Chapter

Demographic Data in the Built Environment and Human Health Studies

Huaqing Wang and Louis G. Tassinary

Abstract

Demographic data is widely used in both built environment and population health studies. Traditional data sources include national, state, and local surveys as well as archived data from longitudinal studies and newly emerging sources such as digitally accessible administrative data and real-time data from mobile devices. The value of these diverse data sets hinges on their accuracy, completeness, reliability, relevance, and timeliness. This chapter reviews the literature published in this field, provides a selective overview of the extant published research based on such data, and offers suggestions for the continuing access and use of such datasets.

Keywords: epidemiology, social context, community, neighborhood, city

1. Introduction

The built environment encompasses aspects of the inhabited constructed world that affect such fundamental human activities and needs as play, rest, work, mobility, shelter, etc. These constructed environments typically consist of homes, buildings, streets, open and green spaces, neighborhoods and communities, grocery stores, healthcare facilities, schools, and other artifacts. Research to date has consistently found that aspects of the built environment affect a person's health in different ways. For example, sedentary habits are reinforced for individuals with limited access to sidewalks or bike paths, and a higher prevalence of obesity is positively related to the proximity of fast-food restaurants to primary dwellings. In addition, urban density generally increases the likelihood of significant air and noise pollution, which in some studies is related to the prevalence of respiratory disease and hearing. Exposure to loud noise can also cause high blood pressure, heart disease, sleep disturbances, and stress. And both laboratory experiments and field-based quasi-experiments have repeatedly shown that exposure to greenspace helps reduce stress or restore attention.

Scholars from the distinct yet converging fields of landscape architecture, city planning, environmental science, psychology, public health, preventive medicine, and geography each provide unique perspectives on the complex multi-level relationships between the built environment and human health. Environmental psychologists

typically conduct research in lab and field settings to explore psychophysiological individual reactions to natural and constructed environments, whereas epidemiologists usually examine such phenomena through aggregate cross-sectional or longitudinal observational studies. Researchers in landscape architecture, city planning, and geography focus primarily on the spatial distribution of certain city elements and human health associations through the conduct of field surveys and experiments in actual human settlements.

Regardless of either the discipline or method, nearly all such research requires some form of social demographic data. Social and demographic factors such as age, gender, race, income, or education all affect human health and therefore need to be controlled or accounted for in any scientific investigation. Some studies control such variation through research design, for example, by an explicit focus on older adults' mental well-being and how it is affected by neighborhood characteristics [1] or an exploration of children's obesity and the influence of their community environment [2]. Alternatively, social demographic variables often moderate the built environment and human health relationships. In these cases, the social demographic variables are controlled through statistical models. For example, the health benefits of exposure to urban greenspace have repeatedly been found to be stronger among lower-income and lower-educated populations [3]. Due to the role of such effects, it is necessary to include social demographic factors in all studies on the built environment and human health relationships.

This chapter offers an overview of the social demographic data that is widely used in the built environment and human health field, through a scoping review of the literature. We will introduce (i) what types of social demographic data have been used; (ii) how to currently obtain these data; (iii) the accuracy, completeness, reliability, relevance, and timeliness of such data; and (iv) offer suggestions for the future use of such datasets in the built environment and population health studies.

2. Methods

A scoping review methodology was selected to best accommodate the wide diversity of studies that have addressed built environment and human health relationships. We searched the literature using Scopus®. We employed "built environment," "health," and "social" as search terms, stopping the search as of January 3, 2023. Subject areas that were not expected to cover the literature examining built environment and health topics (e.g., physics, math, chemistry, etc.) were excluded.

The retrieved literature was screened by reviewing titles and abstracts in terms of relevance to the research question. The four inclusion criteria were that each publication must be an original research article, use a quantitative analysis method, examine built environmental factors, and explore human health outcomes and unhealthy behaviors. The remaining relevant documents were subsequently considered for full-text review and assessed for eligibility. Ineligible documents either did not have a full text available, were not in English, did not concern human health, or did not include variables or analyses relevant to the built environment.

Social demographic variables, the analytical unit of a study, the country in which the studies were conducted, and the types of built environment factors and health-related variables investigated were all codified and reviewed.

3. Overview of the studies on the built environment and human health

The initial search yielded 1398 documents that were screened for relevance. Of these, 927 were excluded, resulting in 471 documents being considered for a full-text review. Of these, 132 were excluded because they were neither in English, original research, analyzed quantitatively, nor included either built environment or human health variables. A total of 339 documents were then reviewed. As shown in **Figure 1**, there has been a marked increase in the number of original articles per year on this topic.

Of the 339 studies, there are 11 articles exploring the topic in two or more countries. The largest proportion of single-country studies was in high-income countries, as defined by the World Bank [4]. Specifically (**Figure 2**), the United States alone published 40.85% of the original research articles. Followed by Australia (34%), Canada (29%), China (21%), and the United Kingdom (20%). 81.25% of the articles were conducted in high-income countries. Only 1.5% of the articles were conducted in low-income countries, including India, Bangladesh, Ghana, Kenya, and Nigeria.

Based on our review, three factors emerged as the most commonly examined in the literature: namely, the effects of promoting physical activity, promoting a healthy food environment, and enhancing contact with nature. Ninety-three percent of the

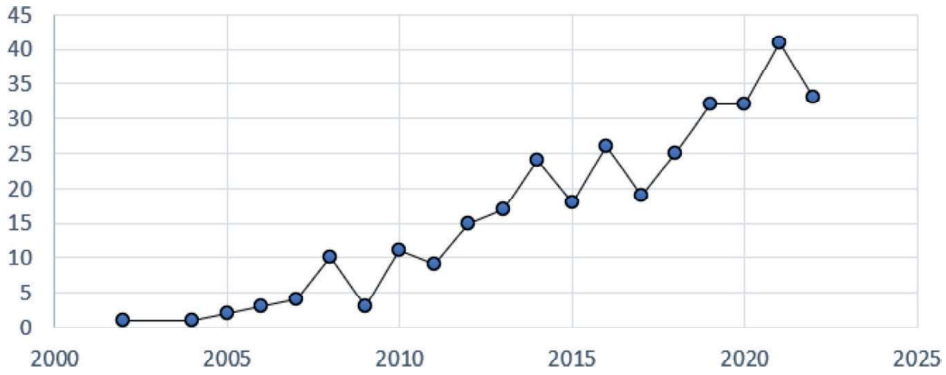

Figure 1.
Number of original research articles published between 2002 and 2022.

