KMANPUB | Health Nexus Vol. 4 No. 4 (2026): 1-12

Article history:
Received 10 April 2026
Revised 01 June 2026
Accepted 05 June 2026
Initial Publication 31 July 2026
Final Publication 01 October 2026

Environment, Ecology, and Urban Plant Diversity: Quantifying the Ecological Inference Limitation in US Green Space and Mental Health Research Using BRFSS and Land Cover Data

Zahidul Islam[1*]

1. Department of Botany, Government Tolaram College, University of Dhaka, Bangladesh

*** Corresponding author email address:** mdzahidul-2018223165@soc.du.ac.bd

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Islam, Z. (2026). Environment, Ecology, and Urban Plant Diversity: Quantifying the Ecological Inference Limitation in US Green Space and Mental Health Research Using BRFSS and Land Cover Data. *Health Nexus, 4*(4), 1-12.
https://doi.org/10.61838/kman.hn.5578

A B S T R A C T

State-level vegetation composition and population mental health share a poorly understood methodological relationship: most prior studies link individual-level health outcomes to aggregate geographic exposures without correcting for the resulting hierarchical data structure. Objective: To test the association between state-level Composite Vegetation Cover Index values and self-reported mental health in a nationally representative US sample, while explicitly quantifying the ecological inference limitation through intraclass correlation and design effect analysis. Methods: We linked 297,195 adults from the 2019 BRFSS to state-level CVCI values derived from NLCD 2019. Negative binomial regression with state-clustered standard errors estimated adjusted associations. Component-wise analysis separately tested tree canopy, shrub/scrub, and herbaceous vegetation. Results: The intraclass correlation coefficient was 0.0029, yielding a design effect of 18.05 and an effective sample size of 16,465 — demonstrating that state-level exposure linkage reduces the apparent n = 297,195 to 16,465 equivalent independent observations. Under clustering-corrected standard errors, the composite CVCI was not significantly associated with mental health (IRR = 1.100, 95% CI: 0.935–1.295, p = 0.251). Component-wise directional patterns were theoretically coherent: herbaceous cover associated with fewer unhealthy days, shrub/scrub with more, though neither reached significance after clustering correction. Conclusions: State-level vegetation composition is not significantly associated with population mental health after proper correction for the hierarchical data structure. This study provides the first formal quantification of the ecological inference limitation in US green space and mental health research and establishes the component-wise analytical framework that county-level replication studies should adopt.

Keywords: *green space; mental health; BRFSS; vegetation composition; ecological inference*

1. Introduction

Mental health disorders are among the leading causes of disability-adjusted life years globally, and the built and natural environment is increasingly recognised as a structural determinant of population mental health that operates independently of individual health-seeking behaviour (1, 2). Within this field, two theoretical frameworks predict that vegetation composition — not merely quantity — shapes mental health outcomes. Stress Restoration Theory (3) holds that natural environments trigger psychophysiological recovery from stress, with the strength of effect dependent on ecological complexity and accessibility. Attention Restoration Theory (4) holds that restorative capacity varies with vegetation structural quality. Both predict that accessible herbaceous and canopy environments are more restorative than dense, inaccessible shrub vegetation (5, 6).

The US literature on this topic has a gap that the present study directly addresses. While Australian (7), British (8), and Canadian (9) studies have examined vegetation and mental health at neighbourhood or dissemination-area scales, no study has examined this association at the national US scale using BRFSS data with explicit attention to the methodological constraints of state-level geographic linkage. This gap matters for a specific reason: the BRFSS public-use file does not include county or sub-state geographic identifiers, restricting exposure linkage to the state level. State-level vegetation values mask enormous within-state heterogeneity and introduce a well-known ecological inference problem that has not previously been formally quantified in the US green space and mental health literature.

A structural feature of US geography compounds this problem. The densely forested Appalachian and rural Western states combine high vegetation cover with elevated poverty, social isolation, and limited healthcare access — a geographic confounding structure that shapes state-level associations independently of any causal vegetation-mental health pathway. Treating such associations as evidence of a vegetation effect without accounting for this confounding and without correcting for the hierarchical data structure risks overstating both the magnitude and significance of findings.

The present study makes three contributions. First, it provides the first formal quantification of the ecological inference limitation in US BRFSS-based green space and mental health research, through intraclass correlation (ICC) and design effect (DEFF) analysis. Second, it introduces a component-wise decomposition of state-level NLCD vegetation cover into three theoretically distinct classes — tree canopy, shrub/scrub, and herbaceous vegetation — that prior single-index studies cannot provide. Third, it establishes the methodological framework that future county-level replication studies should adopt to properly test the vegetation-mental health hypothesis at a spatial scale where causal inference is plausible (10, 11).

2. Methods and Materials

2.1. Study Design and Data Sources

This study employs a cross-sectional ecological-linkage design combining individual-level survey data with state-level land cover metrics. The BRFSS public-use file does not include county or sub-state geographic identifiers; the state is the finest geographic unit available for exposure linkage without restricted-use data access. This constraint is the primary design limitation and is formally quantified in Section 2.5. Data sources: 2019 BRFSS (CDC); NLCD 2019 (USGS Multi-Resolution Land Characteristics Consortium); NCHS Urban-Rural Classification Scheme. The 2019 BRFSS wave was selected as the most recent pre-pandemic year (12).

