Sustainable Cities and Society 99 (2023) 104903

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Rethinking walkability: Exploring the relationship between urban form and neighborhood social cohesion

Andrew Sonta [a,†,*], Xiaofan Jiang [b]

[a] Data Science Institute, Columbia University, 550 West 120th Street, Suite 1401, New York, NY, USA
[b] Intelligent and Connected Systems Lab, Department of Electrical Engineering, Columbia University, New York, NY 10027, USA

ARTICLE INFO

Keywords:
Walkability
Urban design
Social capital/cohesion
Sustainable and resilient cities
People and environment
City structure
Structural equation modeling
New urban/smart growth

ABSTRACT

Recent research has investigated the importance of both walkable urban design and social cohesion. Social cohesion has been shown to have broad social and health benefits, and scholars have hypothesized that walkable urban design can influence cohesion, though evidence remains limited. In this work, we leveraged a data-driven approach that broke down design factors related to walkable design and investigated their impact on cohesion. We used a US-wide open urban form dataset to characterize walkable urban design, and we used an open survey dataset that measured cohesion and demographics with a total sample size of 9670 in six US cities. We leveraged partial least squared structural equation modeling for statistical analysis. We found, controlling for demographics, that land use diversity had a significant positive impact on social cohesion. We also found that physical density, social density, and transit connectedness had significant negative impacts on cohesion, though this association is largely driven by the very dense neighborhoods in cities. These findings shed light on different theories of the built environment, offering insights for designers, engineers, and policymakers interested in the social effects of the built environment.

1. Introduction

The intersection of environmental and social sustainability in the urban built environment has received growing attention in recent years. In the face of climate change and increased urbanization, sustainable mobility—and particularly, active mobility including walkability—has been increasingly recognized as an important goal for the design of cities (Jardim and de Castro Neto, 2022; Loo, 2021; Moreno et al., 2021; Sonta and Jain, 2020; Gao et al., 2022). Working toward walkability in cities involves the provision of mobility infrastructure and social infrastructure in a manner that encourages walking as a viable transportation option (Carr et al., 2010; Liao et al., 2020; Huang and Khalil, 2022). The environmental benefits of walking as a mode of transit are well-understood, but it is important to note that scholars have long postulated that there are social benefits to walkability in cities and neighborhoods as well (Loo et al., 2017; Jun and Hur, 2015; Koohsari et al., 2021; Leyden, 2011; Rogers et al., 2013; Lee and Tan, 2019). A key underlying social benefit commonly associated with walkability is social cohesion, which involves the strength of connections among people. However, due to the relative difficulty of measuring social outcomes at scale, data-driven evidence for the social impacts of walkable urban design is limited (Mazumdar et al., 2018). Nonetheless, it remains an important task to identify how urban mobility infrastructure and urban form impact social outcomes such as social cohesion, as this can help urban designers, city officials, and policymakers better understand how urban environments can be designed and managed for human-oriented goals.

The importance of socially cohesive communities has received growing attention among scholars and practitioners of the built environment. Social cohesion has been argued to be a collective good in and of itself (Coleman, 1994), it has been connected to both physical and mental health (House et al., 1988; Kawachi and Berkman, 2001), it has been shown to improve the ability of communities to respond and adapt to external shocks such as natural disasters (Aldrich and Meyer, 2014; Kawachi and Berkman, 2015), and it is often considered as an important attribute of overall wellbeing (Delhey and Dragolov, 2016). There are many factors that could be expected to impact the cohesiveness of communities; one such factor is the design of the built environment.

* Corresponding author.
E-mail address: andrew.sonta@epfl.ch (A. Sonta).
† Present address: ETHOS Lab, Civil Engineering Institute, École polytechnique fédérale de Lausanne, HBL 1 3B (Halle Bleue), Pass. du Cardinal 13b, 1700 Fribourg, Switzerland.

https://doi.org/10.1016/j.scs.2023.104903
Received 28 April 2023; Received in revised form 4 August 2023; Accepted 28 August 2023
Available online 29 August 2023
2210-6707/

The creation of infrastructure in the built environment involves many policy, design, and engineering decisions that influence the ways in which we experience our cities. One such experience is the social experience, broadly involving interactions and relationships that manifest in built spaces. Different theories have emerged on the best way to provide urban infrastructure vis-à-vis social factors. In recent years, there has been renewed interest in the design philosophy most often referred to as "new urbanism," a framework for design that generally emphasizes dense, walkable, and mixed-use cities (Ellis, 2010). The "15-minute city" concept builds upon the new urbanism philosophy and argues for local neighborhoods and infrastructure that enable residents to reach their required destinations within a 15-mintute walk or bike ride from their homes (Moreno et al., 2021; Allam et al., 2022). Proponents of new urbanism and the 15-minute city have argued that walkable design would encourage more social interaction among city dwellers, helping to improve social cohesion (Lund, 2003). Case-study based research has offered some limited evidence supporting this notion (e.g., that neighborhoods perceived as more walkable also have higher social capital, as reported in a survey) (Rogers et al., 2013). However, a systematic review focusing on the relationships between new urbanism design characteristics and social capital outcomes found general support for this link but also conflicting evidence (i.e., 155 total relationships, but only 66 that were significant and only 84 in the expected direction) (Mazumdar et al., 2018). At the same time, a different design philosophy has argued against pure urbanization, suggesting that high density development would socially overwhelm city dwellers, increasing the feeling of anonymity in cities and making it difficult to form social connections (Nguyen, 2010). The evidence for this line of argument is also limited.

These two philosophies of urban design are often viewed as competing: new urbanists call for dense, walkable cities, and their counterparts warn against it. But it is important to note that these are not diametrically opposed points of view; there are subtleties in each argument. A new urbanist neighborhood involves more than just density—it also critically depends on the diversity of land uses, the physical design of streets and sidewalks, and many other factors. And the critique of pure urbanization does not comment on other aspects of urban design besides density of people. As a result, we need a more nuanced understanding of how urban design factors influence social experiences in cities.

In this work, we identified the outcome of interest to be social cohesion, which we defined through survey questions very commonly used in the social science literature, as originally introduced by Sampson et al. (1997). We note that discussion of social factors in the built environment can use many different terms and can include varied concepts, as discussed in the Background section below. We also defined the independent variables of interest to be those design features of the built environment commonly associated with walkability. Limiting our analysis of the built environment in this way narrowed the scope of analysis for interpretability while reflecting important questions prevalent in theory and in the literature, as we discuss in detail below.

In this paper, we first discuss the relevant literature on human-centric neighborhood sustainability, social cohesion theory and measurement, the connections between social cohesion and the design of walkable urban infrastructure, and metrics used to measure walkability (Section 2: Background). Then, using a survey from six cities measuring social cohesion combined with open data on urban form, we present a nuanced statistical analysis of walkable urban design and social cohesion controlling for demographic factors using the Partial Least Squares Structural Equation Modeling (PLS-SEM) framework (Section 3: Data and Methodology). We present the results of our statistical analysis, which motivated us to look more closely at the interactions between land use diversity and physical density as well as multi-group effects across cities. We also discuss the implications of our findings for the urban planning, engineering, and policy disciplines (Section 4: Results and Discussion). The primary goal of this work is to leverage large open datasets to expand our knowledge on the empirical links between urban design and planning—specifically as it relates to walkable urban form—and social cohesion outcomes that are important to social health, wellbeing, and resilience.

2. Background

2.1. Human-centric neighborhood sustainability

In recent work on sustainable city development, researchers have noted the importance of explicitly studying the human experience as an integral component of overall sustainability. The quality of our experiences and well-being is an important goal in and of itself, but it also forms a nexus with multiple aspects of a sustainable urban environment—urban form, urban energy, the urban heat island, air quality, and flooding, to name a few.

In engineering and science-based research, the interaction between humans and the urban built environment has typically been studied by integrating the human perspective into analysis of the physical aspects of cities. For example, researchers have developed frameworks for analyzing how urban form influences pedestrian exposure to air quality (Miao et al., 2020), flood risk (Zhu et al., 2023), urban greenery (Hua et al., 2022), and the urban heat island (Yu and de Dear, 2022). In these works, which provide valuable insight on human-built interactions, the human perspective is explicitly considered in the context of the built environment, but it is typically not directly quantified. On the other hand, from the sociological perspective, research in this area often involves the collection of data on human factors considering the context of the urban environment, often through specific case studies (Mazumdar et al., 2018; Yang et al., 2023). In this research, we aim to integrate these two perspectives by investigating the links between two large open datasets: urban data theorized to impact social aspects and social data that can be explicitly linked to urban locations.

2.2. Social cohesion theory and measurement

In sociological theory, there has long been interest in characterizing the set of resources that come with the ability to connect with other people (Kawachi and Berkman, 2015). Often falling under the umbrella term social capital (*capital* because it can be viewed as a resource similar to economic forms of capital), this resource has been shown to be associated with positive outcomes in the areas of health and wellbeing (House et al., 1988). This resource is an interdisciplinary concept, with contributions coming from sociology, political science, population health, and other fields, and we continue to lack both a straightforward definition for it as well as an agreed-upon means for measuring it (Kawachi and Berkman, 2015). What is known is that social-network-based connections have been shown to have value for individuals and that there are many ways to conceptualize the pathways for these benefits.

