the subject wore a Global Positioning System (GPS) device; refs. 34, 36]. These differences in neighborhood specification can strongly affect study validity (37, 38).

Molecular epidemiologists have developed analytic approaches to explore similar groups of measures systematically (39). In a genome-wide association study (GWAS), "agnostic" analytic approaches are used to search the whole genome for the strongest genetic associations, which are then assumed to be the best candidates for subsequent research (40). Recently, the agnostic paradigm has been extended to high-throughput "-omic" fields focused on biomarker discovery and environmental sciences, where it has been termed an "environment-wide association study" (EWAS; refs. 40–43). Although agnostic approaches have limited causal interpretation, their systematic nature also enables straightforward replication (39, 44). As neighborhood datasets increasingly resemble GWAS and EWAS datasets, it follows that neighborhood research may similarly benefit by drawing on agnostic research paradigms.

Taking explicit inspiration from the GWAS and EWAS approaches, we propose and illustrate the neighborhood environment-wide association study (NE-WAS) design. We analyze each neighborhood measure's ability to predict total physical activity and each of three specific subtypes of activity, controlling for individual characteristics. Next, we examine the robustness of identified associations to variation in neighborhood definition. Finally, in analyses, we consider two machine learning algorithms that select variables algorithmically (LASSO and Random Forest) and have been used in epigenetic analyses, as potential components of a NE-WAS.

## Materials and Methods

### Subjects and setting

We use data from NYCNames-II, a study of 3,497 residents of New York City aged 65–75 years. Sampling and recruitment for NYCNames-II has been described previously (45). A total of 279 subjects (8%) reported poor health and were excluded, leaving

3,218 subjects for the primary analysis. Briefly, subjects were recruited by phone in 2011 from a telephone list purchased from InfoUSA. Phone numbers had previously been geocoded to census tracts, and numbers were selected to ensure geographic coverage of the city. Final survey weights were then raked [i.e. sample weights were recomputed so the weighted sample approximates the joint distribution of known population characteristics (46)] to New York City population estimates from the 2006–2010 American Community Survey for gender and race/ethnicity and from 2010 Census estimates for educational attainment and borough of residence. For this analysis, we used only data from the baseline interview. Descriptive statistics for the cohort are reported in Table 1

### Measures

During the baseline interview, each subject reported his or her age, educational attainment, health status, income, race/ethnicity, and sex. Because we theorized that the neighborhood environment should only be able to influence physical activity among subjects whose health permitted outdoor physical activity, we excluded those who reported poor health from the primary analysis.

We assessed past-week physical activity using sixteen items derived from the Physical Activity Scale for the Elderly (PASE; refs. 47–49), a validated survey tool designed for use with older adults. The PASE instrument assesses past-week physical activity in a number of domains, including strengthening exercises, sports and recreation, walking, gardening, and housework. The PASE score is a linear combination of all sixteen items that reflects total physical activity [ $r = 0.68$  with a doubly labeled water assessment of physical activity, considered the gold standard metabolic measure of energy expenditure (50)—in one validation study; refs. 47–49]. PASE scores in the included subjects ranged from 0 to 296 and were slightly right-skewed, with a mean of 84 and a median of 77. Thirty-nine percent of the subjects reported daily walking, 23% percent reported gardening, and 57% reported doing heavy housework.

**Neighborhood measures.** During the baseline interview, each subject reported their home address. We geocoded these addresses using GeoSupport, a New York City-specific geocoding tool released by New York's Department of City Planning. Ninety-six percent of addresses were successfully geocoded to a rooftop; the remaining 4% were assigned the age 64–73 population weighted centroid of the reported ZIP code as a home location. For each subject, we defined the residential neighborhood as the land area reachable by city streets within a given distance of the geocoded home location, an area referred to in neighborhood research as a network buffer (51–53). Our primary analysis used 0.25 km network buffers, which contain the area accessible within a 5-minute walk for a 70-year-old woman with a comfortable gait speed within two SDs of the mean (54).

For each subject, we compiled 337 unique neighborhood measures. Specifically, demographic and economic characteristics came from the 2006–2010 American Community Survey. Urban form measures were constructed from TIGER/Line shapefiles describing street layout, the New York Metropolitan Transit Authority's ridership reports, and a LiDAR scan of the city (55, 56). Crime and disorder measures were compiled from a measure of crime risk developed by ESRI, Inc., municipal street cleanliness records, a systematic virtual audit using Google Street

**Table 1.** Selected characteristics of the subjects included in this analysis

Characteristic	Full cohort (N = 3,497)		Fair or better health (N = 3,218)	
	N (%)	Sample weighted %	N (%)	Sample weighted %
Age				
65–68	1,045 (33)	34	956 (33)	33
69–71	664 (21)	20	608 (21)	21
72–75	1,442 (46)	46	1,335 (46)	47
Sex				
Female	2,094 (60)	58	1,907 (59)	57
Male	1,403 (40)	42	1,311 (41)	43
Race/Ethnicity				
Non-Hispanic white	1,800 (51)	47	1,701 (53)	48
Non-Hispanic black	1,073 (31)	26	974 (30)	26
Hispanic	245 (7)	14	209 (6)	13
Other	379 (11)	12	334 (10)	12
Educational attainment				
Less than high school	673 (19)	32	570 (18)	30
Completed high school	949 (27)	29	870 (27)	29
Some college	627 (18)	15	570 (18)	15
Completed college	1,248 (36)	24	1,208 (38)	25
Household income				
Less than \$20,000	1,279 (37)	40	1,097 (34)	37
\$20,000–40,000	842 (24)	24	790 (25)	24
\$40,000–80,000	745 (21)	21	711 (22)	22
More than \$80,000	631 (18)	16	620 (19)	17
Health status				
Excellent	645 (18)	17	645 (20)	18
Good	1,523 (44)	42	1,523 (47)	46
Fair	1,050 (30)	33	1,050 (33)	36
Poor	279 (8)	9	– (–)	–
Activity measures				
Walked 5–7 days in the last week	1,346 (38)	42	1,154 (36)	39
Gardened in the last week	710 (20)	23	686 (21)	24
Performed heavy housework in the last week	1,872 (54)	54	1,793 (56)	57

View imagery, and homicide incident locations as reported by the New York City Police Department to the New York Times (57–60). Parks measures, including boundaries and park cleanliness, were obtained from The New York City Department of Parks and Recreation (61). Pedestrian and cyclist injury counts were compiled from records initially recorded by reporting police officers (62).