2.2. Outcome Variable

The outcome was self-reported mentally unhealthy days in the past 30 days (BRFSS item MENTHLTH). Responses of 88 were recoded to 0; responses of 77 or 99 were treated as missing. The MENTHLTH item is a validated population-level psychological distress indicator (13). The distribution was severely right-skewed with 66.1% zero responses, motivating negative binomial regression.

2.3. Exposure Variable: Composite Vegetation Cover Index

The Composite Vegetation Cover Index (CVCI) was constructed from NLCD 2019 land cover classifications at the state level. The CVCI is the summed proportional state

Health Nexus
E-ISSN: 2981-2569

land area occupied by three vegetation classes: deciduous, evergreen, and mixed forest (NLCD classes 41, 42, 43); shrub/scrub (class 52); and herbaceous grassland (class 71). The CVCI is a summed-proportion measure of combined terrestrial vegetation cover — not a diversity index in the ecological sense, which would capture heterogeneity or balance among classes. Component-wise analysis tests whether the three NLCD classes have distinct mental health associations, directly addressing the composition question that the composite CVCI cannot resolve.

2.4. *Covariates and Missing Data*

Covariates selected a priori: age (standardised); sex; race/ethnicity (Non-Hispanic White [reference], Black, Hispanic, Asian, Other); educational attainment (four levels); employment status (four levels); household income (three levels); self-rated general health; physically unhealthy days; and census region fixed effects. Survey weights (_LLCPWT), stratification (_STSTR), and PSUs (_PSU) were incorporated throughout.

Of 418,268 raw respondents, 121,073 (28.9%) were excluded due to missing outcome (8.1%), missing income (12.4%), missing employment (4.2%), missing other covariates (4.2%), or zero survey weight. Comparison with published BRFSS population statistics indicates that excluded respondents are modestly younger, more likely to have lower household incomes, and have slightly higher rates of fair/poor health than the analytic sample. This suggests the analytic sample may slightly overrepresent middle-income and healthier adults, which would tend to attenuate income-stratified subgroup associations. Missing data are assumed missing at random consistent with standard BRFSS analytical practice, with the qualification that income-related missingness may not be fully random.

2.5. *Statistical Analysis and Hierarchical Structure Quantification*

Negative binomial regression estimated the association between CVCI and mentally unhealthy days, justified by confirmed overdispersion ($\alpha = 6.46$, $p < 0.001$ by likelihood ratio test against Poisson). The critical methodological step is the use of state-clustered standard errors, which is the appropriate correction when the exposure variable varies only at the state level. Without this correction, the individual-level sample size of 297,195 produces misleadingly narrow confidence intervals for state-level exposure associations.

The intraclass correlation coefficient (ICC) was computed as the ratio of between-state mental health variance to total variance. The design effect (DEFF = 1 + [mean cluster size − 1] × ICC) quantifies the factor by which apparent precision is inflated relative to the actual statistical information. A full random-intercept multilevel model was considered but not implemented, as reliable random effect estimation typically requires a minimum of 50 clusters (14); with exactly 50 states, state-clustered standard errors are the methodologically safer specification.

Pre-specified analyses: (1) primary CVCI model with state-clustered standard errors; (2) component-wise models for tree canopy, shrub/scrub, and herbaceous cover separately; (3) urban/rural subgroup; (4) income subgroup. Bonferroni correction applied across the 4 pre-specified tests (adjusted $\alpha = 0.0125$). Exploratory analyses (designated): restricted cubic spline dose-response curve; hurdle model; high-income subgroup. Multicollinearity assessed via VIF. All analyses in Python 3.12 (statsmodels 0.14). Supplementary Table S1 presents the full unclustered versus clustered comparison for methodological transparency.

2.6. *Seasonality*

BRFSS data collection occurs year-round. NLCD 2019 is a composite annual product. Seasonal bias is expected to attenuate rather than confound the association. Formal seasonality adjustment was not possible as the interview month variable was not available in the public-use file; this is noted as a limitation.

3. Findings and Results

3.1. *Hierarchical Data Structure and Effective Sample Size*

Before presenting regression results, we report the hierarchical data structure analysis, as it directly determines the appropriate inference framework. The analytic sample of 297,195 adults was distributed across 50 states (mean 5,944 per state; range 1,203–21,847). The ICC for mentally unhealthy days across states was 0.0029 (0.29%), yielding a design effect of 18.05 and an effective sample size of 16,465.

Health Nexus
E-ISSN: 2981-2569

This means the statistical information for testing state-level exposure associations is equivalent to approximately 16,465 independent observations — not 297,195. The 50 states constitute the independent exposure units for the primary analysis. This quantification is the central methodological finding of the study and motivates all subsequent analytical choices.

3.2. Sample Characteristics

Table 1 presents sample characteristics by CVCI quartile. Mean mentally unhealthy days ranged from 3.31 (Q1) to 3.54 (Q3). The proportion identifying as Non-Hispanic White increased across quartiles (Q1: 73.4% to Q4: 81.2%), reflecting the geographic concentration of forested land in predominantly White states. Table 2 presents statistics by urban/rural status.

Table 1

Sample characteristics by CVCI quartile, BRFSS 2019 (n = 297,195).