The broad notion of social capital is based on network structure and is typically measured when access to social network structure data is available (Moore et al., 2013). While this is often done when studying small groups or organizations, this data can be difficult to obtain as the scope of analysis expands. A related concept known as social cohesion is intended to represent many of the same ideas but is typically measured through surveys that inquire about the emergent properties of network connections (e.g., the trust that individuals put in their neighbors) (Kawachi and Berkman, 2015; Sampson et al., 1997). The key benefit of the social cohesion approach is that it can be measured through large-scale surveys when exact information on social network structure can be difficult to find. While the cohesion approach has been criticized for straying from the network-based definition of social capital (Carpiano, 2008), we note that it is commonly used in research practice and can often be the only feasible measurement item for large-scale analyses. *In this work, we hereafter use the terms neighborhood social cohesion, social* *cohesion, or simply cohesion to represent the primary outcome of interest throughout the study.*

2.3. Connections between walkable urban design and social cohesion

Over the past few decades, theories have emerged around the ways in which the built environment impacts social cohesion and capital. These theories often discuss different aspects of design, including issues related to density, land use diversity, transit access, greenspace, and other factors. Theories of the effect of density and urbanization on social factors are common, varied, and especially relevant to our present moment—in 2014, the U.N. estimated that the share of people living in cities would increase from one-half then to two-thirds by 2050 (United Nations 2014). Additionally, social cohesion in neighborhoods has been shown to have distinct advantages for many issues related to urban design and engineering, including recoveries from infrastructure shocks such as natural disasters (Aldrich and Meyer, 2014).

Ultimately, theories of the relationship between urban form and social cohesion can reasonably be separated into two conceptual frameworks relevant to this study: the new urbanism design paradigm and the critique of urbanization.

The new urbanism design paradigm focuses on the benefits of mixed-use, walkable environments, as opposed to the urban sprawl design phenomenon that emerged in the 1950s with the rise of the personal automobile in America. This theory posits that pedestrian-oriented environments would encourage more interaction on sidewalks as opposed to automobile-oriented environments in which individuals are siloed in their personal vehicles (Leyden, 2011). Even spontaneous, passing interactions, when aggregated over time, could be expected to increase the amount of social cohesion that one identifies with one's neighborhood. These sidewalk interactions have been theorized to foster a web of public respect and trust which forms a resource for the neighborhood (Jacobs, 1961), along the lines of the notions of social cohesion and capital that we have identified above.

On the other hand, a relatively recent body of work has responded to the critical position that the new urbanism design philosophy takes on urban sprawl. These researchers usually identify density of the built environment as a culprit that could explain reductions in social cohesion variables (Koohsari et al., 2021; Nguyen, 2010; Freeman, 2007). One long-standing theory behind this observed phenomenon stems from the notion that a high level of density with many people and activities could overstimulate city dwellers (Simmel, 1903). If it is logical to think that if there are some social benefits to more dynamic neighborhoods that are mixed use and walkable, it is also logical to think that there could be diminishing returns (and even negative effects) if there are an overwhelming number of people in neighborhoods.

A recent review paper (Mazumdar et al., 2018) sought to identify the relationship between social capital and neighborhoods that could be described by new urbanist design characteristics. The researchers reviewed the findings of 23 papers focusing on the built environment and social capital that included relevant case studies. The main takeaway from their work was that while there was some support for the relationship between certain design characteristics and social capital, it was difficult to identify statistically significant findings.

Studies that consider "urban form" in its relationship to social factors typically consider, either explicitly or implicitly, the notion of walkability. Because the concept of walkability is so prevalent in this area of research, we position the concept as central in our study. In this work, we focused our analysis of urban form on walkable urban design, which gives rise to the concept of walkability. We do this for a few key reasons. First, the notion of the connection between walkability and social cohesion has received much theoretical attention in the literature, though evidence remains lacking (Mazumdar et al., 2018). Second, many theories of the relationship between the built environment and social factors involve design features of the built environment that are typically components of aggregated walkability metrics (e.g., land use diversity, density). Lastly, focusing on walkability limits the scope of the study to a point such that the results remain interpretable. Urban form can be described in a large variety of ways and through a variety of lenses, meaning that an analysis of the general concept "urban form" vis-à-vis social cohesion would likely be too broad. Focusing on the different attributes of neighborhood walkability restricts the conceptual space while also providing multiple possibilities for interactions among constituent variables—helping us to potentially parse different sub-components of walkable design.

2.4. Measuring walkability

Different studies and tools have considered the connection between walkable urban form and social cohesion, walkable urban form and other sociological factors, or walkability in and of itself. In these works, different features of the built environment have been used to create different definitions of walkability. Prior to any dedicated walkability metrics scholars such as Talen (Talen, 2005) described walkable neighborhoods as places with high density, diversity of land use, small lots, and connected streets. An early dedicated metric that is common in both research and practice is Walk Score, a metric originally developed by a private company for real estate purposes (Carr et al., 2010). Walk Score is based on the idea that a residence is more walkable if amenities from 13 different categories (e.g., grocery, office) are within a specific walking distance from that residence. In this way, Walk Score is a generalization of many different design characteristics including land use diversity and density. The United States government also publishes a dedicated walkability metric, originally distributed in 2012 and updated more recently in 2021, known as the National Walkability Index (Chapman et al., 2021). Unlike Walk Score, which distills urban form directly into a single metric, the National Walkability Index is functionally an average of three different aspects of design: diversity of land uses, density of physical paths, and connectedness to transit. The ability to use transit is not always included in walkability metrics, but it does indicate the extent to which urban dwellers can access other areas outside of their immediate walking zone while still relying on walking as a key means of transportation.

Based on these walkability metrics, we note that a few key ideas are consistently discussed in the context of walkable urban design. These are diversity of land use, some measure of density (whether describing streets, buildings, or people), as well as occasionally connectedness to transit. These concepts inform our approach to comparing the aspects of walkable urban form to neighborhood-level social cohesion.

2.5. Key gaps

In our review of the literature, we have identified key research gaps that we aim to address in this work. The most common limitation we have observed in previous studies relating walkable urban form to social cohesion is that most studies use relatively small case studies (sample size on the order of 1,000) and consider individual cities in isolation. In contrast, we leveraged open data sources to obtain a sample size on the order of 10,000 across six different cities across the United States, each with varying characteristics. In addition to—and perhaps as a result of—the limited sample sizes we have observed in the literature, there has been conflicting evidence as to the impact of walkable urban form on social cohesion (summarized well in Mazumdar et al. (2018)). We aimed to address this limitation not only by expanding our sample size by leveraging big open geospatially linked data, but also by breaking down the notion of walkability into its constituent components. Two of the more commonly used walkability metrics, Walk Score and the National Walkability Index, blend together multiple ideas such as land use diversity and density, but we note that there are conflicting theories on how these attributed can affect social experiences, as discussed above. Furthermore, there is no physical reason that land use diversity should be correlated to density of street intersections, making it difficult to interpret an overall metric that combines them. By taking a data-driven view of these theorized relationships, we aimed to gain a better understanding of how walkable urban form impacts social cohesion.

3. Data and methodology

In this section, we describe our modeling approach and the data we used to investigate the relationships between walkable urban form and social cohesion, controlling for demographics (overview shown in Fig. 1). Our overarching hypothesis required a methodology for comparing physical design characteristics with social measurements. To make this possible, we identified both physical and social data sources that are geographically specified. We used the Census Block Group (CBG) as the geospatial unit for our analysis. At this geographic scale, we compiled data describing both physical characteristics of the built environment that have been previously related to walkability as well as survey data including responses to questions intended to measure social cohesion. We leveraged Partial Least Squares Structural Equation Modeling (PLS-SEM) to investigate the statistical relationships between the physical environment and social outcomes.

3.1. Data sources

For this modeling approach, we required each of our data sources to be available at the CBG level. The U.S. Census Bureau defines a CBG as a geographic boundary that has a population of 600 to 3,000 people. We used the CBG as our geographic scale for a few key reasons. Both geospatially and socially, it enables aggregation of statistics to what can reasonably be considered a "neighborhood" level, and it is in line with previous research (Freeman, 2007; Andris, 2016). It is also the smallest available geospatial unit for which survey responses and certain urban design characteristics are made public. A larger spatial unit would likely be too large to enable the observation of subtle differences within cities. We relied on two primary data sources:

- The United States Environmental Protection Agency (EPA) publishes a Smart Location Database (SLD) that contains detailed data on physical aspects of the built environment, including their own walkability metric, the specific features used to build that walkability metric, and other related metrics.
- The Baltimore Ecosystem Study (BES) is an ongoing study aimed at ecological understanding of urban areas. One component of the study includes a household telephone survey that includes questions aimed at measuring neighborhood-level social cohesion and is made publicly available.