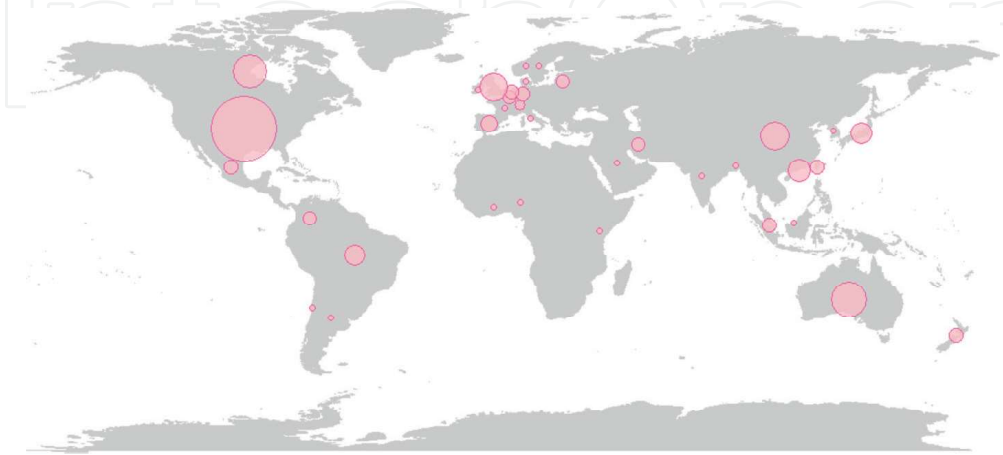

Figure 2.
Map of study countries.

studies used data at the individual level, and some combined with data from the neighborhood level. Only 22 (6%) out of the 339 studies used data at census block groups, census tracts, communities, districts, cities, counties, zip code levels, etc., and adopted an ecological study design. The most widely used built environment variables include walkability of the neighborhood, vacancy rate, land use mix, street connectivity, availability of community facilities, residential density, greenspace and recreational resources, transportation facilities or destinations (i.e., bus, metro, or subway station density), and food environment (both convenience and grocery store density). These variables were derived either from surveys or measured by using Geographic Information System (GIS).

A variety of health outcomes have been reported in the literature. Nearly 15%(n = 50) of the studies explored two or more health outcomes. For example, simultaneously focusing on walking, weight status, and obesity; physician-diagnosed chronic disease and self-rate general health; or physical and mental health. For the studies that focused on single health outcomes, 54.0% (n = 156) investigated physical activity, sedentary behavior, weight, and obesity-related outcomes, whereas slightly more than 15% (n = 44) focused on mental health, cognitive health, and dementia. Other studies explored chronic disease mortality and/or morbidity including hypertension, diabetes, heart disease, asthma, etc. Four studies focus on healthcare system utilization, and 10 studies examined the self-reported quality of life.

4. Role of social demographic variables

Social demographic factors have been included in almost all (336/339) of the studies reviewed. They have been used mostly as control variables in quantitative statistical analyses, yet a few studies went further and explored the moderating role of social-economic factors on the built environment and human health relationships [5]. The most widely used variables include age, gender, education, income, race/ethnicity, marital status, employment, etc. The data-collecting methods and sample sizes limit the ability of any single study to control all these social demographic variables, yet almost all the articles reviewed attempted to control at least some of these social demographic factors (**Figure 3**).

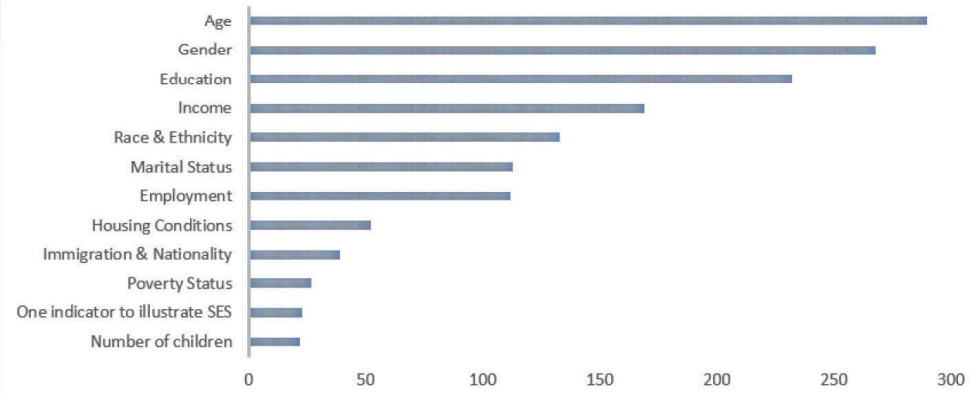

Figure 3.
Frequency of commonly used social demographic variables in the reviewed built environment and human health relationship studies.