We categorized measures into bins according to the aspect of the urban environment each captured (Table 2). These bins are analogous to chromosomes in genomic studies, genes in epigenomic studies, or "class groupings" in an environment-wide association study (42).

Some measures, such as density of vehicle collisions involving pedestrians, were right-skewed. To be consistent with best practices in agnostic studies and maximize comparability between environmental predictors, we transformed such skewed predictors before analysis (42). To assess skew for each measure, we visually compared a histogram of the measure and the measure after log-transformation. We conducted a preliminary investigation of an automated procedure to decide whether to log-transform measures, detailed in Supplementary Fig. S1. We retained log-transformed measures for analysis in place of untransformed measures if the log-transformed measure visually appeared closer to a normal distribution than its untransformed analogue.

**Table 2.** Summary of measures used in the NE-WAS

Domain	Number of measures	Data source(s)	Examples
Demographics and housing characteristics	121	American Community Survey	Population density, % white alone, % boys ages 10–14
Education, employment, and income	102	American Community Survey	% College grad, % in labor force, % in food prep sector
Urban form and walkability	50	American Community Survey, New York State Accident Location Information Service Line Layer, NYC Transit Authority	% walk to work, density of 4-way intersections, Bus stop density, % of roadbed covered by tree canopy
Crime and disorder	35	Esri Crime Risk, Google Street View, New York Times Homicide Map, NYC Sanitation Department Report Cards	Weighted average risk of larceny, Mean neighborhood disorder, % filthy streets
Parks	5	New York City Department of Parks and Recreation	% of land area in large parks
Pedestrian safety	24	New York State Department of Transportation and New York City Police Department	Cyclist injury density in the 1990s, Pedestrian fatality density in the 2000s

For every pair of perfectly correlated measures (for example, proportion of occupied homes occupied by owners and proportion of occupied homes occupied by renters have a correlation coefficient of  $-1$ ), we excluded one of the measures.

Supplementary Table S1 includes a complete list of all 337 contextual measures used in the final analysis, including their underlying data sources, and whether or not we log-transformed the measure before analysis.

### Statistical analysis

Following the GWAS analytic approach, we used linear regression to model PASE score from each neighborhood environment variable individually. In addition, we used logistic regression analogously to estimate the strength of association between each variable and engaging in each of three activities: (i) daily walking, (ii) gardening, and (iii) "heavy housework" (e.g. vacuuming, sweeping, moving furniture; ref. 49). We hypothesized on the basis of prior literature that daily walking would be associated with measures of urban form (63, 64) and that, because lack of outdoor space poses a barrier to gardening in many but not all neighborhoods of New York City, gardening would be associated with housing characteristics (65). We selected heavy housework as a "negative control" (66). That is, because we believe that neighborhood conditions do not affect participation in heavy housework, we can interpret a finding that a large number of neighborhood exposures are associated with housework or that a pattern of exposures similar to the pattern predictive of other activity measures are associated with housework as evidence of residual confounding. Whereas gardening and heavy housework are dichotomous measures in the PASE instrument, daily walking is not; we considered those who reported 5–7 days of past-week walking to be daily walkers. All regression analyses incorporated survey weights and controlled for individual's age, race/ethnicity, educational attainment, income, and home size. After Bonferroni correction for 337 comparisons, we had adequate sample size to detect a change of 3.5 PASE units (roughly 10 minutes of walking per day) for each SD change in neighborhood exposure with a probability of 0.73 (67).

Next, to explore how buffer size affects the pattern of types of measures correlated with activity, we repeated all regression analyses with each measure computed for a 1-km network buffer around the subjects' home address. We then compared the estimated regression parameters for measures at 1 km to regression parameters for measures at 0.25 km to identify instances where neighborhood measures were more predictive at one neighborhood scale as opposed to the other.

Finally, to explore how developments in computer science might inform future NE-WASs, we investigated two algorithmic approaches (LASSO regression and Random Forest) that select the neighborhood characteristics most predictive of physical activity in multivariable models (68, 69). The variables remaining in models tuned using cross-validation can be regarded as the most informative variables (70). We selected these algorithms because they are widely used, software is readily available, and analogous approaches are increasingly common in GWAS studies searching for gene–gene interactions (71) and in epigenetic studies (41). These explorations are detailed in the online supplement.

### Missing data

Relatively little data was missing on physical activity (maximum of 1.8% on any PASE item) or demographic covariates

(16.2% were missing household income data; no other items were missing for more than 10% of subjects), and no data were missing on a neighborhood covariates. Nonetheless, to address potential non-response biases, we performed all survey-weighted regressions on each of 5 datasets where missing values were imputed using multivariate sequential regression as implemented by IVEWARE (72) with all available survey responses included in the prediction model. Following standard practice, we combined the estimates resulting from these models using Rubin's rules (73).

### Sensitivity analyses

To test our regression results' sensitivity to the assumption that neighborhood characteristics were not important for those who reported poor health, we repeated the primary analysis with the full cohort of 3,497 subjects.

### Software

All analyses used 64-bit R for Windows version 3.2.3.