3.1.1. Smart location database

The SLD is a data product produced by the US EPA that summarizes demographic, employment, and built environment variables for every CBG in the United States. It was originally released in 2012, and the current version of the SLD (version 3.0) was released in 2021 (Chapman et al., 2021). The SLD contains over 100 measures as well as an aggregated National Walkability Index built upon three specific concepts measured within the SLD: employment/housing mix, intersection density, and proximity to transit stops. As discussed above, our aim in this research was to analyze the effect that each component of walkable urban design might have on social cohesion, rather than comparing aggregated walkability metrics with cohesion. As a result, we identified four concepts of walkable urban design than can be measured using statistics reported in the SLD: three metrics similar to those used in the National Walkability Index and an additional social density metric. Thus, our four metrics are land use diversity, physical density, social density, and transit connectedness. A summary of all data collected from the SLD for our study can be found in Table 1.

Land use diversity. The metrics we used for land use diversity were based on entropy calculations built upon the mix of households and/or employment categories. The entropy calculation is as follows:

$$H = \frac{1}{N}\sum_{i=1}^{N}\frac{p_i}{P}ln\left(\frac{p_i}{P}\right)$$

where N is the total number of job or household categories, p_i is the number of entities within category i (e.g., households, jobs of a specific type), and P is the total number of entities across all N categories. We used two diversity entropy values from the SLD: mix of commercial uses and mix of all uses. Mix of commercial uses is calculated in the SLD using eight employment categories (retail, office, industrial, service, entertainment, education, healthcare, and public administration), and mix of all uses is calculated using five employment categories (retail, office,

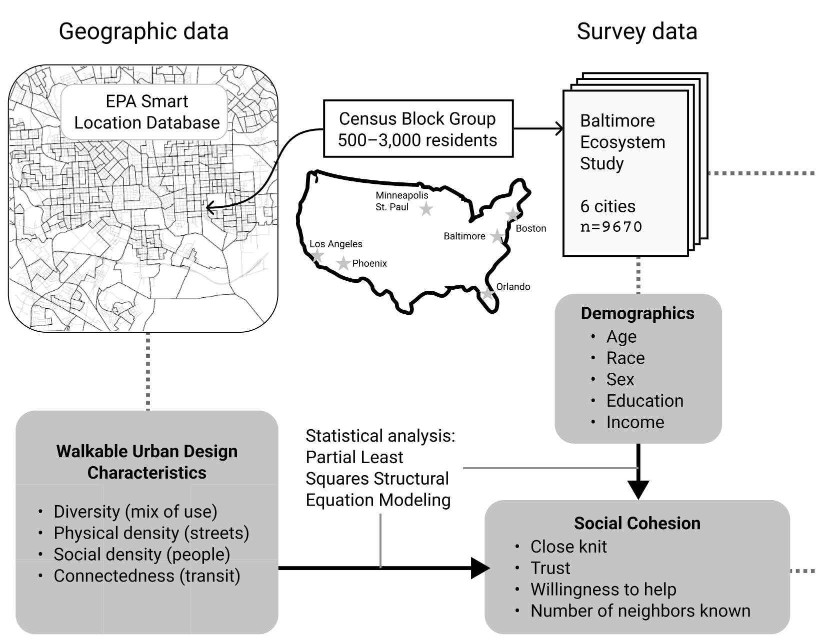

Fig. 1. Overview of methodology.

Table 1
Summary of built environment variables taken from EPA SLD.

Variable name	Minimum	Maximum	Mean	Standard deviation
Land use diversity				
Mix of commercial uses	0.00	0.97	0.48	0.21
Mix of all uses	0.00	1.00	0.56	0.20
Physical density				
Intersection density	0.16	873.43	84.67	75.43
Path density	0.00	48.74	13.47	7.82
Social density				
Population density	0.01	399.91	8.44	10.04
Employment density	0.00	268.60	2.05	6.83
Transit connectedness				
Proximity to transit	0.00	1500.00	993.07	529.05
Transit frequency	0.00	5444.91	46.17	162.52

industrial, service, and entertainment) as well as a category for number of households. These two metrics indicate the level to which different types of land uses exist within a CBG, taking both a detailed look at non-residential uses as well as an overall look at total mix of use. As a result, we used these two entropy metrics as components in an overall diversity metric, as discussed below.

Land use diversity within a relatively fine-grain geospatial area, such as a CBG, is a common component of walkability metrics. This is because it indicates the extent to which different types of activities (e.g., grocery shopping, eating out, going to work) can be carried out within walking distances. Overall, these entropy metrics describe the extent to which the CBG can be described as mixed-use.

Physical density. We used two metrics from the SLD as indicators of physical density: intersection density and path density. While these two metrics are likely to be correlated, they are in fact measuring different things and are both relevant to the concept of physical density of the built environment. Intersection density is a common walkability metric and refers to the number of intersections per square mile, while path density refers to the number of miles of paths per square mile. The SLD includes both metrics for different types of paths: auto-oriented, multi-modal, and pedestrian-oriented. For our analysis, we used the metrics built for multi-model and pedestrian-oriented pathways, excluding auto-oriented pathways.

Social density. We defined the social density metric separately from physical density, in which social density reflects the number of individuals someone within a CBG might expect to be able to interact with as a result of the design of the built environment. We used two metrics to build our social density metric: population density and employment density. Population density is measured in the SLD as people per acre and is a direct result of the number of housing units—and number of people in each housing unit—within a CBG. Employment density is measured as jobs per acre and is a direct result of the number of buildings or spaces that support work within a CBG. These social density indicators can be viewed as both demographic data as well as built environment data, since the design of the neighborhood directly affects the number of residents and workers within a given neighborhood.

When disaggregated spatially to the relatively fine-grain level of the CBG, we argue that these social density variables are indicative of walkable built environment design. The inclusion of the human-oriented variable of social density embeds within our walkability specification the explicit notion of human-built interaction. Walkability depends not just on the bare physical environment, but also the actors within the spaces. For example, physical density and land use diversity can indicate the number of buildings and their mix of use within a neighborhood. But if these buildings are not actually populated either as housing units or workplaces, then the walkability benefits of mixed-use development would be limited. Social density as a design variable can thus be important to walkability.

Transit connectedness. Transit connectedness can be viewed as a walkability indicator because it enables individuals that need to travel across neighborhoods to be able to walk to transit stops within their neighborhoods. We used two SLD metrics as connectedness indicators: proximity to transit and frequency of transit. We measured proximity by taking the negative of the transit distance metric reported in the SLD. Distance is measured as the minimum walk distance (m) between the CBG centroid and the nearest transit stop of any route type. Transit frequency is measured as the frequency of service for each transit route within 0.25 miles during the weekday evening peak period (4:00 pm to 7:00 pm local time), summed for all routes.

A number of CBGs do not have realistic access to transit. These are coded with the values -99999 in the SLD. In an effort to retain these CBGs for analysis, we adjusted these values to the following: 1500m for minimum walking distance to transit (a value greater than the maximum reported value of 1205m and greater than is feasible for frequent transit use), and 0 for transit frequency. These values maintained the reported scales which allowed us to keep non-connected CBGs in the dataset for analysis.

3.1.2. Baltimore ecosystem study

The BES is an ongoing research project that started in 1998 with funding from the US National Science Foundation as a Long-term Ecological Research site. The project grant is administered by the Cary Institute of Ecosystem Studies, and the project is housed at the University of Maryland, Baltimore County. The key goal of the BES is to advance the understanding of urban areas as newer types of ecosystems. Among the core activities of the BES is the BES Household Telephone Survey, which is intended to gather information on environmental knowledge, perceptions, values, and behaviors, as well as how changes in ecosystem structure impact various outcomes including social cohesion. The BES telephone survey covers 6 different cities, despite being originally Baltimore-centric: Baltimore, Boston, Miami, Minneapolis-St. Paul, Phoenix, and Los Angeles. Data and details for the BES telephone survey can be found on the EDI Data Portal.[1] Importantly, each survey response was identified by the respondent's home CBG, which enabled us to fuse the social cohesion survey responses with the SLD, thereby allowing us to attach built environment attributes to each survey response. Through this data fusion process, we ended up with 9,670 total data points containing each attribute of interest, with 26.4% of the valid responses coming from 2006 and 73.6% from 2011.

The BES telephone survey includes three specific questions designed to measure neighborhood social cohesion, each adapted from the seminal Project on Human Development in Chicago Neighborhoods (Sampson et al., 1997) and thereafter used in many studies on neighborhood social cohesion (Bateman et al., 2017; Stein and Griffith, 2015). These were in response to the following phrase: "On a five-point scale, how strongly would you agree or disagree with the following statements about your neighborhood with a score of one being strongly disagree up through five being strongly agree:"

- "This is a close knit neighborhood"
- "People in the neighborhood are willing to help one another"
- "People in this neighborhood can be trusted"

As the number of ties within each individual's social network has also been shown to be an indicator of social cohesion (Kawachi and Berkman, 2015), the BES telephone survey included an additional cohesion-related survey question: "About how many neighbors do you know by name.... 1 (none), 2 (a few), 3 (about half), 4 (most of them), 5 (all of them)." We used the three questions adapted from the Project on Human Development in Chicago Neighborhoods as well as this last question on self-reported and relative tie numbers as our four indicators

[1] https://portal.edirepository.org/nis/mapbrowse?packageid=knb-lter-bes.4000.180

for neighborhood-level social cohesion.