The reason for such inclusion is that the built environment and human health relationship are complex and can be influenced by social demographic variables. Failure to incorporate these demographic factors is generally believed to distort our understanding of how various aspects of the built-in environment affect human health. Age plays a role in that older adults may be more sensitive to environmental factors such as air pollution and extreme temperatures and may have different mobility needs than younger adults [6]. Gender is influential in that men and women may have different patterns of physical activity and access to certain built environment resources, such as public transportation or parks [7, 8]. Race and ethnicity are important because minority groups may experience disparities in access to healthy food options, safe neighborhoods, and green spaces [9–12]. Lower-income individuals may have limited access to resources such as safe housing, healthy food options, and healthcare services, which can impact their health outcomes [13, 14]. Individuals with higher levels of education may have a greater awareness of health behaviors and resources and may be more likely to engage in healthy behaviors [15, 16]. Married individuals may have greater social support and access to resources than single individuals, which can impact their health outcomes [17–19].

Therefore, when social demographic variables are included in studies of the built environment and human health, researchers can better understand how different groups are affected by their environment and identify strategies for promoting health equity. Walkable neighborhoods can promote physical activity and improve health outcomes. However, when social demographic variables are included in the analysis, the relationship varies. The relationship was stronger for women than for men, and the association varied by race and ethnicity [20, 21]. Male cardiovascular disease and respiratory disease mortality rates decreased with increasing green space, but no significant associations were found for women [22]. Perceived walkability was directly linked to the happiness of people aged 36–45 and, to a lesser extent, those aged 18–35 years of age [23]. Studies regarding access to healthy food and obesity relationships reported social demographic variables such as income and race can influence access to healthy food [13, 24]. The association between supermarket access and obesity varied by income level, with the relationship being stronger for lower-income individuals [25, 26]. The relationship between green spaces and mental health was weaker for higher socioeconomic groups [27]. Greenspace exposure had no correlation with birth weight, until when maternal education was considered; an increase in birth weight was detected among the lowest education level group [28].

It is important to recognize that the inclusion of social demographic variables when investigating the relationship between the built environment and health varies depending on the specific health outcome being studied, the demographic group being considered, the built environment factors being examined, and even the place where the studies are conducted. A study focusing specifically on birth outcomes did not include gender as a variable, as all the participants are women biologically [29]. A study examining the environment and changes in physical activity of high school students did not control age, given high school students are generally within the same age group [30]. Studies on children and adolescents often include the education level of parents as a controlling factor instead of the grade level of the participants, and there is generally no need to control marital status [31–34]. Some studies from Australia used the Index of Relative Socio-Economic Advantage and Disadvantage to control social economic status as this index is readily available and widely used in Australia [35–37].

Given the importance of examining social demographic variables in the built environment and human health studies, it is also vital to effectively obtain social demographic data for research. There are five major sources in the literature. The first one is accessing data from national, state, and local surveys. The second one is to obtain administrative data. The third is to utilize data from previous longitudinal studies. The fourth way is to survey by using questionnaires, either organized via paper-based, online, telephone-based, or face-to-face questionnaires. The new trend is data collected from health-related applications on mobile devices. Some studies obtained data from two or more of these sources [38].

5. Access data from the national, state, and local surveys

Socio-demographic data is widely available from many national, state, and local surveys. Government institutions from developed countries usually collect health and related social demographic data periodically and often offer freely accessible data to the public. The data collected by such institutes are generally considered to be of high quality and reliable. These countries usually established rigorous standards and procedures to ensure the accuracy, completeness, and reliability of their data, including the use of standardized data collection methods and quality control procedures. Many developing countries, however, either do not conduct such surveys, or the survey results are of poor quality, making them either unavailable or unreliable for academic research purposes [39–41].

In the United States, one of the widely used built environment and human health databases is the Behavioral Risk Factor Surveillance System (BRFSS). It is a national telephone survey system used by the Centers for Disease Control and Prevention (CDC) to collect data on health behaviors, chronic health conditions, and the use of preventive services among the adult population. A variety of studies have utilized such data in the built environment and human health field [42–45]. The BRFSS collects data from a random sample of noninstitutionalized adults in each state and the District of Columbia. The survey is conducted monthly and has been in operation since 1984. Some of the topics covered by the BRFSS include tobacco use, physical activity, diet, chronic disease, access to health care, and preventive health practices.

BRFSS also collects social demographic data as part of its survey. The demographic data collected by the BRFSS includes age, gender, race/ethnicity, education level, employment status, income, and marital status, among others. The demographic information collected is used to provide a descriptive profile of the respondents and to help identify and address health disparities in the population. By collecting information on social and demographic characteristics, the BRFSS can help to understand how health behaviors and outcomes are related to individual characteristics and can inform the development of targeted public health interventions. The database can be accessed through CDC's official website (https://www.cdc.gov/brfss/). This data is matched and combined with the built environment data, usually based on zip codes when used in research.

The National Center for Health Statistics (NCHS) is a division of the Centers for Disease Control and Prevention (CDC) and is the nation's principal health statistics agency. A study examining associations between environmental quality and preterm birth in the United States utilized data from NCHS on 24,483,348 participants [29]. NCHS provides a wide range of health data and information, including (a) vital statistics: data on births, deaths, marriages, and divorces. (b) National Health Interview

Survey (NHIS): a large, ongoing survey that collects information on the health of the U.S. population. (c) National Health and Nutrition Examination Survey (NHANES): a nationally representative survey of the U.S. population that provides data on health and nutritional status, including information on diet, physical activity, and exposure to environmental contaminants. Similar surveys exist in other countries. South Korea, for example, also conducts a national-level survey (i.e., Korea National Health and Nutrition Examination Survey [KNHNES]) [46]. (d) Mortality data: data on causes of death in the United States. (e) Health care utilization: data on the use of health care services, including hospital stays, physician visits, and prescription drug use. A study focusing on greenspace and historical redlining used such data to explore their impact on emergency department utilization [47]. (f) Health behavior and risk factors: data on behaviors and risk factors that impact health, including smoking, alcohol use, and physical activity. (g) Chronic conditions: data on the prevalence and impact of chronic conditions, such as heart disease, diabetes, and cancer.