## Results

### Characteristics associated with physical activity

In linear regression models controlling for individual covariates, measures of neighborhood resident socioeconomic position and physical disorder were most strongly associated with total physical activity. Specifically, the proportion of residents living in households with incomes less than half the poverty level was the most strongly associated with PASE score, with an estimated decrease of 0.85 (95% CI: 0.56–1.14) PASE score units per 1% increase in proportion of residents living in households with incomes less than half the federal poverty line. This association size is equivalent to 10 minutes less of daily walking per 4% decrease in proportion of residents living below half the federal poverty line. The remaining four of the top five measures included three other measures of resident socioeconomic position, all showing correlations between higher numbers of higher-income residents and more physical activity among the NYC NAMES II subjects, and one disorder measure, showing well-maintained windows, a marker of building upkeep, to be correlated with more activity (Table 3). After Bonferroni correction, no measure of resident demographics, parks, urban form, or pedestrian and cyclist safety were associated with PASE score.

Logistic regression analyses focused only a single type of activity identified many more significant neighborhood correlates than analyses targeting total activity (Fig. 1, given in color in the online supplement). Measures of high neighborhood socioeconomic status were the strongest predictors of gardening, whereas a wide range of neighborhood characteristics predicted walking 5–7 days in the previous week (Table 3). Reassuringly, no neighborhood measures were predictive of heavy housework after Bonferroni correction.

### Neighborhood characteristics and buffer size

Using 1 km rather than 0.25-km buffers led to more variables being significant after Bonferroni correction, but neither neighborhood measures for the 1-km buffers nor for the 0.25-km buffers were uniformly more strongly correlated with total PASE score (Fig. 2, given in color in the online Supplementary Data). Of the 337 neighborhood measures available at both scales, regression coefficients for associations with PASE scores changed signs

**Table 3.** Specific neighborhood measures identified as most predictive for several physical activity outcomes using regression models

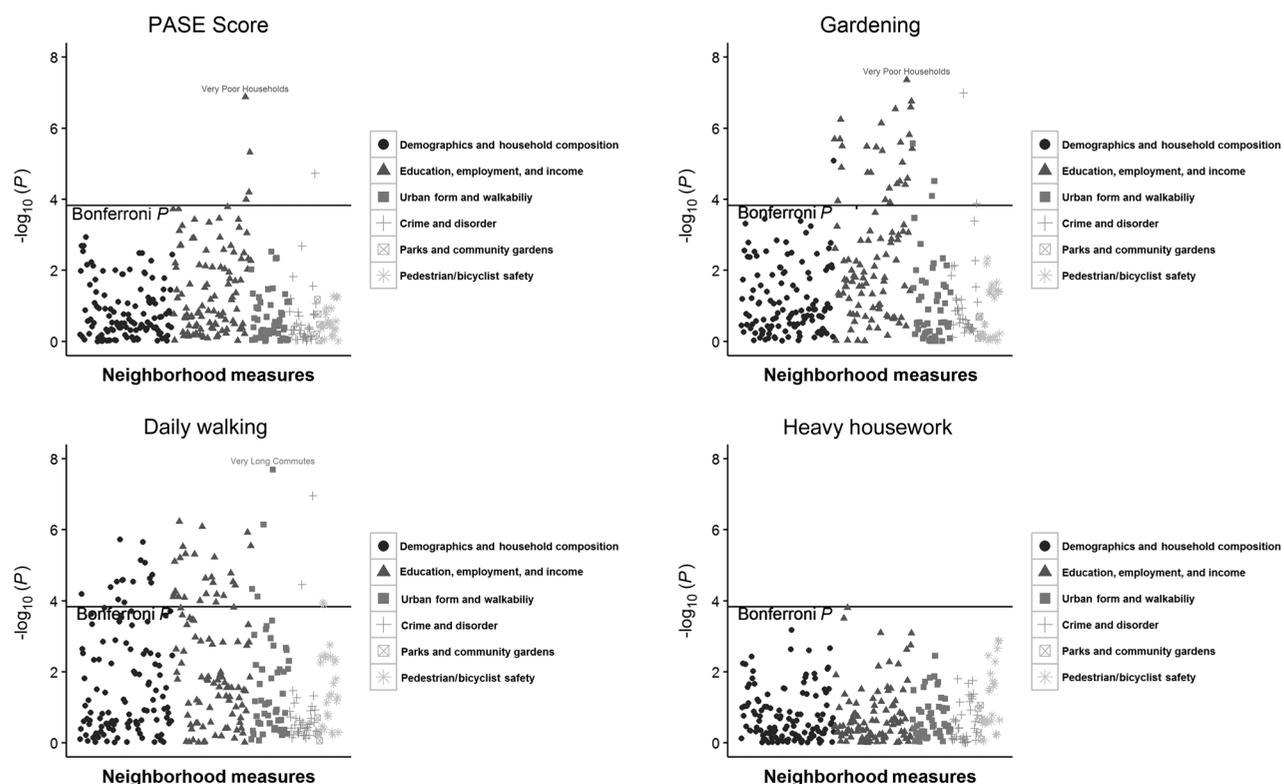
	PASE Score	Gardening	Walking daily	Heavy housework
Count of measures that remained significant after Bonferroni correction	5 (1.5%)	33 (9.8%)	49 (14.4%)	0 (0.0%)
Top 5 statistically significant neighborhood measures (by <i>P</i> value of coefficient)	People living in households with incomes less than half the poverty level (–) People living in households with incomes below the poverty line (–) No problems with windows in HVS survey (+) People living in households with incomes more than twice the poverty level (+) People living in households with incomes between half and three-quarters of the poverty level (–)	People living in households with incomes less than half the poverty level (–) Neighborhood Physical Disorder (–) People living in households with incomes below the poverty line (–) People living in households above twice the poverty line (+) People living in households with any interest, dividend, or rental income (+)	Proportion of residents with 60- to 90-minute travel time to work (–) Broken windows in HVS survey (–) Proportion of adult population with at least some college education (+) Proportion of working adult population commuting by car, truck, or van (–) Proportion of adult population working in professional or management industries (+)	–

NOTE: All analyses control for subject age, race/ethnicity, educational attainment, household income, gender, and home type. Abbreviation: HVS, New York City Housing and Vacancy Survey