In addition to the social cohesion questions, the telephone survey also included important self-reported demographic information, which we used as control variables for statistical analysis. The demographic data available for each survey response were as follows: education level, race, income, age, and sex. Each of the variables from the BES that we used in this study is summarized in Table 2.

Table 2
Summary of social cohesion and demographic variables from BES household telephone survey (n=9670).

Variable name	Survey question	Response range	Mean	Standard deviation
Social cohesion				
Close knit	How strongly would you agree or disagree with the following statements about your neighborhood: This is a close-knit neighborhood.	1 (strongly disagree), 2, 3, 4, 5 (strongly agree)	3.55	1.26
Trust	How strongly would you agree or disagree with the following statements about your neighborhood: People in this neighborhood can be trusted.	1 (strongly disagree), 2, 3, 4, 5 (strongly agree)	4.02	1.09
Willingness to help	How strongly would you agree or disagree with the following statements about your neighborhood: People in the neighborhood are willing to help one another.	1 (strongly disagree), 2, 3, 4, 5 (strongly agree)	3.92	1.14
Number of neighbors known	About how many neighbors do you know by name?	1 (None), 2 (A few), 3 (About half), 4 (Most of them), 5 (All of them).	2.95	1.01
Demographics				
Education	What is the highest grade of school you have had the opportunity to complete?	1 (less than high school), 2 (high school graduate), 3 (some college), 4 (college graduate), 5 (postgraduate work)	3.51	1.13
Race	Do you consider yourself to be…	1 (White), 0 (Asian, Black, Hispanic, Native American, Other)	0.80	0.40
Income	What is your income?	1 (under $15K), 2 ($15K to $25K), 3 ($25K to $35K), 4 ($35K to $50K), 5 ($50K to $75K), 6 ($75K to $100K), 7 ($100K to $150K), 8 (over $150K)	4.28	1.64
Age	Please stop me when I reach the category that includes your age.	1 (under 35), 2 (35 to 44), 3 (45 to 54), 4 (55 to 64), 5 (65 or over)	3.21	1.27
Sex	(Respondents chose the response they most identified with.)	0 (Female), 1 (Male)	0.59	0.49

3.2. Modeling approach: structural equation modeling with partial least squares

We required a statistical tool that enabled us to relate our measured variables to the underlying concepts they represent and thereafter analyze the statistical relationships among these concepts. We used Partial Least Squares Structural Equation Modeling (PLS-SEM), a framework for statistical modeling that combines factor analysis with path analysis. In factor analysis, one relates an underlying concept, or latent variable (e.g., social cohesion), to a number of indicators that are theorized to represent aspects of the latent variable (e.g., responses to questions on a survey); this is known as the measurement model. In this research, many of the concepts we are interested in modeling can be readily thought of as latent variables expressed through indicators. Given a set of latent variables, path analysis enables multivariate regression in which some latent variables are affected by others (e.g., social cohesion affected by density of the built environment); this is known as the structural model.

The standard SEM approach is also known as covariance-based SEM, which aims to minimize the distance between the model's covariance matrix and the observed covariance matrix using maximum likelihood estimation. An alternative approach, PLS-SEM (also known as PLS Path Modeling), uses ordinary least squares regression to maximize the explained variance of the target endogenous latent variables. Covariance-based SEM involves strict data assumptions, including that the measured data follow normal distributions, which is a difficult assumption to meet with non-experimental data. Additionally, covariance-based SEM requires that the measurement model be in the reflective mode. Scholars have noted that PLS-SEM can be preferred to CB-SEM when the following conditions are met (Hair et al., 2019; Mehmetoglu and Venturini, 2021):

- Distributions are nonnormal (while this is not a reason to choose PLS-SEM in and of itself, PLS-SEM has been shown to perform well when assumptions about data distributions are not met)
- The research is exploring theory development, rather than confirming an existing theory
- One or more of the latent variables is formatively measured
- When the research is based on secondary data rather than a controlled experiment

As each of these points applies to our modeling approach, we elected to use PLS-SEM in this research. An additional strength of PLS-SEM is that it can create latent variable scores, which enables follow up analysis. This is particularly helpful when exploring the implications of new theory development (Hair et al., 2019). One disadvantage of the PLS-SEM approach for statistical analysis is that since it is nonparametric, significant testing requires bootstrapping to create confidence intervals for certain model parameters, such as weights in the measurement model and coefficients in the structural model.

3.3. Model definition

The model definition for our PLS-SEM is shown graphically in Fig. 2. In the subsections below, we define explicitly our measurement model and structural model.

3.3.1. Measurement model

Based on our review of the literature and the urban form data available at the CBG, we defined four urban form latent concepts to be our primary independent variables: land use diversity, physical density, social density, and transit connectedness. These four concepts are typically included in overall walkability metrics, such as the National Walkability Index and Walk Score. For these four concepts, we used formative measurement with indicators taken from the EPA SLD. We chose to use formative measurement because, for each exogenous latent

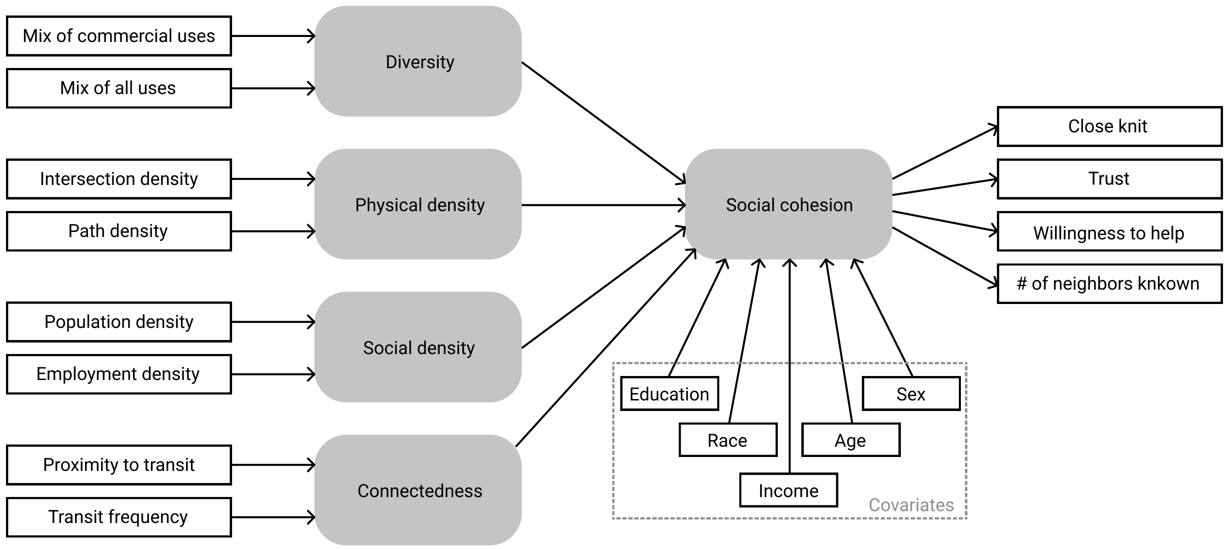

Fig. 2. PLS-SEM model definition. The exogenous latent variables in the structural model are measured formatively (arrow from indicator to latent variable), whereas social cohesion is measured reflexively (arrow from latent variable to indicator).

variable, the measured indicators are more appropriately thought of as combining to *form* the latent concept, rather than reflecting subtle differences in the latent concept. For example, social density in our model is formed by population density and employment density. While it may be the case that these two indicators are correlated with one another, they are in fact separate concepts measuring different things. But, when taken together, they create an overall picture of the number of people expected to utilize a particular neighborhood, which we call social density in our analysis. The measurement model for the formatively measured exogenous variables is reflected in the overall proposed PLS-SEM model (Fig. 2).

For our endogenous variable, social cohesion, we used reflective measurement with four indicators from the BES. Here, reflective measurement is more appropriate because the survey questions are all designed to reflect an overarching concept of social cohesion. These four indicators—close knit, trust in neighbors, willingness to help, and number of neighbors known—are responses to the survey questions described above.