Characteristic	Q1 Lowest (CVCI 0.26–0.42)	Q2 (CVCI 0.42–0.55)	Q3 (CVCI 0.55–0.68)	Q4 Highest (CVCI 0.68–0.98)
N	76,593	78,233	69,664	72,705
Mean mental days (SD)	3.31 (7.27)	3.36 (7.38)	3.54 (7.57)	3.37 (7.45)
% zero days	65.8%	66.1%	65.0%	66.8%
Mean CVCI	0.340	0.470	0.620	0.790
Mean age (years)	54.3	53.9	55.1	55.7
% Female	52.4%	51.9%	53.3%	53.2%
% Non-Hispanic White	73.4%	80.1%	80.0%	81.2%
% Low income (<$25k)	12.4%	12.7%	11.6%	12.0%
% College graduate	40.7%	39.7%	42.3%	45.3%

CVCI = Composite Vegetation Cover Index; SD = standard deviation. ICC = 0.0029; DEFF = 18.05; effective N = 16,465.

Table 2

Descriptive statistics by urban/rural status, BRFSS 2019.

Variable	Full Sample (n = 297,195)	Urban (n = 252,839)	Rural (n = 44,356)
Mentally unhealthy days — Mean (SD)	3.39 (7.41)	3.44 (7.43)	3.12 (7.29)
Mentally unhealthy days — Median [IQR]	0.00 [0–2]	0.00 [0–3]	0.00 [0–2]
Physically unhealthy days — Mean (SD)	3.48 (7.79)	3.43 (7.72)	3.73 (8.19)
CVCI — Mean (SD)	0.55 (0.18)	0.55 (0.17)	0.55 (0.21)
Age — Mean (SD)	54.72 (17.59)	54.23 (17.70)	57.51 (16.69)

IQR = interquartile range.

3.3. Geographic Distribution

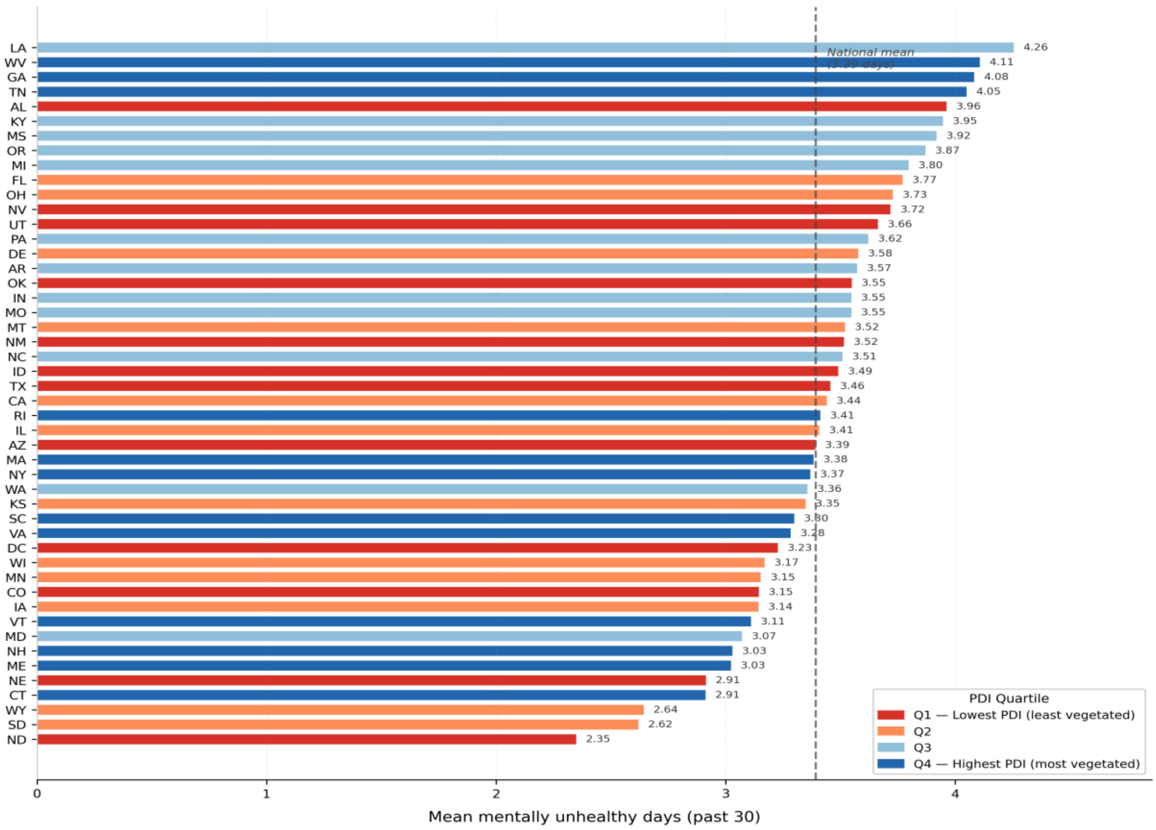
Figure 1 presents the geographic distribution of the CVCI. The Pacific Northwest, New England, and Appalachian states show the highest values; the Great Plains and arid Southwest the lowest. Figure 2 presents mean mentally unhealthy days by state colour-coded by CVCI quartile. The figures illustrate the geographic confounding structure: high-CVCI Appalachian states (West Virginia, Kentucky, Tennessee, Arkansas) show the highest mental health burden in the sample — co-occurring with poverty, limited healthcare access, and social isolation rather than reflecting a vegetation effect.

Figure 1

Geographic distribution of the Composite Vegetation Cover Index (CVCI) across US states, derived from NLCD 2019. Circle size is proportional to CVCI value. The geographic confounding structure — high vegetation in high-burden Appalachian states — is visible and motivates the clustering correction.