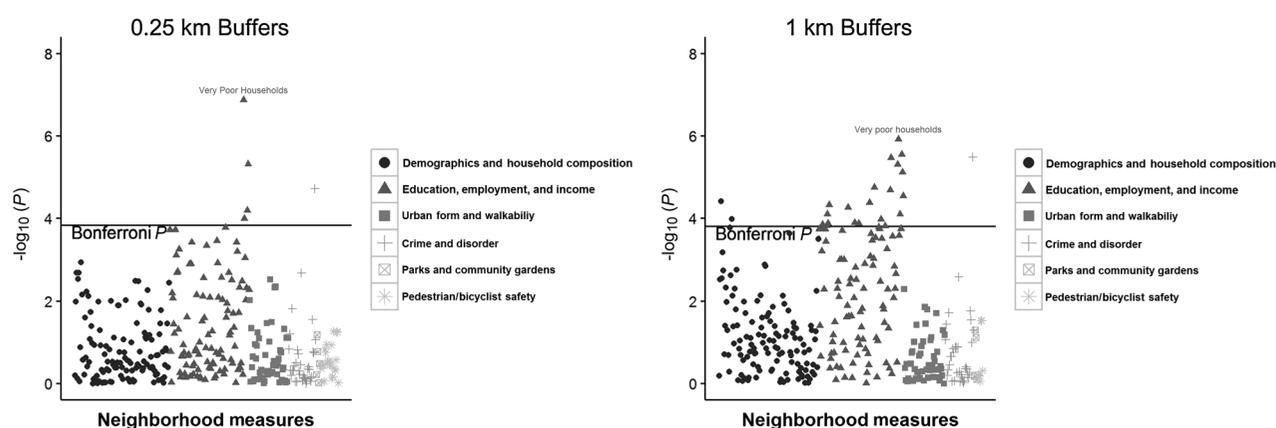
for 38 (11%) of the neighborhood measures. None of the measures that changed sign were nominally significant at a *P* value of 0.05 at either neighborhood buffer scale. There was no clear pattern as to which neighborhood measures were better correlated with PASE scores at the two scales (Table 4).

#### Algorithmic variable selection

The best fitting LASSO regression models for each outcome incorporated roughly the same number of neighborhood variables as were significant in conventional regression for the respective outcome (3 for total PASE score, 45 for gardening, 22 for walking,

**Figure 1.**

Manhattan plots showing the strengths of associations between individual neighborhood variables and various physical activity outcomes as measured by the PASE after controlling for age, sex, race/ethnicity, educational attainment, income, and housing type.



**Figure 2.**

Manhattan plots showing the strength of associations between individual neighborhood variables and total PASE score at smaller and larger buffer sizes after controlling for age, sex, race/ethnicity, educational attainment, income, and housing type.

and 0 for heavy housework). However, the specific variables that were selected were highly sensitive to model tuning parameters, limiting substantive interpretability. The variables ranked as highly important in Random Forests frequently all belonged to the same neighborhood measurement domain (e.g., all variables important for gardening were related to housing characteristics) but did not match the variables selected by LASSO regression or conventional regression. The final LASSO model explained 10.1% of variation in PASE score; in contrast, the Random Forest explained  $-3.6\%$  of PASE variation, suggesting the final model was worse than chance. Results from algorithmic variable selection are discussed in more detail in the online Supplementary Data.

### Sensitivity analyses

Subjects who were excluded from the primary analysis owing to poor health were more likely to be female, to be racial/ethnic minorities, to have lower household incomes, and to be less educated (Table 1). However, the sensitivity analysis conducted using the full cohort identified the same top 5 measures, albeit in a different order (Supplementary Table S2).

## Discussion

In this analysis, we explored a novel agnostic NE-WAS approach to selecting the neighborhood measures most strongly

associated with total physical activity, as well as specifically with walking, gardening, and housework. In our study, the most strongly predictive measure of total physical activity was proportion of residents living in households with incomes below half the federal poverty threshold, equating to \$11,056 for a family of four. Neighborhood socioeconomic and disorder measures were most associated with total activity. Socioeconomic measures also strongly predicted gardening, whereas measures of commute distance and commute times were more relevant for walking. As expected, no neighborhood measures significantly predicted heavy housework. Overall, the NE-WAS approach appears promising, and our findings suggest NE-WAS may be appropriate for other neighborhood-associated health conditions as well, such as obesity (64), breast cancer (74), or cardiac arrest (75).

More neighborhood environment measures were significantly associated with our specific outdoor activity measures, walking and gardening, than with physical activity as a whole, while no neighborhood measures significantly predicted heavy housework. Our findings thus serve as empirical support for prior calls, typically made on theoretical grounds, to consider influences on differing domains of activity separately (24, 25, 76–78).

There are several interpretations for our finding that neighborhood socioeconomic measures were more consistently associated with activity measures than neighborhood characteristics (e.g. access to parks) that have more direct theoretical relevance to

**Table 4.** The top 5 measures for each neighborhood definition that were more significant predictors of total PASE score than using the alternate neighborhood definition

	Measure	Difference in $-\log_{10}P$ value
Better at 0.25-km scale	Proportion of population living in households (+)	2.24
	Proportion of population who are naturalized citizens (–)	1.68
	Proportion of population living in households with income below half the poverty level (–)	1.57
	Density of 3-way intersections (+)	1.28
	Proportion of vacant housing units offered for rent (–)	1.27
Better at 1.0-km scale	Proportion of households with incomes between 25K and 30K (–)	2.77
	Proportion of adult residents with a professional degree or more (+)	2.56
	Proportion of households with incomes between 30K and 35K (–)	2.56
	Proportion of family households living below the poverty line with a male householder and no children under age 18 (–)	2.45
	Proportion of total population aged 10 to 14 years	2.42

NOTE: Plus or minus indicates the direction of association between the neighborhood measure and PASE score.

specific forms of outdoor activity. It may be that residents of higher socioeconomic status neighborhoods have used their resources to shape neighborhoods to offer more support for different forms of activity among older adults (79), including dedicated outdoor space that supports gardening and well-maintained sidewalks or amenities such as benches and public restrooms that older adults cite as necessary supports for walking (80). A complementary explanation is that residual confounding due to incomplete control for individual socioeconomic position is responsible for this association. Higher socioeconomic position older adults are typically more physically active (81, 82), and tend to live in neighborhoods with other high socioeconomic position individuals. Our analysis controlled for household income and educational attainment, but neither fully captures socioeconomic position among older adults (83).