NCHS data is publicly available and can be accessed through the NCHS website (https://www.cdc.gov/nchs/). Many of the data sets collected by the National Center for Health Statistics (NCHS) include social demographic information such as age, race/ethnicity, education, income, and other demographic characteristics. For example, the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES) both collect data on social demographics as part of their surveys. Similarly, vital statistics data on births, deaths, marriages, and divorces include demographic information such as age, race/ethnicity, and other characteristics of the individuals involved.

Another major resource is data available through the American Community Survey (ACS). The ACS is a nationwide survey conducted by the U.S. Census Bureau to collect data on the social, economic, and housing characteristics of the U.S. population. The ACS is designed to provide data on small geographic areas, including neighborhoods and communities, and is conducted continuously throughout the year. The survey collects data on a range of topics, including age, race/ethnicity, income, education, employment, housing, and others. The ACS is one of the largest surveys conducted by the U.S. government and provides important information on the social and economic characteristics of the U.S. population. The ACS data is available to the public and can be accessed through the U.S. Census Bureau's website (www.census.gov). To be useful, however, ACS data needs to be combined with built environment data and health data at the same spatial location and scale. For example, a study focusing on neighborhood-built environments, obesity risks, and racial composition combined data from ACS as well as the Southeastern Pennsylvania Household Health Survey [48].

Some states and localities conduct surveys that collect data on specific health and environmental issues. The California Health Interview Survey (CHIS) is a well-known example. CHIS is a large, statewide health survey conducted by the UCLA Center for Health Policy Research. The survey is designed to collect data on the health and well-being of California residents and is one of the largest state health surveys in the nation. The survey collects data on a range of topics, including health behaviors, chronic conditions, access to health care, health insurance coverage, and more. The survey also collects social demographic information such as age, race/ethnicity, education, income, and others. CHIS data is used by researchers to study the determinants of health and health disparities. For example, a study examining the influence of urban tree canopy on health utilized the data from CHIS [49]. The CHIS data is publicly available and can be accessed through the UCLA Center for Health Policy

Research's website (https://www.uclahealthpolicy.org/). This website provides data sets, documentation, and tools for accessing and analyzing the data, including the CHIS Data Online tool, which allows users to create custom data tables and visualizations. The data is available for a variety of geographic levels, including state, county, and subcounty areas, and can be used to study health and healthcare issues at the state and local levels.

6. Administrative data

Six percent (n = 21) of the examined studies utilized administrative data, yet a growing number of studies are using such data in recent years in health-related studies [50]. Administrative data, including data from Medicaid, Medicare, and electronic health records (EHRs), can provide a wealth of information on the health status, utilization, and costs of specific populations. Administrative data is generated from the day-to-day operations of healthcare organizations, such as hospitals and clinics, and is used to manage patient care, billing, and reimbursement. One advantage of administrative data is that it is often comprehensive, covering a large number of individuals and including a wide range of health information. A study examining built environmental factors and body mass index trajectories among youth used EHRs from 2001 to 2012 from 163,820 youth aged 3–18 years from 1288 communities [51]. Medicaid and Medicare data can provide information on the health status, utilization, and costs of older adults and individuals with disabilities, while the data can provide detailed information on patient diagnoses, medications, and other health-related information. This allows researchers to examine particular health measures, like changes in left ventricular ejection fraction, that are otherwise difficult to capture without such data sets [52].

Administrative data often include social demographic information such as age, race/ethnicity, gender, and other demographic characteristics. It is important to note, however, that the completeness and accuracy of the demographic information in administrative data can vary and may be subject to data quality issues concerning reliability and validity. As the data is generated from the day-to-day operations of healthcare organizations, errors can occur during the data collection and recording process and may occasionally fail to capture all relevant information. Administrative data may be generated using different data standards and definitions, which can impact its comparability across different data sources and over time. Compared to the official survey data collected by government agencies mentioned above, administrative data requires more effort from the researcher to ensure data integrity.

Additionally, privacy and security concerns must be carefully considered when using administrative data for research. The data must be protected following privacy and security regulations, such as the Health Insurance Portability and Accountability Act (HIPAA). This is a US federal law that sets standards for the privacy and security of protected health information (PHI) in the healthcare industry. It requires healthcare providers, health plans, and healthcare clearinghouses to maintain the privacy and security of individuals' PHI and to provide individuals with certain rights concerning their PHI. HIPAA also establishes penalties for noncompliance and provides for enforcement of its requirements by the Department of Health and Human Services. This use of this data will likely require data use agreements, data security measures, and other measures to protect the confidentiality and security of the data. Overall, when used appropriately, administrative data is a valuable source of information for researchers.

The Centers for Medicare & Medicaid Services (CMS), a federal agency within the US Department of Health and Human Services (HHS), is the primary source of Medicare and Medicaid data. CMS collects, maintains, and provides access to a variety of data on Medicare and Medicaid beneficiaries, providers, and payments. The website (see https://www.cms.gov/) provides access to a range of data sets, including the Medicare Provider Utilization and Payment Data, the Medicaid Statistical Information System, and the National Health Expenditure Accounts. Researchers can also submit a Freedom of Information Act (FOIA) request or a data use agreement request to obtain specific Medicare or Medicaid data sets.

There are a variety of sources that researchers can contact directly to obtain electronic health records. They are (a) Healthcare providers: hospitals, clinics, and individual practitioners, that maintain and use EHRs to manage patient information. Data is available but typically only accessible by certain professionals, such as physicians. In such situations, active collaborations with staff physicians are an option. A study exploring the role of neighborhood characteristics and the racial disparities in childhood obesity obtained EHRs of 44,810 children seen at 14 Massachusetts pediatric practices in 2011–2012 [53]. (b) Health information exchanges (HIEs): HIEs are organizations that collect and securely exchange health information between healthcare providers. HIEs can provide access to EHR data for research and other purposes, subject to certain conditions and requirements. (c) Government agencies such as the National Institutes of Health (NIH) and the Agency for Healthcare Research and Quality (AHRQ), maintain and provide access to EHR data for research and other purposes. (d) Commercial vendors, such as Epic Systems, Cerner, and Allscripts, offer EHR systems for healthcare providers and organizations and may also provide access to EHR data for research and other purposes.