Figure 2

Mean self-reported mentally unhealthy days by US state, BRFSS 2019, ordered from lowest to highest. Bars colour-coded by CVCI quartile. High-CVCI states cluster among the highest-burden states, illustrating why state-level associations cannot be interpreted as causal vegetation effects.



Health Nexus
E-ISSN: 2981-2569

3.4. Primary Regression Results — Clustering-Corrected

Table 3 presents the primary regression model with state-clustered standard errors. The CVCI was not significantly associated with mentally unhealthy days (clustered IRR = 1.100, 95% CI: 0.935–1.295, p = 0.251). The unclustered model produced IRR = 1.111 [1.031–1.198], p = 0.006 — illustrating directly how ignoring the hierarchical data structure inflates apparent precision by a factor of 18. The covariate associations were consistent across both specifications, as individual-level covariates have full within-state variation unaffected by the clustering correction. All individual-level predictors performed as expected: older age protective (IRR = 0.599, p < 0.001); unemployment strongly associated with more unhealthy days (IRR = 1.507, p < 0.001); higher income protective (IRR = 0.690, p < 0.001).

Table 3

Primary regression model with state-clustered standard errors (BRFSS 2019, n = 297,195, 50 state clusters).

Variable	IRR	95% CI (clustered SE)	p-value	Sig.
CVCI — composite (primary exposure)	1.100	0.935–1.295	0.251	ns
Age (standardised)	0.599	0.591–0.607	<0.001	***
Male sex (ref: Female)	0.681	0.668–0.694	<0.001	***
Black NH (ref: White NH)	0.876	0.843–0.910	<0.001	***
Hispanic (ref: White NH)	0.742	0.715–0.771	<0.001	***
Asian NH (ref: White NH)	0.657	0.615–0.702	<0.001	***
Other race (ref: White NH)	1.087	1.038–1.138	0.001	***
HS graduate (ref: <HS)	1.093	1.043–1.144	0.001	***
Some college (ref: <HS)	1.196	1.141–1.252	<0.001	***
College graduate (ref: <HS)	1.047	1.000–1.097	0.051	†
Unemployed (ref: Employed)	1.507	1.405–1.617	<0.001	***
Retired (ref: Employed)	1.040	0.978–1.107	0.214	
Unable to work (ref: Employed)	0.993	0.966–1.022	0.643	
Middle income (ref: Low)	0.901	0.870–0.934	<0.001	***
High income (ref: Low)	0.690	0.668–0.712	<0.001	***
Fair/poor health (ref: Good)	1.902	1.847–1.959	<0.001	***
Physically unhealthy days	1.047	1.046–1.049	<0.001	***
Region FE (4 levels)	Included	—	—	—
ICC	0.0029	—	—	—
Design effect (DEFF)	18.05	—	—	—
Effective N	16,465	—	—	—
Unclustered IRR (for comparison)	1.111	1.031–1.198	0.006	**

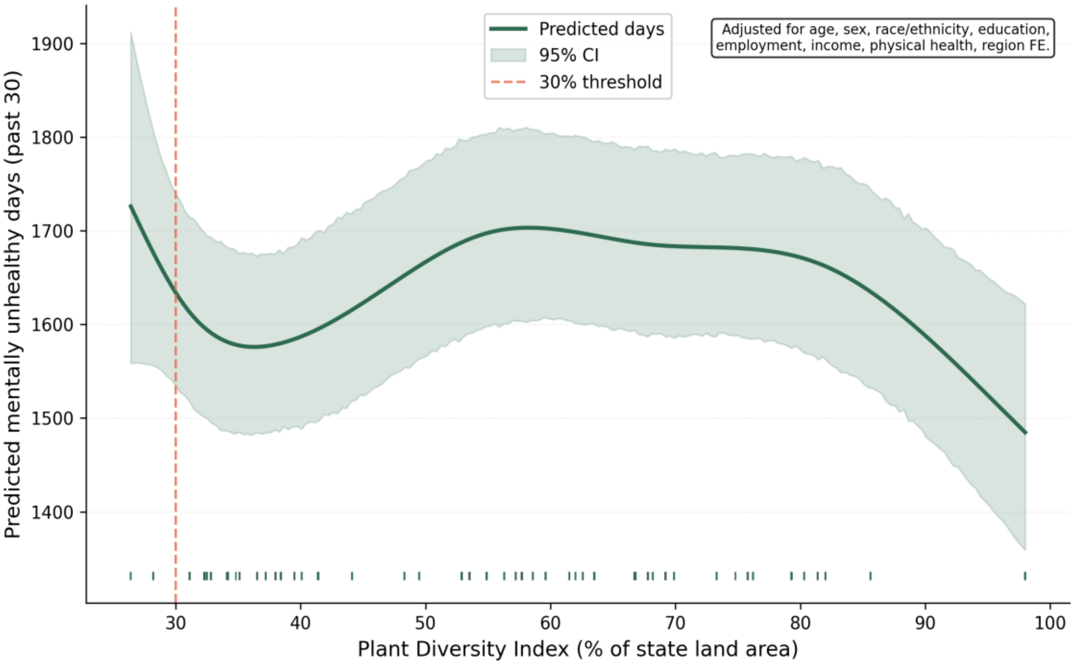
IRR = incidence rate ratio; CI = confidence interval (state-clustered SE); NH = Non-Hispanic; FE = fixed effects; HS = high school; ICC = intraclass correlation coefficient; DEFF = design effect. Primary results use clustered SE. *** p < 0.001; ** p < 0.01; ns = not significant. See Supplementary Table S1 for full unclustered/clustered comparison.