3.3.2. *Structural model*

Our structural model simply states that social cohesion is affected by each of the four features of walkable urban form as well as demographic covariates, as shown in Fig. 2. Because the demographic covariates do not define a latent variable, either formatively or reflectively, we created a dummy latent variable for each covariate (e.g., “Latent Education” formed only by “Education”) and included a path directly from each dummy demographic latent variable to social cohesion. We omitted this technical detail from Fig. 2. To estimate the model, we used the “SEMinR” package in the statistical programming language “R” (Ray et al., 2021). Our code for analysis is publicly available on Github.[2]

4. Results and discussion

Evaluation of the PLS-SEM starts with validity of the measurement model, to ensure that the latent variables are indeed represented by the chosen indicators. Once the measurement model is deemed valid, the structural model can be analyzed for statistical relationships.

[2] https://github.com/asonta/rethinking-walkability

4.1. *Measurement model validity*

Table 3 shows the metrics used to assess validity of the measurement model. The metrics differ depending on whether the latent variable is defined reflectively or formatively, as informed by established guidelines on PLS-SEM use (Hair et al., 2019; Mehmetoglu and Venturini, 2021). For the four independent variables in the measurement model, which were measured formatively, we assessed measurement model validity by investigating collinearity of the formative indicators as well as statistical significance of the indicator weights. We used the Variance Inflation Factor (VIF) to evaluate collinearity of indicators, as is common in PLS-SEM. A VIF value above 5 indicates high collinearity among indicators, and guidelines state that ideally VIF values should be close to 3 or below (Hair et al., 2019; Becker et al., 2015). This condition was met for diversity, social density, and connectedness, while physical density exhibited a VIF of 3.73. Because this value is below the recommended maximum threshold of 5, we elected to accept these modeling results for

Table 3
Measurement model validity.

Latent variable Indicator variable	Weight	95% weight confidence interval	VIF
Independent variables			
(Built environment attributes)			
Diversity			1.19
Mix of commercial uses	0.948*	(0.811, 1.04)	
Mix of all uses	0.115	(-0.117, 0.345)	
Physical density			3.73
Intersection density	0.853*	(0.616, 1.07)	
Path density	0.168	(-0.085, 0.420)	
Social density			1.03
Population density	0.966*	(0.927, 0.997)	
Employment density	0.145*	(0.014, 0.257)	
Connectedness			1.16
Proximity to transit	0.678*	(0.550, 0.785)	
Transit frequency	0.526*	(0.400, 0.648)	
Latent variable Indicator variable	**Loading**	**Cronbach’s alpha**	**AVE**
Dependent variable			
Social cohesion		0.789	0.482
Close knit	0.399		
Trust	0.990		
Willingness to help	0.670		
Number of neighbors known	0.583		

further analysis. Because PLS-SEM is a nonparametric method, statistical significance in general can be estimated through bootstrapping, as discussed above. After bootstrapping the model with 1000 iterations, we were able to create a 95% confidence interval for the indicator weights used in the formative measurement model. We found that some of these confidence intervals did include 0, as indicated in the table. However, we note that a non-significant weight does not necessarily indicate that the indicator should be removed (Hair et al., 2019). Instead, for indicators with non-significant weights, the indicator's outer loading should be considered (Cenfetelli and Bassellier, 2009). For the three indicators with nonsignificant weights, each had an outer loading higher than 0.5, as recommended (Mix of all uses: 0.50, Path density: 0.90). For these reasons, we accepted the validity of the formative components of the measurement model.

To assess the validity of the reflectively measured dependent latent variable, social cohesion, we examined indicator loadings, consistency reliability using Cronbach's alpha, and convergent reliability using the average variance extracted (AVE) metric. The standard criterion for indicator loadings is that each loading should be at 0.7 or above, which indicates that more than 50% of the indicator's variance is captured by the latent construct. We found that only one of the indicators (trust) met this threshold. However, we note that indicator loadings between 0.4 and 0.7 are often accepted, particularly in early-stage theory development, which we believe applies to this study (Hair et al., 2016). Furthermore, we note that the close knit, trust, and willingness to help variables are established indicators for social cohesion in the social science literature, which we believe provides convincing rationale for retaining each indicator in the model. For consistency reliability, Cronbach's alpha should be greater than 0.7, which we found to be the case. And finally, the AVE metric should be 0.5 or above. While 0.482 is below the 0.5 threshold, it is quite close, and we deemed this result acceptable for continued analysis of the structural part of the model.

4.2. Structural model path analysis

Key results from the structural model path analysis are shown in Table 4. As a first step in assessing the structural model, it is important to ensure that there exists no collinearity among independent variables. Using the VIF metric, we found that no critical collinearity issues exist in the structural model (VIF below 5).

The most important results from the structural model are the path coefficients, which allow us to examine the relationships between the independent and dependent variables, and the 95% confidence internal, which allows us to determine the significant relationships between urban form and cohesion. We found that eight of the nine independent latent variables had path coefficients significantly different from zero at the 95% confidence level. Each of the demographic covariates except education level had a significant impact on reported social cohesion. On average, as both income and age increased, respondents reported higher cohesion. We note that it is generally accepted that social isolation poses a major health risk for older adults (Nicholson, 2012; Cornwell et al., 2008), though some recent work has found that older age is associated with divergent factors for different forms of social connectedness (Cornwell et al., 2008). While we found an average positive relationship between age and cohesion, which was an important effect to control for in the research, we note that this was an average effect, and this work did not focus on the demographic of older adults. For the other covariates, white respondents and female respondents reported higher cohesion than nonwhite and male respondents.

Table 4
Structural model results.

Latent variable	VIF	Cohesion path coefficient	95% confidence interval
Diversity	1.05	0.051*	(0.029, 0.072)
Physical Density	2.01	-0.050*	(-0.079, -0.019)
Social Density	2.02	-0.085*	(-0.125, -0.055)
Connectedness	2.33	-0.038*	(-0.071, -0.001)
Education	1.22	0.011	(-0.013, 0.032)
Race	1.14	0.066*	(0.043, 0.088)
Income	1.35	0.158*	(0.134, 0.182)
Age	1.07	0.166*	(0.145, 0.187)
Sex	1.01	0.041*	(0.020, 0.063)

* indicates that 0 is not contained within the 95% confidence interval $R^2 = 0.101$

The urban design features we included as independent variables were each found to be statistically significant, but they differ in the effect they had on social cohesion. Land use diversity was shown to have a positive impact on social cohesion. On the other hand, both physical density and social density were shown to have significant negative effects on social cohesion. The effect of connectedness to transit was also negative, with greater access to transit being associated with less social cohesion. While these coefficients are not large, it is important to note that they would likely not be expected to have high values. Social cohesion is a complex phenomenon influenced by many factors, one of which may be the design of the built environment. By showing that these relationships are statistically significant, we find evidence that these influences do in fact exist, which is an important finding irrespective of the raw coefficient values.

A particularly striking finding is the opposite impacts of diversity and density, which simultaneously lends support to two arguments that are often seen as competing:

- The positive impact of land use diversity lends partial support to the theories associated with the new urbanism design philosophy. As we move from purely residential communities (places where you often need a car for daily activities) toward more mixed-use, walkable places, it can be expected that more socially meaningful contact becomes possible. This effect could arise from a few different mechanisms. One is that higher diversity makes walking a more viable transportation option, because more destinations/amenities would be within walking distance. As more people utilize sidewalks through walking as opposed to streets through driving, there become more opportunities for interaction among city users, which in turn would be expected to increase cohesion. Another possible mechanism is that increased land use diversity simply creates more vibrancy within neighborhoods (more shops, more event spaces, etc.), and the uses of neighborhood parcels themselves create the opportunities for meaningful social contact.
- At the same time, however, the negative impact of physical and social diversity also lends partial support to the theories that have critiqued pure urbanization. As neighborhoods become denser, we may be seeing the hypothesized effect of increased anonymity. Highly dense neighborhoods, in terms of both the number of buildings and the number of people within those buildings, could overwhelm our social experiences. If, for example, one is walking down a street, one might expect different outcomes depending on the number of other individuals using that space. If no one else is on the street, interaction is not possible. If a very large number of individuals are also using that space, interaction is possible, but meaningful interaction may not be feasible. If some moderate number of other individuals is using that sidewalk, it may be easier to engage in socially meaningful interaction—one may start to recognize the same person over and over, and eventually engage in conversation, for example. There may be socially diminishing returns from increased density, an idea that our findings would support.

The theoretical behaviors embedded in these theories operate on two different key aspects of design—density and diversity. This analysis sheds empirical light on this density/diversity dichotomy, a light that helps us to understand the complexities of our urbanism theories, and one that invites further analysis of the effects of diversity and density.

4.3. Investigating diversity and density

The opposite effects of diversity and density on cohesion invited further investigation of the variable relationships. In Fig. 3, we show the data relationships between diversity, physical density, and social cohesion (we omit social density here for readability), along with a locally estimated scatterplot smoothing (LOESS) regression line to aid with visualization. The latent variables were calculated using the weight factors of their constituent indicators. We also applied a log transform to the physical density variable for visualization purposes. We found that the negative relationship between physical density and cohesion is much stronger for high-density neighborhoods than it is for lower-density neighborhoods (as indicated by the concave shape of the curve). This would suggest that our highest-density neighborhoods are driving the overall negative relationship between density and cohesion. We also see a surprising relationship between diversity and density. For lower-density neighborhoods (on the left-hand side of the plot), we see that increasing density is associated with increasing diversity. In other words, as we start to increase physical density, we generally increase the amount of diversity of buildings. However, for high-density neighborhoods, we see that increasing density is actually associated with a decrease in diversity. This would suggest that in areas with many buildings already, as we increase the density of the urban fabric, we generally reinforce the existing mix of land use (e.g., more office buildings in areas with many office buildings in place). One important takeaway from this analysis is that we have a large opportunity to improve the social experience of the densest parts of our cities.