7. Secondary data from previous longitudinal studies

Using data from previous longitudinal studies for a cross-sectional built environment and human health study can be a valuable approach, as it can provide a historical perspective on the relationship between the built environment and human health. Previous longitudinal health studies may or may not include social demographic variables. The type and amount of data collected in a study can vary depending on the specific research questions and methods used. Typically, however, social demographic variables can be important factors, and many longitudinal health studies routinely collect such data.

Relying on secondary data, however, can present challenges. For example, the data may not be representative of current conditions, the measurement of the built environment and health outcomes may have changed over time, and the data may fail to dovetail with the specific research questions and methods. It is, therefore, important to carefully review the data and consider its limitations when conducting the analysis. When selecting the relevant data, factors such as the time period of the study, the geographic location, the type of health outcomes measured, and the type of built environment data collected must be examined, as well as if it is even possible to spatially and temporally link the health and social demographic data with the built environment data. It is also important to determine what variables are included and whether they are suitable for the specific research questions and methods being used in the current study. In some cases, additional data sources or secondary data analysis may be necessary to obtain the required social demographic variables.

There are several sources where a researcher can check websites or contact directly to inquire about obtaining data from previous longitudinal studies. Sometimes, such data are not free, and extramural funding is needed to purchase access. The sources are (a) Research institutions and universities often conduct and archive data from longitudinal studies. For example, one study focusing on the weight status of low-income elementary school children obtained data from the third author from her NIH-funded grant [54]. In a study from China, the researchers used survey data collected by the Public Health College of Fudan University [55]. (b) Government agencies, such as the NIH and the CDC, often conduct and archive data from longitudinal studies. A study examining childhood obesity obtained data from the NIH environmental influences on child health outcomes (ECHO) program [56]. (c) Professional organizations and societies, such as the American Public Health Association (APHA) and the American Medical Association (AMA), may conduct and archive data from longitudinal studies. (d) There are several publicly available databases, such as the National Longitudinal Study of Adolescent to Adult Health (Add Health: https://addhealth.cpc.unc.edu/), that provides access to data from longitudinal studies.

8. Firsthand data collected through questionnaires by researchers

Questionnaires can be designed to collect data from all kinds of aspects. It can be used to gather information on the exposure of individuals to various built environment factors, such as access to green spaces, exposure to air pollution, or availability of sidewalks for active transportation [57]. It can also be used to gather information on physical activity and sedentary behavior, including frequency, intensity, and duration, and the built environment factors that influence these behaviors [58]. Researchers can also gather information on the perceptions and attitudes of individuals toward their neighborhood and the built environment, including perceived safety, walkability, and social cohesion [59, 60]. Studies could use questionnaires to collect information on various health outcomes, including physical and mental health, health behaviors, and health-related quality of life [61, 62]. Also, questionnaires can be used to gather demographic information, such as age, gender, education, and income, which can be used to control potential confounding variables in the analysis.

In this review, approximately 40% (n = 138) of the published original research articles collected firsthand social demographic data through a custom-designed questionnaire. These studies are typically conducted at the individual level and are usually able to collect all the social demographic factors that the researchers are interested in. Due to the reason that the questionnaire focuses on a single study, the sample size is usually relatively small compared to the above-mentioned national or state surveys. A study focusing on spatial accessibility to physical activity facilities and food outlets and overweight among French youth randomly selected 3293 students from eastern France middle schools [63]. Questionnaires are especially useful when the available secondary data is insufficient or its quality is questionable.

For developing countries where the above-mentioned secondary data is unavailable or unreliable, collecting data via questionnaire is the major method to conduct built environment and human health studies. In most cases, such questionnaires usually focus on one geographic local area. For example, a study from China exploring social capital, built environment, and mental health used data collected from 591 participants in Nanjing City [64]. Yet occasionally, it also allows researchers to

conduct across-country studies where the same set of questionnaires has been used across multiple different districts, communities, or even countries, allowing the comparison between geographic areas [65]. For instance, A study examining the impact of neighborhood environments on sedentary time collected data across 10 countries [66]. The questionnaire can also be easily modified if, during the study, the researcher determines it is necessary to change the questionnaire. Questionnaires are relatively easy to administer and can reach a sufficient number of participants in a short period of time, making it a convenient method for collecting data. In addition, questionnaires are often relatively inexpensive to administer; in some cases, questionnaires can be designed and distributed purely online; if there are no incentives for the participants, then it is virtually free.

Custom questionnaires allow researchers to dig deeper into the social demographic characteristics of a particular group of participants, going beyond the routinely collected variables such as age, gender, race, education, income, and employment. For example, studies focusing on school students could collect information regarding which grade the participants are at [67]. Other research examining immigrants' health could collect factors such as country of birth and how long they have been in the original country and destination country, as the built environment in different countries could vary significantly [68].