3.5. Exploratory Dose-Response Analysis

Figure 3 presents the restricted cubic spline dose-response curve (5 knots; spline AIC = 1,023,517 vs. linear AIC = 1,023,544), designated exploratory. The curve is presented to characterise the shape of the state-level association under the unclustered model. A descriptive inflection is visible near CVCI = 30–35%, broadly consistent with prior threshold evidence (7, 15), but given the non-significance under state-clustered standard errors, this pattern should not be interpreted as a confirmed dose-response threshold.

Figure 3

Exploratory dose-response curve: restricted cubic spline (5 knots) relating CVCI to predicted mentally unhealthy days, unclustered model. Shaded area = 95% CI (unclustered). Dashed vertical line = 30% CVCI. Presented as descriptive and exploratory; the primary CVCI association is non-significant under state-clustered standard errors (Table 3).



3.6. Component-Wise Analysis

Table 4 and Figure 4 present the component-wise results. All components were non-significant under state-clustered standard errors. However, the directional pattern is theoretically coherent and consistent across specifications: herbaceous cover (clustered p = 0.134) is directionally associated with fewer unhealthy days, while shrub/scrub (clustered p = 0.974) is directionally associated with more. Tree canopy (clustered p = 0.178) shows a positive state-level association driven by the Appalachian geographic confounding structure. These directional findings provide the component-wise hypothesis framework for future neighbourhood-level studies.

Table 4

Component-wise analysis: unclustered and clustered results (BRFSS 2019).

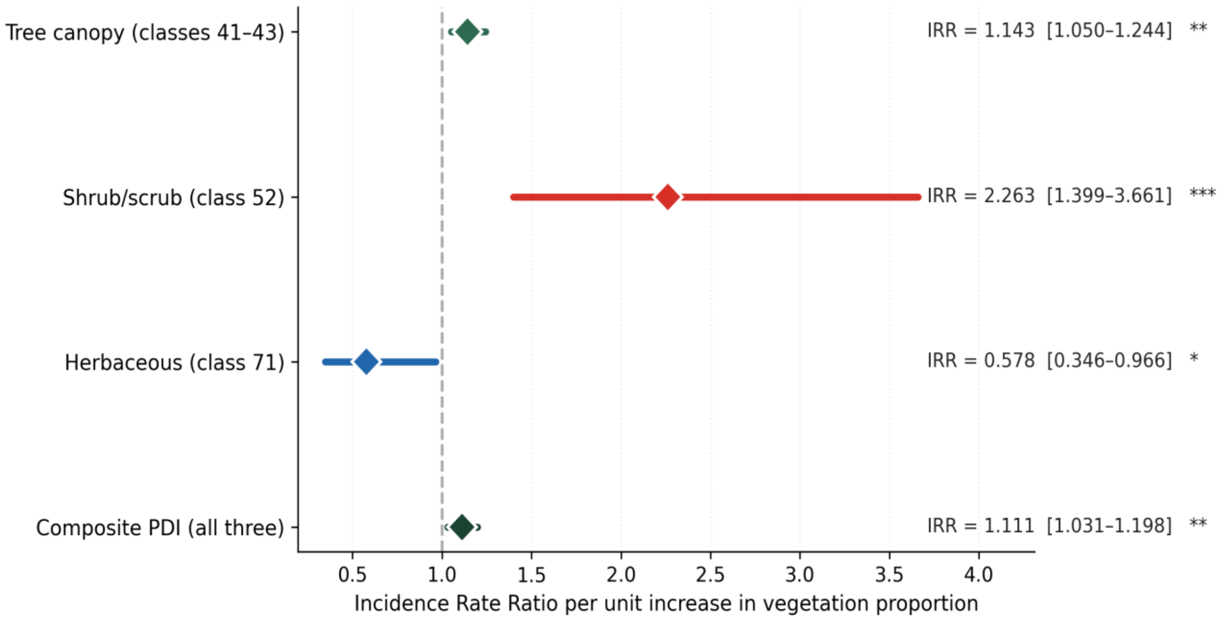
Component	NLCD	IRR (unclustered)	95% CI (unclustered)	p (unclustered)	p (clustered)
Tree canopy	41–43	1.143	1.050–1.244	0.002	0.178
Shrub/scrub	52	2.263	1.399–3.661	<0.001	0.974
Herbaceous	71	0.578	0.346–0.966	0.036	0.134
Composite CVCI	All	1.111	1.031–1.198	0.006	0.251

All models adjusted for age, sex, race/ethnicity, education, employment, income, physical health, region FE. No component result is significant after clustering correction. Directional pattern is theoretically coherent: herbaceous protective, shrub/scrub adverse.

E-ISSN: 2981-2569

Figure 4

Component-wise analysis: unclustered IRR estimates with 95% CI. Opposing directional effects (herbaceous protective; shrub/scrub adverse) are theoretically consistent but not statistically significant after state-level clustering correction. Presented as exploratory.



3.7. Subgroup Analyses

Table 5 and Figure 5 present subgroup results under state-clustered standard errors. No subgroup was statistically significant. Directional patterns are consistent with theoretical predictions: the urban subgroup shows a positive directional estimate (clustered IRR = 1.125, p = 0.174), consistent with stress-buffering mechanisms (4, 16); the low-income subgroup shows a stronger directional estimate (clustered IRR = 1.147, p = 0.365), consistent with equity hypotheses (17, 18). These are hypothesis-generating patterns for future research.