While the NE-WAS approach explicitly draws an analogy between genetics and neighborhoods, we caution, as others have, that there are vital differences between genomes and modifiable exposures like neighborhoods (44). Most importantly, unlike the SNPs that act as independent variables in a GWAS, wherein there are few if any correlations between polymorphisms on separate chromosomes, the correlation structure underlying neighborhood characteristics is strong, complex, and potentially causally circular (84). However, in proteomics and metabolomics research, wherein measured molecules do show strong and complex intercorrelations, identified molecules are considered to be markers of a process rather than causes of the process and a separate scientific approach, pathway analysis, has developed to integrate knowledge from agnostic analyses to develop and test causal hypotheses (85, 86). There are analogous systems science-derived integrations of knowledge in neighborhood research (e.g., ref. 87), although such approaches are still in their infancy. Nonetheless, we anticipate that in this sense the NE-WAS approach is more akin to an omics approach than a GWAS: the value of the NE-WAS stems not from a precise estimate of the causal effect of some neighborhood characteristic but rather from the ability to systematically identify targets for future exploration and to reveal reproducible patterns in associations across cohorts (39).

While this analysis addressed neighborhood factors correlated with physical activity, the NE-WAS approach could be applied to explore other contextual research questions. For example, NE-WAS might help to systematically explore the appropriate measures and geographic levels at which to understand the disparities in cancer incidence (35). NE-WAS may also be of value for standardizing neighborhood definitions or selecting the spatial scale at which a neighborhood construct is most relevant (24). Finally, NE-WAS with pooled or multisite studies would allow a systematic assessment of correlation patterns between geographic regions, quantifying variation in susceptibility to neighborhood environment risk factors (88).

In general, algorithmic variable selection resulted in substantively uninterpretable models. We caution, however, that our investigation was limited to two techniques that were designed for prediction rather than for explanation and happen to embed variable selection within the predictive modeling approach. Future NE-WASs might explore not only more aggressive exclusion of collinear measures but also other algorithmic approaches, including multifactor dimension reduction (89), a technique that explicitly aims to identify interactions that might not be uncovered by conventional analytic approaches.

Further analysis of correlation structures among neighborhood predictors, which were out of scope for this analysis, may shed further light on the most appropriate analytic techniques for future NE-WASs (90, 91).

This study had several notable strengths. First, the relatively large and population-based sample of older adults residing in a very well-characterized urban environment allowed for relatively precise estimates of associations between neighborhood characteristics and activity outcomes. Second, the use of a survey measure that included items assessing types of activity allowed us to incorporate analyses that target activity measures representing a range of hypothesized susceptibility to neighborhood influence (24, 77). Third, without a theoretical basis to guide variable selection, agnostic studies are at risk of identifying strongly confounded variables. In this light, our "negative control" finding that no environmental factors were associated with heavy housework after Bonferroni correction provides some, albeit incomplete, evidence against residual confounding (66).

However, our results should be viewed in light of several limitations. First, the 337 neighborhood measures we analyzed comprise only the measures of New York City's urban environment that were readily available to the research team. Future NE-WASs might productively undertake a more systematic exploration of neighborhood measures used in the literature to select a comprehensive set of measures to study, potentially incorporating neighborhood measures of no theoretical relevance as further negative controls. Second, we compared only two neighborhood definitions, 0.25-km network buffers and 1.0-km network buffers. It has been repeatedly noted that no single definition captures the construct of a neighborhood (24, 36); indeed, the meaning of neighborhood may be different for different measures, in different contexts, and for different subgroups (76). Future NE-WASs might broadly compare more buffer sizes for a single measure. Third, while New York City comprises a range of urban environments, including pockets of sidewalk-free post-war "sprawl," it nonetheless contains a much more pedestrian-oriented environment than the United States as a whole, and a population at greater extremes of the socioeconomic spectrum. It may be productive to compare results from this NE-WAS to future NE-WASs conducted in environments more representative of the contexts in which most American older adults reside. Fourth, as in any agnostic study, our substantive findings should be viewed with caution until replicated in other cohorts (92). Fifth, the Bonferroni correction we used to account for multiple comparisons is likely overly conservative; future NE-WASs might explore estimating the false discovery rate instead (93). Finally, as in most neighborhood studies, we were unable to determine whether statistical adjustment for participant race/ethnicity and socioeconomic status fully account for residential self-selection (94).

In conclusion, the NE-WAS is a promising approach to empirically identify neighborhood measures most strongly related to measurable outcomes, including not only cancer-preventing behaviors such as physical activity, but also health outcomes such as cancer incidence. In this NE-WAS, neighborhood socioeconomic characteristics were more consistently associated with physical activity than measures of crime, parks, and pedestrian safety. We anticipate performing NE-WASs in other cohorts, other geographic contexts, and with other outcomes, to determine the replicability of the approach, to improve handling of multicollinearity, and to deepen substantive findings (39).

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Authors' Contributions**

**Conception and design:** S.J. Mooney, J.R. Beard, A.G. Rundle  
**Development of methodology:** S.J. Mooney, G.J. Kennedy, J.R. Beard, A.G. Rundle  
**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** M. Cerdá, G.J. Kennedy, J.R. Beard, A.G. Rundle  
**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** S.J. Mooney, J.R. Beard, A.G. Rundle  
**Writing, review, and/or revision of the manuscript:** S.J. Mooney, S. Joshi, M. Cerdá, G.J. Kennedy, J.R. Beard, A.G. Rundle  
**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** S. Joshi, A.G. Rundle  
**Study supervision:** M. Cerdá, A.G. Rundle

**Acknowledgments**

The authors thank Thelma Mielenz, Alfred Neugut, Shuang Wang, and Ryan Demmer for their helpful comments on an earlier version of this work.