One limitation to using questionnaires in the built environment and human health studies is that the health measures are based on self-reported signs and symptoms [69] and thus are subject to the vagaries of memory and an accurate understanding of the questions. With respect to studying sensitive topics, however, questionnaires may be preferred because they do not require physical examination or intervention. Below is a questionnaire constructed to illustrate this manner of data collection. These questions are just an example and can be adapted to meet the specific needs of a study. It is important to note that the response options must be culturally sensitive and

Data Types	Example Questions	Example Response Options
Age	What is your current age?	18–24, 25–34, 35–44, 45–54, 55–64, 65, or older
Gender	What is your gender?	Male, Female, Other
Education	What is the highest level of education you have completed?	Less than high school, High school/GED, Some college, College degree, Graduate degree
Employment	What is your current employment status?	Employed full-time, Employed part-time, Unemployed, Retired, Other
Income	What is your annual household income?	Less than $20,000, $20,000–$39,999, $40,000–$59,999, $60,000–$79,999, $80,000 or more
Ethnicity	What is your ethnicity/race?	White, Black or African American, Asian, Native American or Alaska Native, Native Hawaiian or Pacific Islander, Other
Marital status	What is your marital status?	Single, Married, Divorced, Widowed, Separated
Number of children	How many children do you have?	None, 1, 2, 3, 4, or more
Residential location	What is your current residential location?	Urban, Suburban, Rural

Table 1.
Example questions and response options for collecting social demographic data.

appropriate for the target population, and it may be necessary to include additional demographic variables that are relevant to your study. For instance, it is important to collect race/ethnicity data in the United States, yet it is much less worthwhile to collect such data in big cities in China due to the ethnic health disparity between Han and non-Han is believed to depend more on the regions where people live than on actual ethnic differences (**Table 1**) [70].

9. Emerging trend: data collected from health-related applications on mobile devices

Smartphones, smartwatches, and fitness bands often come with accelerometers and GPS facilities which track data regarding physical activity, as well as geographic locations. With the growing utilization of smartphone fitness apps, many researchers now collect such data for the built environment and human health studies. The amount of data that can be collected in his manner is often quite large and is generated in real-time. A study from China published in 2021 examining built environment and physical activity obtained such data for more than 3 million users [71]. Another study from Canada explored neighborhood characteristics and running, and purchased user data from Strava Metro, which allows the researchers to analyze a total of 242,265 segments [72]. Strava anonymized and aggregated user activities in a spatial file made up of line segments representing routes. Segments contain count information for Strava users and running activities using bins of five to protect individual privacy. Strava categorizes users by gender and age groups which offer a minimum level of demographic information.

Such methods often provide large amounts of physical activity data with relatively accurate location information, which researchers can match with built environment measures. Anonymized aggregated data, however, typically obscures the fine-grained social demographic information necessary to examine the moderating effects of age, gender, income, etc., on the built environment and human health relationships. In addition, data collected through mobile apps may not be representative of the general population; rather, it might be more likely to represent a physically active or health awareness population.

While with care in mind, researchers can obtain a rich source of information about individual behavior, preferences, and experiences in the built environment from mobile devices. With the right apps, data can be collected on the physical activity levels of users, such as the number of steps taken, distance traveled, and calories burned. This data can be used to study the relationship between physical activity and the built environment, such as the availability of parks and sidewalks. They can also facilitate the collection of data on travel behavior, such as mode of transportation, travel time, and the routes taken. This data can be used to study the impact of the built environment on travel behavior, affording the exploration of transit accessibility and pedestrian-friendly design. They facilitate the collection of data on health outcomes, such as sleep quality, stress levels, and mood. This data can be used to study the relationship between the built environment and health outcomes, such as the impact of green space on stress reduction. They can also help in the collection of data on user preferences and experiences, such as ratings of parks and public spaces. This data can be used to study the impact of the built environment on user experience and to inform decisions about the design and maintenance of public spaces.

10. The way forward

In the realm of the built environment and human health research, the inclusion of social demographic variables and the utilization of reliable and valid data sources hold paramount importance. When collecting social demographic data, it is beneficial to consider (1) accuracy – ensuring correct and precise data; (2) completeness – including all relevant information in the dataset; (3) reliability – consistency and stability of the data over time and sources; (4) relevance – aligning the data with research objectives and questions; (5) timeliness – using up-to-date and current data; and (6) magnitude – the scale and abundance of the data. **Table 2** offers a detailed description of these characteristics of social demographic data collected from the major sources.

Moving forward, a key future direction lies in expanding and refining data collection methods. This involves the integration of emerging technologies, such as wearable devices, mobile applications, and sensor networks, to capture real-time and objective data on individuals' behaviors, exposures, and environmental attributes. Furthermore, the integration of multiple data sources might be a potentially valuable way to enrich the breadth and depth of social demographic variables. For example, by combining traditional survey-based approaches with health data collected from mobile devices, geospatial data, and digital footprints, researchers can create comprehensive databases that capture diverse aspects of individuals' lives.

Government and health data-collecting agencies could foster the integration of data from different sources to create comprehensive datasets. Emphasize interoperability to facilitate seamless data sharing and exchange between different platforms and systems, enabling researchers to leverage multiple data sources effectively. This integration fosters a more holistic understanding of the social, economic, and environmental contexts in which individuals interact with their built environment, enabling researchers to uncover nuanced associations and identify potential mechanisms through which social demographics influence health outcomes.

New advanced analytics and methodologies could be developed and utilized, such as machine learning, spatial analysis, and data mining, to extract meaningful insights from complex social demographic datasets. Embrace innovative methodologies that can handle large-scale and diverse data sources effectively. Researchers could also engage in validation studies to assess the reliability of new data sources and methodologies, comparing them against well-established measures. By establishing strong foundations of data reliability and validity, the field can build a solid evidence base to inform policymaking, urban planning, and public health interventions.

Data Source	Accuracy	Completeness	Reliability	Relevance	Timeliness	Magnitude
National, state, and local surveys	***	***	***	***	**	***
Administrative data	**	*	**	*	**	**
Previous longitudinal studies	***	**	***	**	*	*
Questionnaires	***	***	***	***	**	*
Apps on mobile devices	*	*	*	***	***	***

Table 2.
Summary of data characters from different sources.