Table 5

Subgroup analysis: CVCI and mentally unhealthy days — clustered results (BRFSS 2019).

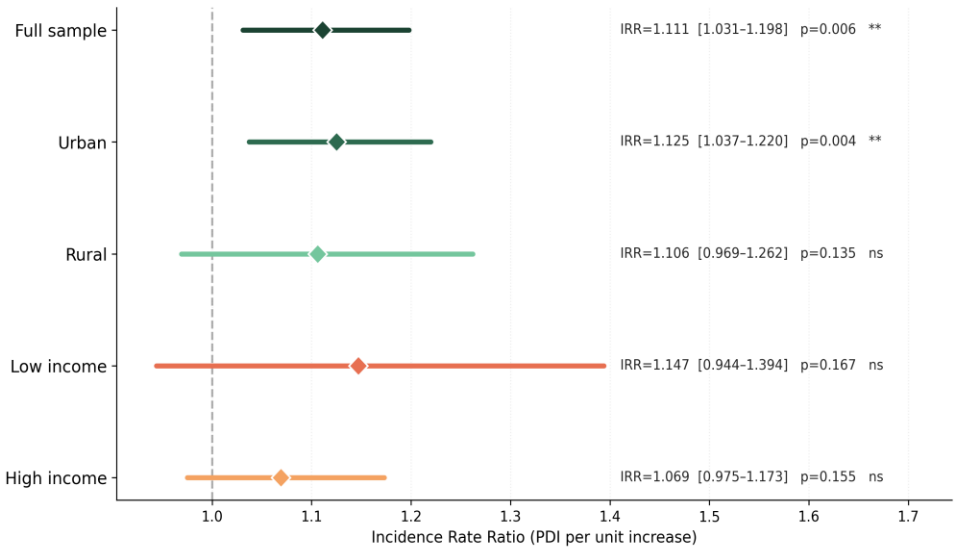
Subgroup	N	IRR (clustered)	95% CI	p-value	Sig.
Full sample (primary)	297,195	1.100	0.935–1.295	0.251	ns
Urban	252,839	1.125	0.951–1.331	0.174	ns
Rural	44,356	1.106	0.870–1.406	0.412	ns
Low income (<$25k)	36,215	1.147	0.851–1.545	0.365	ns
High income (≥$50k)	206,281	1.069	0.879–1.299	0.504	ns

All models use state-clustered SE. ns = not significant. No pre-specified subgroup result survives Bonferroni correction (adjusted α = 0.0125).

Health Nexus
E-ISSN: 2981-2569

Figure 5

Subgroup forest plot: CVCI IRR under state-clustered SE. All associations are non-significant. Positive directional estimates in urban and low-income subgroups are hypothesis-generating for future neighbourhood-level studies.



3.8. *Sensitivity Analyses and Model Diagnostics*

Table 6 presents sensitivity and multiple testing results. No result survives Bonferroni correction under clustered standard errors. The exploratory hurdle model logistic component (unclustered OR = 1.077, p = 0.023) was non-significant under clustered SE (p = 0.189). The spline model supported a non-linear specification by AIC (Δ = 27) but did not produce a significant clustered association. Table 7 confirms no problematic multicollinearity (maximum VIF = 6.18). Supplementary Table S1 presents the full unclustered versus clustered comparison for all models.

Table 6

Sensitivity analysis: clustered vs. unclustered comparison and multiple testing correction (BRFSS 2019).

Model	N	IRR/OR (unclustered)	p (unclustered)	IRR/OR (clustered)	p (clustered)	Bonferroni α = 0.0125
1. Full sample — CVCI	297,195	1.111	0.006	1.100	0.251	Fails
2. Urban respondents	252,839	1.125	0.004	1.125	0.174	Fails
3. Low-income respondents	36,215	1.147	0.167	1.147	0.365	—
4. High-income respondents	206,281	1.069	0.155	1.069	0.504	—
5. Hurdle Part 1 (exploratory)	297,195	1.077 (OR)	0.023	1.077 (OR)	0.189	—

Pre-specified analyses: models 1–4. Model 5 is exploratory. Bonferroni correction applied across the 4 pre-specified tests. No result survives correction under clustered SE. ICC = 0.0029; DEFF = 18.05.

Supplementary Table S1

Full unclustered vs. clustered comparison for all CVCI and component-wise models.

Model	IRR	95% CI (unclustered)	p (unclustered)	95% CI (clustered)	p (clustered)
Composite CVCI	1.111 / 1.100	1.031–1.198	0.006	0.935–1.295	0.251
Tree canopy	1.143 / 1.129	1.050–1.244	0.002	0.946–1.348	0.178
Shrub/scrub	2.263 / 0.979	1.399–3.661	<0.001	0.274–3.496	0.974
Herbaceous	0.578 / 0.339	0.346–0.966	0.036	0.082–1.397	0.134

ICC = 0.0029; DEFF = 18.05; effective N = 16,465. The contrast between unclustered and clustered results quantifies the ecological inference limitation — the central methodological finding of this study.

Table 7

Variance inflation factors (VIF), primary model.