4.4. Exploring diversity and density interaction effects

Based on the apparent relationship between physical density and diversity, we also conducted a post-hoc analysis using a PLS-SEM model that included an interaction term between physical density and diversity but was otherwise identical to our original analysis. Our hypothesis for this additional analysis was that any effects that density might have on cohesion would be mediated by diversity. For example, for a high-density region, we might expect different effects on cohesion, depending on whether the region is low or high diversity—an expectation that stems from our exploratory analysis above. The model specification and model results can be seen in Fig. 4.

We omit in this paper many of the details of measurement model validity for this alternate model, but we note that none of the metrics used to assess model validity above changed significantly with the addition of the interaction term. It is important to note that the addition of the interaction term also does not introduce multicollinearity into the structural model, with the VIF of the interaction term being 1.053. With the inclusion of the interaction term, the model fit as measured through variance explained increased by a small amount (R^2 increase from 0.101 to 0.102), which indicates that the interaction term slightly improves the model fit. We also found that the interaction term (diversity × physical density) was significant at 0.024 (95% confidence interval: 0.004, 0.049), which supports our hypothesis that diversity mediates the effect that physical density has on social cohesion. In other words, as physical density increases, we see different effects on social cohesion depending on how the diversity of those neighborhoods changes alongside the increase in density. Neighborhoods with high diversity and high density would have higher social cohesion than neighborhoods with low diversity and high density. This finding lends support to the argument that the densest parts of our cities—which would likely include the central business districts of the cities analyzed—could be improved by making them more mixed use. Furthermore, this finding implies that in the densest parts of our cities, if we can implement design with more land use diversity, we may not necessarily be limiting social cohesion. While the interaction analysis demonstrates the opportunities that exist regarding highly dense areas, we would also like to emphasize the findings that relate to the more common areas of lower density. Here, the negative effects of density seem to be less pronounced (Fig. 3b), and when considering the direction and significance of the interaction term, we argue that the association between density and cohesion should continue to be investigated. One the other hand, the positive association between mix of uses and cohesion seems to persist across all neighborhood typologies, suggesting that adjusting mix of uses appears to be a design and policy lever with wide applicability.

We note that we also explored the possibility of interaction between social density and diversity, but we found that the inclusion of this interaction did not improve overall model fit, and the interaction term coefficient was insignificant at the 95% confidence level.

4.5. Multigroup effects for different cities

Our unique dataset includes data for six different cities, which offers the opportunity to further investigate whether different relationships appear in different cities. Therefore, we also conducted a post-hoc analysis to understand if different cities have different relationships when compared to the overall trends for the other cities in the dataset. To do this, we used multi-group analysis in PLS-SEM (Cheah et al., 2023). We ran six tests, where in each test we separated the data into two groups: one group containing data from only one of the cities, and the other group containing the data from all other five cities. We repeated this, isolating each of the cities in turn. In each test, the group with the individual cities still contained a large amount of data (Baltimore: 3779, Boston: 1227, Los Angeles: 1079, Orlando: 1033, Phoenix: 1254, Twin Cities: 1298). We used this procedure to attempt to identify the differences between average effects and individual cities' effects. In Fig. 5, we report how the coefficients for the structural model change when each city is placed in a group separate from the rest of the dataset.

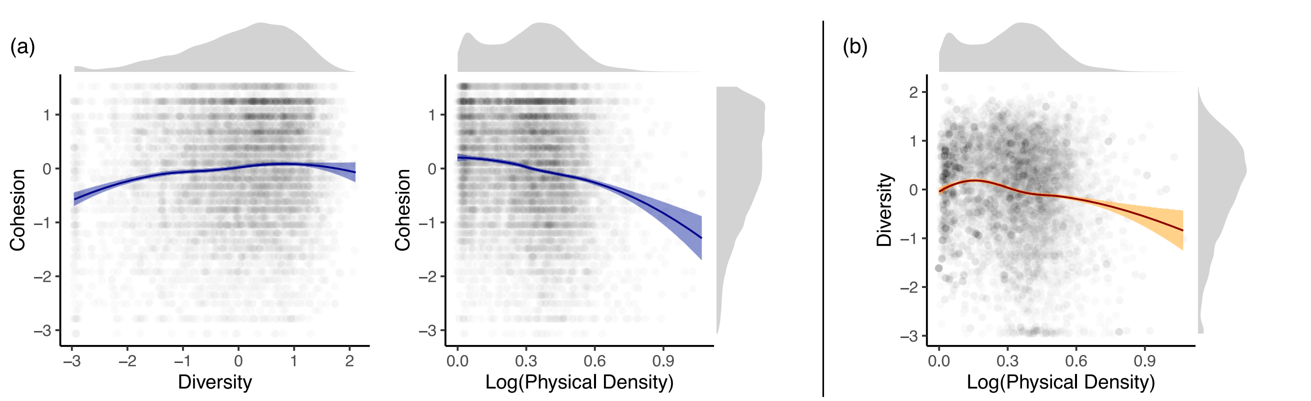

Fig. 3. Data relationships between (a.) cohesion and both land use diversity and physical density, and (b.) land use diversity and physical density. Relationships are plotted as pairwise scatterplots with LOESS regressions (span of 0.80 and confidence intervals of 0.95).

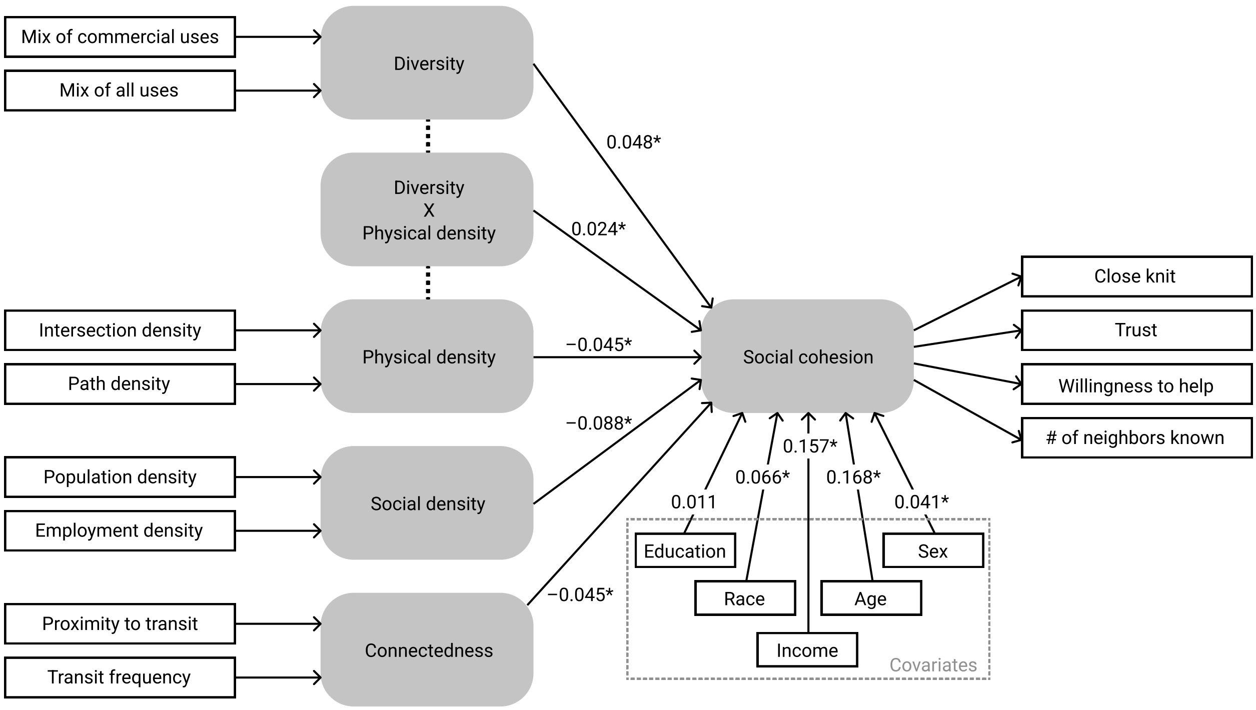

Fig. 4. PLS-SEM model with diversity and physical density interaction term with structural model path analysis results. A (*) beside the regression coefficient indicates that zero is not within the 95% confidence interval after bootstrapping. $R^2 = 0.102$.