**Grant Support**

S.J. Mooney was supported by National Institute of Child Health and Human Development (NICHD) grant 5T32HD057822-07. S. Joshi, M. Cerdá, J.R. Beard, G.J. Kennedy, and A.G. Rundle were supported by National Institute for Mental Health grant 5R01MH085132-05.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 18, 2016; revised January 9, 2017; accepted January 27, 2017; published OnlineFirst February 2, 2017.

**References**

- Eheman C, Henley SJ, Ballard-Barbash R, Jacobs EJ, Schymura MJ, Noone AM, et al. Annual Report to the Nation on the status of cancer, 1975-2008, featuring cancers associated with excess weight and lack of sufficient physical activity. *Cancer* 2012;118:2338-66.
- Rogers CJ, Colbert LH, Greiner JW, Perkins SN, Hursting SD. Physical activity and cancer prevention. *Sports Med* 2008;38:271-96.
- Rundle A. Mechanisms underlying the effects of physical activity on cancer. Physical activity, dietary calorie restriction, and cancer. New York: Springer; 2011. p. 143-63.
- Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body fatness and cancer—viewpoint of the IARC working group. *N Engl J Med* 2016;375:794-8.
- Cramp F, Daniel J. Exercise for the management of cancer-related fatigue in adults. *Cochrane Database Syst Rev* 2012;11:CD006145.
- Ibrahim EM, Al-Homaidh A. Physical activity and survival after breast cancer diagnosis: meta-analysis of published studies. *Med Oncol* 2011;28:753-65.
- Holmes MD, Chen WY, Feskanich D, Kroenke CH, Colditz GA. Physical activity and survival after breast cancer diagnosis. *JAMA* 2005;293:2479-86.
- Schmitz KH, Holtzman J, Courneya KS, Masse LC, Duval S, Kane R. Controlled physical activity trials in cancer survivors: a systematic review and meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2005;14:1588-95.
- Greenlee H, Shi Z, Molmenti CLS, Rundle A, Tsai WY. Trends in obesity prevalence in adults with a history of cancer: results from the US National Health Interview Survey, 1997 to 2014. *J Clin Oncol* 2016;34:3133-3140.
- Parry C, Kent EE, Mariotto AB, Alfano CM, Rowland JH. Cancer survivors: a booming population. *Cancer Epidemiol Biomarkers Prev* 2011;20:1996-2005.
- Centers for Disease Control and Prevention. How much physical activity do older adults need? 2015. Available from: [http://www.cdc.gov/physicalactivity/basics/older\\_adults/index.htm](http://www.cdc.gov/physicalactivity/basics/older_adults/index.htm)
- Carlson SA, Fulton JE, Schoenborn CA, Loustalot F. Trend and prevalence estimates based on the 2008 physical activity guidelines for Americans. *Am J Prev Med* 2010;39:305-13.
- Tucker JM, Welk GJ, Beyler NK. Physical activity in US adults: compliance with the physical activity guidelines for Americans. *Am J Prev Med* 2011;40:454-61.
- Sallis J, Bauman A, Pratt M. Environmental and policy interventions to promote physical activity. *Am J Prev Med* 1998;15:379-97.
- Office of Disease Prevention and Health Promotion. Healthy people 2020; 2010. Washington, DC: U.S. Department of Health and Human Services. Available from: <http://www.healthypeople.gov/2010/>.
- American Cancer Society. Healthy Eating and Active Living 2016; Available from: <https://www.aacscn.org/what-we-do/healthy-eating-and-active-living>
- Pucher J, Dijkstra L. Promoting safe walking and cycling to improve public health: lessons from the Netherlands and Germany. *Am J Public Health* 2003;93:1509-16.
- Boarnet MG, Day K, Anderson C, McMillan T, Alfonso M. California's Safe Routes to School program: impacts on walking, bicycling, and pedestrian safety. *J Am Plan Assoc* 2005;71:301-17.
- Van Cauwenberg J, Van Holle V, Simons D, Deridder R, Clarys P, Goubert L, et al. Environmental factors influencing older adults' walking for transportation: a study using walk-along interviews. *Int J Behav Nutr Phys Act* 2012;9:85.
- Moran M, Van Cauwenberg J, Hercky-Linnewiel R, Cerin E, Deforche B, Plaut P. Understanding the relationships between the physical environment and physical activity in older adults: a systematic review of qualitative studies. *Int J Behav Nutr Phys Act* 2014;11:79.
- Stokols D. Translating social ecological theory into guidelines for community health promotion. *Am J Health Promot* 1996;10:282-98.
- Brownson RC, Boehmer TK, Luke DA. Declining rates of physical activity in the United States: what are the contributors? *Annu Rev Public Health* 2005;26:421-43.
- Lorenc T, Clayton S, Neary D, Whitehead M, Petticrew M, Thomson H, et al. Crime, fear of crime, environment, and mental health and wellbeing: mapping review of theories and causal pathways. *Health Place* 2012;18:757-65.
- Ding D, Gebel K. Built environment, physical activity, and obesity: what have we learned from reviewing the literature? *Health Place* 2012;18:100-5.
- Foster S, Giles-Corti B. The built environment, neighborhood crime and constrained physical activity: an exploration of inconsistent findings. *Prevent Med* 2008;47:241-51.
- Doyle S, Kelly-Schwartz A, Schlossberg M, Stockard J. Active community environments and health: the relationship of walkable and safe communities to individual health. *J Am Plan Assoc* 2006;72:19-31.
- Wilcox S, Castro C, King AC, Housemann R, Brownson RC. Determinants of leisure time physical activity in rural compared with urban older and ethnically diverse women in the United States. *J Epidemiol Community Health* 2000;54:667-72.
- Bader MD, Ailshire JA. Creating measures of theoretically relevant neighborhood attributes at multiple spatial scales. *Sociol Methodol* 2014;44:322-68.
- Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol* 2005;34:680-7.
- Auchincloss AH, Gebreab SY, Mair C, Roux AVD. A review of spatial methods in epidemiology, 2000-2010. *Annu Rev Public Health* 2012;33:107.
- Mooney SJ, Westreich DJ, El-Sayed AM. Commentary: epidemiology in the era of big data. *Epidemiology* 2015;26:390-4.
- Rundle A, Rauh VA, Quinn J, Lovasi G, Trasande L, Susser E, et al. Use of community-level data in the National Children's Study to establish the representativeness of segment selection in the Queens Vanguard Site. *Int J Health Geographics* 2012;11:1.
- Basta LA, Richmond TS, Wiebe DJ. Neighborhoods, daily activities, and measuring health risks experienced in urban environments. *Soc Sci Med* 2010;71:1943-50.
- Hirsch JA, Winters M, Clarke P, McKay H. Generating GPS activity spaces that shed light upon the mobility habits of older adults: a descriptive analysis. *Int J Health Geographics* 2014;13:1.

35. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian S, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project. *Am J Epidemiol* 2002;156:471–82.
36. Rundle AG, Sheehan DM, Quinn JW, Bartley K, Eisenhower D, Bader MM, et al. Using GPS data to study neighborhood walkability and physical activity. *Am J Prev Med* 2015;50:e65–72.
37. Weden MM, Bird CE, Escarce JJ, Lurie N. Neighborhood archetypes for population health research: is there no place like home? *Health Place* 2011;17:289–99.
38. Fotheringham AS, Wong DWS. The modifiable areal unit problem in multivariate statistical-analysis. *Environ Plann A* 1991;23:1025–44.
39. Ioannidis J. Exposurewide epidemiology: revisiting Bradford Hill. *Stat Med* 2015;35:1749–62.
40. Rundle A, Ahsan H, Vineis P. Better cancer biomarker discovery through better study design. *Eur J Clin Invest* 2012;42:1350–9.
41. Sun H, Wang S. Penalized logistic regression for high-dimensional DNA methylation data with case-control studies. *Bioinformatics* 2012;28:1368–75.
42. Patel CJ, Bhattacharya J, Butte AJ. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS ONE* 2010;5:e10746.
43. Lind PM, Risérus U, Salihovic S, Bavel Bv, Lind L. An environmental wide association study (EWAS) approach to the metabolic syndrome. *Environ Int* 2013;55:1–8.
44. Ioannidis JP, Loy EY, Poulton R, Chia KS. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. *Sci Translat Med* 2009;1:7ps8.
45. Mooney SJ, Joshi S, Cerdá M, Quinn JW, Beard JR, Kennedy GJ, et al. Patterns of physical activity among older adults in new york city: a latent class approach. *Am J Prev Med* 2015;49:e13–e22.
46. Lumley T. Analysis of complex survey samples. *J Stat Software* 2004;9:1–19.
47. Schuit AJ, Schouten EG, Westerterp KR, Saris WH. Validity of the physical activity scale for the elderly (PASE): according to energy expenditure assessed by the doubly labeled water method. *J Clin Epidemiol* 1997;50:541–6.
48. Washburn RA, McAuley E, Katula J, Mihalko SL, Boileau RA. The physical activity scale for the elderly (PASE): evidence for validity. *J Clin Epidemiol* 1999;52:643–51.
49. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. *J Clin Epidemiol* 1993;46:153–62.
50. Shephard RJ, Aoyagi Y. Measurement of human energy expenditure, with particular reference to field studies: an historical perspective. *Eur J Appl Physiol* 2012;112:2785–815.
51. Lovasi GS, Bader MD, Quinn J, Neckerman K, Weiss C, Rundle A. Body mass index, safety hazards, and neighborhood attractiveness. *Am J Prev Med* 2012;43:378–84.
52. Lovasi GS, Jacobson JS, Quinn JW, Neckerman KM, Ashby-Thompson MN, Rundle A. Is the environment near home and school associated with physical activity and adiposity of urban preschool children? *J Urban Health* 2011;88:1143–57.
53. Rainham D, McDowell I, Krewski D, Sawada M. Conceptualizing the healthscape: contributions of time geography, location technologies and spatial ecology to place and health research. *Soc Sci Med* 2010;70:668–76.
54. Bohannon RW. Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. *Age Ageing* 1997;26:15–9.
55. Purciel M, Neckerman KM, Lovasi GS, Quinn JW, Weiss C, Bader MD, et al. Creating and validating GIS measures of urban design for health research. *J Environ Psychol* 2009;29:457–66.
56. Lovasi GS, O'Neil-Dunne JP, Lu JW, Sheehan D, Perzanowski MS, MacFaden SW, et al. Urban tree canopy and asthma, wheeze, rhinitis, and allergic sensitization to tree pollen in a New York City Birth Cohort. *Environ Health Perspect* 2013;121:494.
57. Lovasi GS, Schwartz-Soicher O, Neckerman KM, Konty K, Kerker B, Quinn J, et al. Aesthetic amenities and safety hazards associated with walking and bicycling for transportation in New York City. *Ann Behav Med* 2013;45:76–85.
58. Lovasi GS, Schwartz-Soicher O, Quinn JW, Berger DK, Neckerman KM, Jaslow R, et al. Neighborhood safety and green space as predictors of obesity among preschool children from low-income families in New York City. *Prev Med* 2013;57:189–93.
59. Mooney SJ, Bader MDM, Lovasi GS, Neckerman KM, Teitler JO, Rundle AG. Validity of an ecometric neighborhood physical disorder measure constructed by virtual street audit. *Am J Epidemiol* 2014;180:626–35.
60. The New York Times. Homicide map. Available from: <https://www.nytimes.com/interactive/projects/crime/homicides/map>
61. Stark JH, Neckerman K, Lovasi GS, Quinn J, Weiss CC, Bader MD, et al. The impact of neighborhood park access and quality on body mass index among adults in New York City. *Prev Med* 2014;64:63–8.
62. Mooney SJ, DiMaggio CJ, Lovasi GS, Neckerman KM, Bader MD, Teitler JO, et al. Use of Google Street View to assess environmental contributions to pedestrian injury. *Am J Public Health* 2016;106:462–9.
63. Saelens BE, Sallis JF, Frank LD. Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Ann Behav Med* 2003;25:80–91.
64. Rundle A, Diez Roux AV, Freeman LM, Miller D, Neckerman KM, Weiss CC. The urban built environment and obesity in New York City: a multilevel analysis. *Am J Health Promot* 2007;21:326–34.
65. Echenique MH, Hargreaves AJ, Mitchell G, Namdeo A. Growing cities sustainably: does urban form really matter? *J Am Plan Assoc* 2012;78:121–37.
66. Lipsitch M, Tchetgen ET, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010;21:383.
67. Dupont WD, Plummer WD Jr. Power and sample size calculations for studies involving linear regression. *Control Clin Trials* 1998;19:589–601.
68. Breiman L. Random forests. *Machine Learning* 2001;45:5–32.
69. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc Series B (Methodol)* 1996;58:267–88.
70. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Machine Learning Res* 2003;3:1157–82.
71. Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am J Hum Genet* 2010;86:6–22.
72. Raghunathan TE, Solenberger PW, Van Hoewyk J. IVEware: imputation and variance estimation software. Ann Arbor, MI: University of Michigan; 2002. Available from: <http://www.isr.umich.edu/src/smp/ive/>.
73. Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons; 2004.
74. Akinyemiju TF, Genkinger JM, Farhat M, Wilson A, Gary-Webb TL, Tehranifar P. Residential environment and breast cancer incidence and mortality: a systematic review and meta-analysis. *BMC Cancer* 2015;15:1.
75. Mooney SJ, Grady ST, Sotoodehnia N, Lemaitre RN, Wallace ER, Mohanty AF, et al. In the wrong place with the wrong SNP: the association between stressful neighborhoods and cardiac arrest within beta-2-adrenergic receptor variants. *Epidemiology* 2016;27:656–62.
76. Feng J, Glass TA, Curriero FC, Stewart WF, Schwartz BS. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place* 2010;16:175–90.
77. Lovasi GS, Grady S, Rundle A. Steps forward: review and recommendations for research on walkability, physical activity and cardiovascular health. *Public Health Rev* 2012;33:484.
78. Saelens BE, Handy SL. Built environment correlates of walking: a review. *Med Sci Sports Exercise* 2008;40:S550.
79. Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav* 1995;80–94.
80. Mahmood A, Chaudhury H, Michael YL, Campo M, Hay K, Sarte A. A photovoice documentation of the role of neighborhood physical and social environments in older adults' physical activity in two metropolitan areas in North America. *Soc Sci Med* 2012;74:1180–92.
81. Kamphuis CB, van Lenthe FJ, Giskes K, Huisman M, Brug J, Mackenbach JP. Socioeconomic differences in lack of recreational walking among older adults: the role of neighbourhood and individual factors. *Int J Behav Nutr Phys Activity* 2009;6:1.
82. Tucker-Seeley RD, Subramanian S, Li Y, Sorensen G. Neighborhood safety, socioeconomic status, and physical activity in older adults. *Am J Prev Med* 2009;37:207–13.
83. Robert S, House JS. SES differentials in health by age and alternative indicators of SES. *J Aging Health* 1996;8:359–88.