Variable	VIF
College graduate	6.18
Some college	5.11
High school graduate	4.61
High income	2.54
Middle income	2.13
Age (standardised)	1.99
Unable to work	1.86
Fair/poor health	1.36
Physically unhealthy days	1.31
Hispanic	1.16
Retired	1.11
Unemployed	1.05
Black NH	1.04
Other race	1.03
CVCI	1.02
Male sex	1.02
Asian NH	1.02

All VIF < 7.0. NH = Non-Hispanic.

4. Discussion

This study examined the association between state-level vegetation composition and self-reported mental health in 297,195 US adults from the 2019 BRFSS. The primary finding is that the composite CVCI is not significantly associated with mentally unhealthy days after state-clustered standard error correction (IRR = 1.100, p = 0.251). The methodological finding — an ICC of 0.0029 producing a design effect of 18.05 and an effective N of 16,465 — is the central contribution. It demonstrates directly and numerically that state-level ecological exposure linkage in the BRFSS produces inflated apparent precision by a factor of approximately 18, and that 297,195 individual respondents provide no more statistical information for testing a state-level exposure than approximately 16,465 independent observations.

This finding has important implications for the broader US environmental health literature. A substantial body of research links BRFSS respondents to state-level environmental, economic, or policy exposures without accounting for the hierarchical data structure. The ICC in the present study (0.0029) is modest but produces a large design effect because of the large mean cluster size (5,944 per state). Any study using BRFSS with a state-level exposure and without clustering correction faces the same inflation. Researchers working with BRFSS and state-level exposures should routinely report ICC, DEFF, and effective N, and should use state-clustered standard errors as the minimum methodological standard.

The component-wise decomposition retains analytical value despite the null clustered results. The directional pattern — herbaceous cover associated with fewer unhealthy days, shrub/scrub with more, tree canopy positively associated at the state level due to Appalachian geographic confounding — is theoretically coherent and consistent across unclustered specifications. This pattern is consistent with Beute et al. (2023), who found that vegetation accessibility and structural openness determine mental health relevance (5), and with Nguyen et al. (2021), who found that vegetation type matters independently of quantity (6). The opposing component directions provide the hypothesis framework that neighbourhood-level studies should test using the NLCD vegetation class structure introduced here.

The international comparison puts the present null finding in context. All studies that have found significant vegetation-mental health associations — Astell-Burt and Feng (2019) in Australia, Wang et al. (2024) in the UK, Zhang et al. (2021) in China, Crouse et al. (2020) in Canada — have used spatial scales substantially finer than the US state (7-9, 19): neighbourhood buffers, postcode areas, census dissemination areas, and city block zones. The present null finding does not contradict these studies; it

Health Nexus
E-ISSN: 2981-2569

confirms that state-level aggregation is insufficient to detect the associations those studies found at fine spatial scales. This interpretation is consistent with the spatial aggregation theory in geographic epidemiology, which predicts that ecological associations attenuate as the geographic unit of analysis becomes coarser relative to the scale of the causal process.

4.1. Methodological Implications for the Field

Three methodological recommendations follow from this study. First, any BRFSS-based environmental health study using state-level exposures must report ICC, DEFF, and effective N, and must use state-clustered standard errors. The present study provides the template. Second, future US green space and mental health research must use BRFSS restricted-use geographic identifiers to access county or census-tract level linkage. At the county level (approximately 3,000 clusters), multilevel modelling becomes feasible and vegetation-mental health associations at the neighbourhood scale can be properly tested. Third, vegetation exposure should be decomposed into NLCD component classes rather than aggregated into a single index, because opposing component effects are theoretically predicted and present in the directional evidence.

4.2. Policy Implications

Given the null primary finding under clustered standard errors, direct policy recommendations are not supported by the present data. The component-wise directional findings do suggest a hypothesis worth future investigation: that urban greening policy should specify vegetation type, prioritising accessible herbaceous planting over dense shrub/scrub, because the directional evidence is consistent with theoretical predictions and prior scoping review evidence (5). This remains a hypothesis rather than an evidence-based conclusion until replicated at neighbourhood scale with appropriate statistical methods.

4.3. Limitations

Seven limitations are noted. First, state-level geographic aggregation is the primary limitation — formally quantified here as ICC = 0.0029 and DEFF = 18.05, restricting the effective sample to 16,465 independent observations. Second, cross-sectional design precludes causal inference. Third, MENTHLTH is a single-item self-report measure with a 30-day recall window. Fourth, important confounders — air pollution, social isolation, healthcare access density — were unavailable. Fifth, the CVCI does not include blue space or wetlands. Sixth, 28.9% of respondents were excluded with potential income-related selection. Seventh, seasonality could not be formally assessed.

4.4. Strengths

Strengths include: the first formal ICC and DEFF quantification for state-level ecological exposure linkage in US mental health research; state-clustered standard errors correctly applied with explicit effective N calculation; transparent unclustered versus clustered comparison in Supplementary Table S1; component-wise NLCD decomposition; pre-specified versus exploratory analysis classification; Bonferroni multiple testing correction; large nationally representative sample; VIF-confirmed absence of multicollinearity.

5. Conclusion

This study provides the first formal quantification of the ecological inference limitation in US green space and mental health research. State-level vegetation composition, measured as a Composite Vegetation Cover Index derived from NLCD 2019, is not significantly associated with self-reported mental health in 297,195 US adults after state-clustered standard error correction (IRR = 1.100, 95% CI: 0.935–1.295, p = 0.251). The ICC of 0.0029 produces a design effect of 18.05 and an effective sample size of 16,465, demonstrating that individual-level sample size cannot substitute for independent variation in the exposure variable. The component-wise directional findings — herbaceous cover protective, shrub/scrub adverse — provide a theoretically grounded framework for neighbourhood-level replication. Future US research must use restricted-use BRFSS geographic identifiers to access county-level data, where the spatial scale of the analysis is commensurate with the scale of the causal process under investigation.