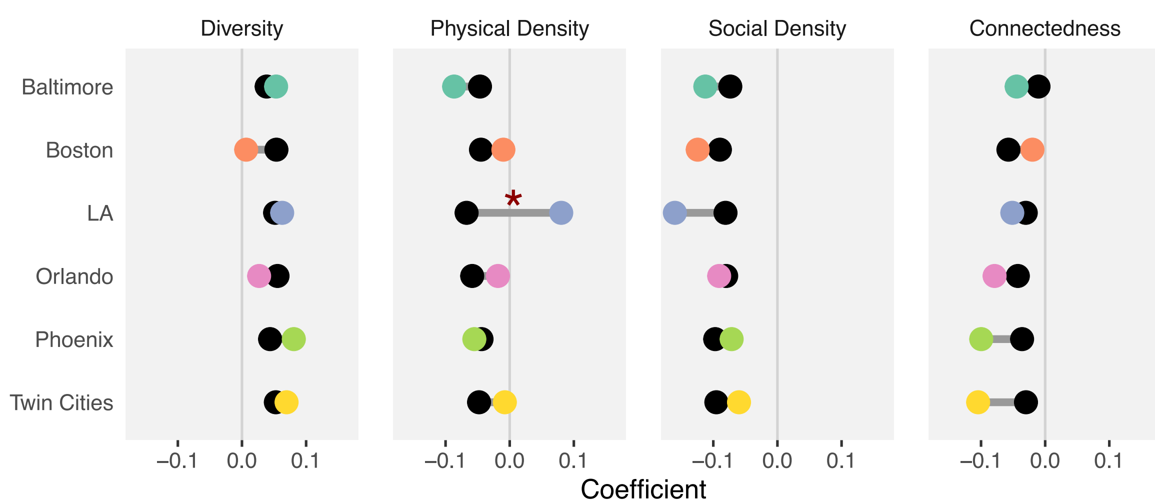

Fig. 5. Results of PLS-SEM multi-group analysis. Each pairwise plot in the matrix shows the differences between the PLS-SEM structural model coefficients (by column) for different cities (by row). The colored dot is for the individual city, and the black dot is for the other five cities combined.

Here again, each model was bootstrapped, enabling estimation of 95% confidence intervals and estimations of the statistical significance of differences between groups. We found through bootstrapping that the only significant difference occurred between groups when analyzing Los Angeles separately, and only for physical density. Here, the coefficient for the Los Angeles data was 0.080, while it was -0.068 for the rest of the data. We found none of the other group differences to be significant, and importantly, the signs for the other differences did not change as well. We believe that this finding demonstrates that our data has a reasonable amount of consistency across cities and that the findings seem generally applicable across the different cities. At the same time, the finding related to physical density in Los Angeles raises questions about the effect of physical density on cohesion both within Los Angeles and in the other cities. The specific reasons for this difference related to physical density is outside the scope of this study, but we note that if we only had the data from Los Angeles in our study, we may have drawn different conclusions. This could be one reason that a previous systematic review of studies that investigated social cohesion and urban form found inconsistencies in the literature (Mazumdar et al., 2018), as discussed above. While acknowledging the limitations of our work and the need for future studies, we note that our findings offer new evidence at a large scale and demonstrate the benefits of using large, open datasets for such analysis.

4.6. Limitations and future work

The work presented here demonstrates that the notion of walkability is best decomposed into its constituent parts when its effects on social outcomes are considered. While the findings are significant and have important implications for planning, engineering, and policy decisions, we do note that there are certain limitations to this work that should be clearly stated.

One inherent limitation is that only six metropolitan areas in the United States were considered. While these cities vary by geographic region, climate, and many other factors, it is important to note that any analysis of urban form would benefit from the inclusion of data from more cities around the globe. Furthermore, while the six cities and roughly 10,000 data points means that this study, to the best of our knowledge, is the largest analysis of urban form vis-à-vis social cohesion, any study would benefit from the inclusion of more data. For this reason, we believe that this broad area of research would benefit from the ability to extract social cohesion data from data sources associated with the big-data revolution, such as GPS traces from smartphones, *in-situ* sensors, and other sources.

Furthermore, our PLS-SEM model exhibited weak explanatory power ($R^2 = 0.102$). We note that while this value is low, caution should be taken when interpreting R^2 values in social research, particularly when the analysis is focused on the effects of variables rather than predictive power. Researchers have noted that small values of this magnitude are acceptable in such contexts, particularly when modeling effects emergent from human behavior (Hair et al., 2019; Abelson, 1985; Lewis-Beck and Skalaban, 1990). One reference on PLS-SEM modeling states that high R^2 values "in a model that predicts human attitudes, perceptions, and intentions likely indicate an overfit" (Hair et al., 2019). While we argue this is acceptable in exploratory-stage research, it does indicate that much of the variance in social cohesion cannot be described by the features we captured—built environment characteristics and demographic covariates. This limitation is to be expected, as social cohesion is a complex social outcome, but we note that it does mean there is significant room for model development as well as new theories that could better explain the pathways linking the experience of the built environment to social outcomes.

5. Conclusion

In this research, we leveraged open data sources on both neighborhood-level social cohesion and walkable urban design characteristics to explore the relationship between the provision of urban infrastructure via urban form and the social outcome of cohesion. Through a statistical analysis, we found that different aspects of walkable urban design have opposite effects on cohesion, controlling for demographics. We found land use diversity to be positively associated with cohesion, lending support to the theories of new urbanism that emphasize mixed-use development. At the same time, however, we found that physical density, social density, and transit connectedness were each negatively associated with cohesion, providing evidence that density inhibits cohesion. Through additional analysis, we found that the highly dense parts of the cities analyzed were driving negative associations with cohesion. This particular finding demonstrates that reevaluating our infrastructure and urban form in highly dense areas has the potential to improve the social outcomes associated with cities. We also found that the effect of density is mediated by diversity, indicating that if our dense neighborhoods are also diverse, the negative effects are significantly reduced. Ultimately, the findings from this analysis demonstrate the value in rethinking the way we discuss "walkability" in the context of social cohesion. Because walkable urban form and mobility infrastructure is comprised of a complex set of attributes that do not necessarily align, we benefit when taking a nuanced look at the individual attributes. When considering urban form and the provision of infrastructure—including active mobility infrastructure, urban amenities, and land uses—findings from this research suggest the importance of the mix of uses when it comes to the social experience of cohesion, particularly in dense areas. These findings can aid the realms of urban planning, engineering, and policy when it comes to striving toward more cohesive and resilient communities.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data is publicly available and the code for analysis is publicly available at: https://github.com/asonta/rethinking-walkability.

Acknowledgement

This work has been supported by the Data Science Institute at Columbia University, by EPFL, and by the National Science Foundation under Grant Number CNS-1943396. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Columbia University, NSF, or the U.S. Government or any of its agencies.

References

Jardim, B., & de Castro Neto, M. (2022). Walkability indicators in the aftermath of the COVID-19 pandemic: A systematic review. *Sustainability, 14*(17). https://doi.org/10.3390/su141710933. Art. no. 17Jan.

Loo, B. P. Y. (2021). Walking towards a happy city. *Journal of Transport Geography, 93*, Article 103078. https://doi.org/10.1016/j.jtrangeo.2021.103078. May.

Moreno, C., Allam, Z., Chabaud, D., Gall, C., & Pratlong, F. (2021). Introducing the '15-minute city': Sustainability, resilience and place identity in future post-pandemic cities. *Smart Cities, 4*(1). https://doi.org/10.3390/smartcities4010006. Art. no. 1Mar.

Sonta, A. J., & Jain, R. K. (2020). Optimizing neighborhood-scale walkability. In *Proceedings of the computing in civil engineering* (pp. 454–461). https://doi.org/10.1061/9780784482438.058. May.

Gao, W., Qian, Y., Chen, H., Zhong, Z., Zhou, M., & Aminpour, F. (2022). Assessment of sidewalk walkability: Integrating objective and subjective measures of identical context-based sidewalk features. *Sustainable Cities and Society, 87*, Article 104142. https://doi.org/10.1016/j.scs.2022.104142. Dec.

Carr, L. J., Dunsiger, S. I., & Marcus, B. H. (2010). Walk score™ as a global estimate of neighborhood walkability. *American Journal of Preventive Medicine, 39*(5), 460–463.

Liao, B., van den Berg, P. E. W., van Wesemael, P. J. V., & Arentze, T. A. (2020). Empirical analysis of walkability using data from the Netherlands. *Transportation Research Part D: Transport and Environment, 85*, Article 102390. https://doi.org/10.1016/J.TRD.2020.102390. Aug.

W. Huang and E.B. Khalil, "Walkability optimization: Formulations, algorithms, and a case study of Toronto." arXiv, Dec. 09, 2022. doi: 10.48550/arXiv.2212.05192.

Loo, B. P., Mahendran, R., Katagiri, K., & Lam, W. W. (2017). Walking, neighbourhood environment and quality of life among older people. *Current Opinion in Environmental Sustainability, 25*, 8–13. https://doi.org/10.1016/j.cosust.2017.02.005. Apr.