84. Oakes JM. The (mis) estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med* 2004;58:1929–52.
85. Wu X, Al Hasan M, Chen JY. Pathway and network analysis in proteomics. *J Theoret Biol* 2014;362:44–52.
86. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 2012;8:e1002375.
87. Yang Y, Roux AVD, Auchincloss AH, Rodriguez DA, Brown DG. A spatial agent-based model for the simulation of adults' daily walking within a city. *Am J Prev Med* 2011;40:353–61.
88. Jencks C, Mayer SE. The social consequences of growing up in a poor neighborhood. In: Lynn LE, McGeary MGH, editors. *Inner-city poverty in the United States*. Washington, DC: National Academy Press; 1990. p. 111–86.
89. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 2001;69:138–47.
90. Patel C, Ioannidis J. Studying the elusive environment in large scale. *JAMA* 2014;311:2173–4.
91. Patel CJ, Ioannidis JP. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *J Epidemiol Community Health* 2014;68:1096–100.
92. Ioannidis J. This I believe in genetics: discovery can be a nuisance, replication is science, implementation matters. *Front Genet* 2013;4:33.
93. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B (Methodol)* 1995;289–300.
94. Van Dyck D, Cardon G, Deforche B, Owen N, De Bourdeaudhuij I. Relationships between neighborhood walkability and adults' physical activity: how important is residential self-selection? *Health Place* 2011; 17:1011–4.

# Cancer Epidemiology, Biomarkers & Prevention

## Contextual Correlates of Physical Activity among Older Adults: A Neighborhood Environment-Wide Association Study (NE-WAS)

Stephen J. Mooney, Spruha Joshi, Magdalena Cerdá, et al.

*Cancer Epidemiol Biomarkers Prev* 2017;26:495-504. Published OnlineFirst February 2, 2017.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/1055-9965.EPI-16-0827](https://doi.org/10.1158/1055-9965.EPI-16-0827)

**Supplementary Material** Access the most recent supplemental material at:  
<http://cebp.aacrjournals.org/content/suppl/2017/02/02/1055-9965.EPI-16-0827.DC1>

**Cited articles** This article cites 84 articles, 5 of which you can access for free at:  
<http://cebp.aacrjournals.org/content/26/4/495.full#ref-list-1>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cebp.aacrjournals.org/content/26/4/495>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.