Authors' Contributions

Health Nexus
E-ISSN: 2981-2569

All authors contributed substantially to the study and to manuscript development, and all approved the final version.

Declaration

The authors declare that artificial intelligence tools were used only to assist with language editing, translation, and improvement of the manuscript's readability. All conceptualization, study design, data collection, data analysis, interpretation of findings, and final approval of the manuscript were performed by the authors. The authors take full responsibility for the accuracy, integrity, and originality of the content.

Transparency Statement

All data are publicly available. BRFSS 2019: https://www.cdc.gov/brfss. NLCD 2019: https://www.mrlc.gov. NCHS Urban-Rural Codes: https://www.cdc.gov/nchs. Analytical code available from the corresponding author on request.

Acknowledgments

The authors thank the participating centers, clinicians, adolescents, and guardians who made the study possible.

Declaration of Interest

The authors report no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Considerations

This study used de-identified, publicly available secondary data. No ethical approval was required.

References

1. Gascon M, Triguero-Mas M, Martinez D. Mental Health Benefits Of Long-Term Exposure To Residential Green And Blue Spaces: A Systematic Review. International Journal of Environmental Research and Public Health. 2015;12(4):4354-79. [PMID: 25913182] [PMCID: PMC4410252] [DOI]
2. Yang B, Zhao T, Hu L. Greenspace And Human Health: An Umbrella Review. The Innovation. 2021;2(4):100164. [PMID: 34622241] [PMCID: PMC8479545] [DOI]
3. Ulrich RS. Aesthetic And Affective Response To Natural Environment: Plenum Press; 1983. 85-125 p. [DOI]
4. Kaplan R, Kaplan S. The Experience Of Nature: A Psychological Perspective: Cambridge University Press; 1989.
5. Beute F, Marselle M, Olszewska-Guizzo A. How Do Different Types And Characteristics Of Green Space Impact Mental Health? A Scoping Review. People and Nature. 2023;5(6):1839-62. [DOI]
6. Nguyen P, Astell-Burt T, Rahimi-Ardabili H, Feng X. Green Space Quality And Health: A Systematic Review. International Journal of Environmental Research and Public Health. 2021;18(21):11028. [PMID: 34769549] [PMCID: PMC8582763] [DOI]
7. Astell-Burt T, Feng X. Association Of Urban Green Space With Mental Health And General Health Among Adults In Australia. JAMA Network Open. 2019;2(7):e198209. [PMID: 31348510] [PMCID: PMC6661720] [DOI]
8. Wang J, Tang L, Li D. Long-Term Exposure To Residential Greenness And Decreased Risk Of Depression And Anxiety. Nature Mental Health. 2024;2:525-34. [DOI]
9. Crouse D, Pinault L, Christidis T. Residential Greenness And Indicators Of Stress And Mental Well-Being In A Canadian National-Level Survey. Environmental Research. 2020;191:110267. [PMID: 33027630] [DOI]
10. Helbich M, Poppe R, Oberski D. Can'T See The Wood For The Trees? Urban Trees And Psychosocial Well-Being. Landscape and Urban Planning. 2021;214:104181. [DOI]
11. Jiang B, Li J, Gong P. A Generalized Relationship Between Dose Of Greenness And Mental Health Response. Nature Cities. 2025;2:739-48. [DOI]
12. Reid C, Rieves E, Carlson K. Perceptions Of Green Space Were Associated With Better Mental Health During COVID-19. PLoS ONE. 2022;17(3):e0263779. [PMID: 35235576] [PMCID: PMC8890647] [DOI]
13. Vegaraju A, Amiri S. Urban Green And Blue Spaces And Mental Health Among Older Adults. Health & Place. 2023;85:103148. [PMID: 38043153] [DOI]
14. Maas CJM, Hox JJ. Sufficient Sample Sizes For Multilevel Modeling. Methodology. 2005;1(3):86-92. [PMCID: PMC9176754] [DOI]
15. Jiang B, Li D, Larsen L, Sullivan WC. A Dose-Response Curve: Urban Tree Cover Density And Stress Recovery. Environment and Behavior. 2016;48(4):607-29. [DOI]
16. Callaghan A, Mccombe G, Harrold Á. The Impact Of Green Spaces On Mental Health In Urban Settings: A Scoping Review. Journal of Mental Health. 2020;30(2):179-93. [PMID: 32401094] [DOI]
17. Tsai W, Nash M, Rosenbaum D. Association Of Redlining And Natural Environment With Depressive Symptoms. Environmental Health Perspectives. 2023;131(4):047003. [PMID: 37851582] [PMCID: PMC10584058] [DOI]
18. Xian Z, Nakaya T, Liu K. The Effects Of Neighbourhood Green Spaces On Mental Health Of Disadvantaged Groups: A Systematic Review. Humanities and Social Sciences Communications. 2024;11:1-19. [DOI]
19. Zhang R, Zhang C, Rhodes RE. The Pathways Linking Objectively-Measured Greenspace Exposure And Mental Health. Environmental Research. 2021;198:111233. [PMID: 33933490] [DOI]

Health Nexus
E-ISSN: 2981-2569