Jun, H. J., & Hur, M. (2015). The relationship between walkability and neighborhood social environment: The importance of physical and perceived walkability. *Applied Geography, 62*, 115–124. https://doi.org/10.1016/J.APGEOG.2015.04.014. Aug.

Koohsari, M. J., et al. (2021). Traditional and novel walkable built environment metrics and social capital. *Landscape and Urban Planning, 214*, Article 104184. https://doi.org/10.1016/J.LANDURBPLAN.2021.104184. Oct.

K.M. Leyden, "Social capital and the built environment: The importance of walkable neighborhoods," 10.2105/AJPH.93.9.1546, vol. 93, no. 9, pp. 1546–1551, Oct. 2011, doi: 10.2105/AJPH.93.9.1546.

Rogers, S., Gardner, K., & Carlson, C. (2013). Social capital and walkability as social aspects of sustainability. *Sustainability, 5*(8), 3473–3483. https://doi.org/10.3390/su5083473. Aug.

Lee, J. H., & Tan, T. H. (2019). Neighborhood walkability or third places? Determinants of social support and loneliness among older adults. *Journal of Planning Education and Research*. https://doi.org/10.1177/0739456X19870295, 0739456X1987029Aug.

Mazumdar, S., Learnihan, V., Cochrane, T., & Davey, R. (2018). The built environment and social capital: A systematic review. *Environment and Behavior, 50*(2), 119–158. https://doi.org/10.1177/0013916516687343. Feb.

Coleman, J. S. (1994). *Foundations of social theory*. Harvard University Press.

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science, 241*(4865), 540–545. https://doi.org/10.1126/SCIENCE.3399889

Kawachi, I., & Berkman, L. F. (2001). Social ties and mental health. *Journal of Urban Health, 78*(3), 458–467. https://doi.org/10.1093/JURBAN/78.3.458, 2001 78:3.

D.P. Aldrich and M.A. Meyer, "Social capital and community resilience:," https://doi.org.ezproxy.cul.columbia.edu/10.1177/0002764214550299, vol. 59, no. 2, pp. 254–269, Oct. 2014, doi: 10.1177/0002764214550299.

Kawachi, I., & Berkman, L. F. (2015). Social capital, social cohesion, and health,". *Social epidemiology* (pp. 290–319). Oxford University Press. https://doi.org/10.1093/MED/9780195377903.003.0008

Delhey, J., & Dragolov, G. (2016). Happier together. Social cohesion and subjective well-being in Europe. *International Journal of Psychology, 51*(3), 163–176. https://doi.org/10.1002/IJOP.12149. Jun.

C. Ellis, "The new urbanism: Critiques and rebuttals," https://doi.org/10.1080/1357480022000039330, vol. 7, no. 3, pp. 261–291, Oct. 2010, doi: 10.1080/1357480022000039330.

Allam, Z., Bibri, S. E., Chabaud, D., & Moreno, C. (2022). The theoretical, practical, and technological foundations of the 15-minute city model: Proximity and its environmental, social and economic benefits for sustainability. *Energies, 15*(16). https://doi.org/10.3390/en15166042. Art. no. 16Jan.

Lund, H. (2003). Testing the claims of new urbanism. *Journal of the American Planning Association*.

D. Nguyen, "Evidence of the impacts of urban sprawl on social capital:," https://doi-org.ezproxy.cul.columbia.edu/10.1068/b35120, vol. 37, no. 4, pp. 610–627, Aug. 2010, doi: 10.1068/b35120.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*(5328), 918–924. https://doi.org/10.1126/SCIENCE.277.5328.918. Aug.

Miao, C., et al. (2020). How the morphology of urban street canyons affects suspended particulate matter concentration at the pedestrian level: An *in-situ* investigation. *Sustainable Cities and Society, 55*, Article 102042. https://doi.org/10.1016/j.scs.2020.102042. Apr.

Zhu, Z., Gou, L., Liu, S., & Peng, D. (2023). Effect of urban neighbourhood layout on the flood intrusion rate of residential buildings and associated risk for pedestrians. *Sustainable Cities and Society, 92*, Article 104485. https://doi.org/10.1016/j.scs.2023.104485. May.

Hua, J., et al. (2022). Investigating pedestrian-level greenery in urban forms in a high-density city for urban planning. *Sustainable Cities and Society, 80*, Article 103755. https://doi.org/10.1016/j.scs.2022.103755. May.

Yu, Y., & de Dear, R. (2022). Thermal respite for pedestrians in overheated urban environments – introduction of a dynamic analysis of outdoor thermal comfort. *Sustainable Cities and Society, 86*, Article 104149. https://doi.org/10.1016/j.scs.2022.104149. Nov.

Yang, C., Shi, S., & Runeson, G. (2023). Towards sustainable urban communities: Investigating the associations between community parks and place attachment in master-planned estates in Sydney. *Sustainable Cities and Society, 96*, Article 104659. https://doi.org/10.1016/j.scs.2023.104659. Sep.

Moore, S., Salsberg, J., & Leroux, J. (2013). Advancing social capital interventions from a network and population health perspective. *Global Perspectives on Social Capital and Health*, 189–203. https://doi.org/10.1007/978-1-4614-7464-7_8. Jan.

Carpiano, R. M. (2008). Actual or potential neighborhood resources for health. *Social Capital and Health*, 83–93. https://doi.org/10.1007/978-0-387-71311-3_5

United Nations, "World urbanization prospects 2014," 2014. doi: (ST/ESA/SER.A/366).

J. Jacobs, The death and life of great American cities, vol. 71. 1961. doi: 10.2307/794509.

L. Freeman, "The effects of sprawl on neighborhood social ties: An explanatory analysis," https://doi.org/10.1080/01944360108976356, vol. 67, no. 1, pp. 69–77, 2007, doi: 10.1080/01944360108976356.

G. Simmel, "The metropolis and mental life," The metropolis and mental life, 1903.

Talen, E. (2005). Evaluating good urban form in an inner-city neighborhood: An empirical application. *Journal of Architectural and Planning Research*, 204–228.

J. Chapman, E. Fox, W. Bachman, L. Frank, J. Thomas, and A.R. Reyes, "Smart location database technical documentation and user guide (Version 3.0)," United States environmental protection agency, Jun. 2021.

Andris, C. (2016). Integrating social network data into GISystems. *International Journal of Geographical Information Science*, 1–23. https://doi.org/10.1080/13658816.2016.1153103. Mar.

Bateman, L. B., et al. (2017). Examining neighborhood social cohesion in the context of community-based participatory research: Descriptive findings from an academic-community partnership. *Ethnicity & Disease, 27*, 329. https://doi.org/10.18865/ED.27.S1.329. Suppl 1Nov.

R.E. Stein and C. Griffith, "Resident and police perceptions of the neighborhood: Implications for community policing," https://doi.org/10.1177/0887403415570630, vol. 28, no. 2, pp. 139–154, Feb. 2015, doi: 10.1177/0887403415570630.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203. Jan.

M. Mehmetoglu and S. Venturini, "Structural equation modelling with partial least squares Using Stata and R," Structural Equation Modelling with Partial Least Squares Using Stata and R, Feb. 2021, doi: 10.1201/9780429170362/STRUCTURAL-EQUATION-MODELLING-PARTIAL-LEAST-SQUARES-USING-STATA-MEHMET-MEHMETOGLU-SERGIO-VENTURINI.

Ray, S., Danks, N., & Valdez, A. C (2021). SEMinR: Domain-specific language for building, estimating, and visualizing structural equation models in R. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3900621. Aug.

Becker, J. M., Ringle, C. M., Sarstedt, M., & Völckner, F. (2015). How collinearity affects mixture regression results. *Marketing Letters, 26*(4), 643–659. https://doi.org/10.1007/S11002-014-9299-9. Dec.

Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly: Management Information Systems, 33*(4), 689–707. https://doi.org/10.2307/20650323

Hair, J., Hult, G. T., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. SAGE Publications, Inc.

Nicholson, N. R. (2012). A review of social isolation: An important but underassessed condition in older adults. *Journal of Primary Prevention, 33*(2), 137–152. https://doi.org/10.1007/s10935-012-0271-2. Jun.

Cornwell, B., Laumann, E. O., & Schumm, L. P. (2008). The social connectedness of older adults: A national profile*. *American Sociological Review, 73*(2), 185–203. https://doi.org/10.1177/000312240807300201

Cheah, J. H., Amaro, S., & Roldán, J. L. (2023). Multigroup analysis of more than two groups in PLS-SEM: A review, illustration, and recommendations. *Journal of Business Research, 156*, Article 113539. https://doi.org/10.1016/j.jbusres.2022.113539. Feb.

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin, 97*(1), 129–133. https://doi.org/10.1037/0033-2909.97.1.129

Lewis-Beck, M. S., & Skalaban, A. (1990). The R-Squared: Some straight talk. *Political Analysis, 2*, 153–171. https://doi.org/10.1093/pan/2.1.153. Jan.