11. Conclusion

In conclusion, this chapter has provided a comprehensive overview of the demographic data used in built environment and population health studies. From traditional data sources like national surveys to emerging data sources such as administrative databases and mobile devices, the chapter highlights the importance of data quality and its impact on decision-making. The chapter has also discussed the trends in demographic data usage and variable selection and offers suggestions for accessing and using different datasets in the built environment and population health studies. Understanding the relative strengths and limitations of accessible and burgeoning diverse data sets is crucial to making informed decisions and affording valid field research. This chapter hopefully serves as a useful resource for researchers and practitioners seeking to use demographic data to deepen our understanding of the important and complex relationships between the built environment and population health.

Author details

Huaqing Wang[1*] and Louis G. Tassinary[2]

1 Utah State University, Logan, Utah, USA

2 Texas A&M University, College Station, Texas, USA

*Address all correspondence to: huaqing.wang@usu.edu

References

[1] Pun VC, Manjourides J, Suh HH. Close proximity to roadway and urbanicity associated with mental ill-health in older adults. Science of the Total Environment. 2019;**658**:854-860

[2] Spence JC, Cutumisu N, Edwards J, et al. Influence of neighbourhood design and access to facilities on overweight among preschool children. International Journal of Pediatric Obesity. 2008;**3**:109-116

[3] Wang H, Tassinary LG. Effects of greenspace morphology on mortality at the neighbourhood level: A cross-sectional ecological study. Lancet Planet Health. 2019;**3**:e460-e468

[4] New World Bank country classifications by income level: 2021-2022. Available from: https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2021-2022. 2021 [Accessed 20 February 2023]

[5] Sun P, Lu W, Song Y, Gu Z. Influences of built environment with hilly terrain on physical activity in Dalian, China: An analysis of mediation by perceptions and moderation by social environment. International Journal of Environmental Research and Public Health. 4 Dec 2019;**16**(24):4900. DOI: 10.3390/ijerph16244900

[6] Hong Y-C. Aging society and environmental health challenges. Environmental Health Perspectives. 2013;**121**:a68-a69

[7] Azevedo MR, Araújo CLP, Reichert FF, et al. Gender differences in leisure-time physical activity. International Journal of Public Health. 2007;**52**:8

[8] Cohen DA, Williamson S, Han B. Gender differences in physical activity associated with urban Neighborhood parks: Findings from the National Study of Neighborhood parks. Womens Health Issues. 2021;**31**:236-244

[9] Block JP, Scribner RA, DeSalvo KB. Fast food, race/ethnicity, and income: A geographic analysis. American Journal of Preventive Medicine. 2004;**27**:211-217

[10] Price JH, Khubchandani J, McKinney M, et al. Racial/ethnic disparities in chronic diseases of youths and access to health Care in the United States. BioMed Research International. 2013;**2013**:787616

[11] Richardson LD, Norris M. Access to health and health care: How race and ethnicity matter. Mount Sinai Journal of Medicine. 2010;**77**:166-177

[12] Li M, Zhong R, Zhu S, Ramsay LC, Li F, Coyte PC. Access to community living infrastructure and its impact on the establishment of community-based day care centres for seniors in rural China. International Journal of Environmental Research and Public Health. 6 Jun 2018;**15**(6):1184. DOI: 10.3390/ijerph15061184

[13] Ziso D, Chun OK, Puglisi MJ. Increasing access to healthy foods through improving food environment: A review of mixed methods intervention studies with residents of low-income communities. Nutrients. 2022;**14**:2278

[14] Kushel MB, Gupta R, Gee L, et al. Housing instability and food insecurity as barriers to health care among low-income Americans. Journal of General Internal Medicine. 2006;**21**:71-77

[15] Fletcher JM, Frisvold DE. Higher education and health investments: Does more schooling affect preventive health care use? Journal of Human Capital. 2009;**3**:144-176

[16] Ross CE, Wu C. The links between education and health. American Sociological Review. 1995;**60**:719-745

[17] Joung IMA, van der Meer JBW, Mackenbach JP. Marital status and health care utilization. International Journal of Epidemiology. 1995;**24**:569-575

[18] Verbrugge LM. Marital status and health. Journal of Marriage and the Family. 1979;**41**:267-285

[19] August KJ, Sorkin DH. Marital status and gender differences in managing a chronic illness: The function of health-related social control. Social Science & Medicine 1982. 2010;**71**:1831-1838

[20] Choi A, Blanco L, Hays RD. Race and ethnicity differences in walking and associations with neighborhood perceptions among older adults in California. Journal of Applied Gerontology: The Official Journal of the Southern Gerontological Society. 2022;**41**:2499-2510

[21] Althoff T, Sosič R, Hicks JL, et al. Large-scale physical activity data reveal worldwide activity inequality. Nature. 2017;**547**:336-339

[22] Richardson EA, Mitchell R. Gender differences in relationships between urban green space and health in the United Kingdom. Social Science & Medicine. 2010;**71**:568-575

[23] Leyden KM, Hogan MJ, D'Arcy L, Bunting B, Bierema S. Walkable neighborhoods: Linkages between place, health, and happiness in younger and older adults. Journal of the American Planning Association. 11 Apr 2023:1-4

[24] Hill JL, Chau C, Luebbering CR, Kolivras KK, Zoellner J. Does availability of physical activity and food outlets differ by race and income? Findings from an enumeration study in a health disparate region. International Journal of Behavioral Nutrition and Physical Activity. 6 Sep 2012;**9**:105. DOI: 10.1186/1479-5868-9-105

[25] Moore LV, Die Roux AV. Associations of neighborhood characteristics with the location and type of food stores. American Journal of Public Health. 2006;**96**:325-331

[26] Drewnowski A, Aggarwal A, Hurvitz PM, et al. Obesity and supermarket access: Proximity or Price? American Journal of Public Health. 2012;**102**:e74-e80

[27] Maas J, Verheij RA, Groenewegen PP, et al. Green space, urbanity, and health: How strong is the relation? Journal of Epidemiology and Community Health. 2006;**60**:587-592

[28] Dadvand P, de Nazelle A, Figueras F, et al. Green space, health inequality and pregnancy. Environment International. 2012;**40**:110-115

[29] Rappazzo KM, Messer LC, Jagai JS, Gray CL, Grabich SC, Lobdell DT. The associations between environmental quality and preterm birth in the United States, 2000-2005: A cross-sectional analysis. Environmental Health. 9 Jun 2015;**14**:50. DOI: 10.1186/s12940-015-0038-3

[30] Molina-García J, Queralt A, Castillo I, et al. Changes in physical activity domains during the transition out of high school: Psychosocial and environmental correlates. Journal

of Physical Activity & Health. 2015;**12**:1414-1420

[31] Wong BY-M, Ho S-Y, Lo W-S, et al. Longitudinal relations of perceived availability of neighborhood sport facilities with physical activity in adolescents: An analysis of potential moderators. Journal of Physical Activity & Health. 2014;**11**:581-587

[32] Carlson JA, Sallis JF, Kerr J, et al. Built environment characteristics and parent active transportation are associated with active travel to school in youth age 12-15. British Journal of Sports Medicine. 2014;**48**:1634-1639

[33] Vanhelst J, Béghin L, Salleron J, et al. A favorable built environment is associated with better physical fitness in European adolescents. Preventive Medicine. 2013;**57**:844-849

[34] Bringolf-Isler B, Grize L, Mäder U, et al. Built environment, parents' perception, and children's vigorous outdoor play. Preventive Medicine. 2010;**50**:251-256

[35] Mazumdar S, Learnihan V, Cochrane T, Phung H, O'Connor B, Davey R. Is walk score associated with hospital admissions from chronic diseases? Evidence from a cross-sectional study in a high socioeconomic status Australian city-state. BMJ Open. 8 Dec 2016;**6**(12):e012548. DOI: 10.1136/bmjopen-2016-012548

[36] Abe T, Carver A, Sugiyama T. Associations of neighborhood built and social environments with frailty among mid-to-older aged Australian adults. Geriatrics & Gerontology International. 2021;**21**:893-899

[37] Collyer C, Bell MF, Christian HE. Associations between the built environment and emotional, social and physical indicators of early child development across high and low socioeconomic neighbourhoods. International Journal of Hygiene and Environmental Health. Jun 2022;**243**:113974. DOI: 10.1016/j.ijheh.2022.113974

[38] Jaime PC, Duran AC, Sarti FM, et al. Investigating environmental determinants of diet, physical activity, and overweight among adults in Sao Paulo, Brazil. Journal of Urban Health. 2011;**88**:567-581

[39] Poor data hurts African countries' ability to make good policy decisions. Quartz. Available from: https://qz.com/africa/762729/poor-data-is-hurting-african-countries-ability-to-make-good-policy-decisions/. 2016 [Accessed 20 February 2023]

[40] Runde D. The data revolution in developing countries has a long way to go. Forbes; 2017. Available from: https://www.forbes.com/sites/danielrunde/2017/02/25/the-datarevolution-in-developing-countrieshas-a-long-way-to-go/. [Accessed 20 February 2023]

[41] In low-income countries fundamental data issues remain for COVID-19 response. TReNDS. Available from: https://www.sdsntrends.org/blog/covid19andlowincome-countries?locale=en. 2020 [Accessed 20 February 2023]

[42] Xu Y, Wang F. Built environment and obesity by urbanicity in the U.S. Health & Place. 2015;**34**:19-29

[43] Briggs AC, Black AW, Lucas FL, Siewers AE, Fairfield KM. Association between the food and physical activity environment, obesity, and cardiovascular health across Maine counties. BMC Public Health.

3 Apr 2019;**19**(1):374. DOI: 10.1186/s12889-019-6684-6

[44] Cunningham SA, Patel SA, Beckles GL, et al. County-level contextual factors associated with diabetes incidence in the United States. Annals of Epidemiology. 2018;**28**: 20-25.e2

[45] West ST, Shores KA, Mudd LM. Association of available parkland, physical activity, and overweight in America's largest cities. Journal of Public Health Management and Practice. 2012;**18**:423-430

[46] Choi Y, Yoon H. Do the Walkability and Urban Leisure Amenities of Neighborhoods Affect the Body Mass Index of Individuals? Based on a Case Study in Seoul, South Korea. International Journal of Environmental Research and Public Health. 20 Mar 2020;**17**(6):2060. DOI: 10.3390/ijerph17062060

[47] Wang H, Li D. Emergency department visits for mental disorders and the built environment: Residential greenspace and historical redlining. Landscape and Urban Planning. 2023;**230**:104568

[48] Li K, Wen M, Henry KA. Residential racial composition and black-white obesity risks: Differential effects of neighborhood social and built environment. International Journal of Environmental Research and Public Health. 2014;**11**:626-642

[49] Ulmer JM, Wolf KL, Backman DR, et al. Multiple health benefits of urban tree canopy: The mounting evidence for a green prescription. Health & Place. 2016;**42**:54-62

[50] Ulrich EH, So G, Zappitelli M, Chanchlani R. A review on the application and limitations of administrative health care data for the study of acute kidney injury epidemiology and outcomes in children. Frontiers in Pediatrics. 27 Oct 2021;**9**:742888. DOI: 10.3389/fped.2021.742888

[51] Poulsen MN, Glass TA, Pollak J, Bandeen-Roche K, Hirsch AG, Bailey-Davis L, et al. Associations of multidimensional socioeconomic and built environment factors with body mass index trajectories among youth in geographically heterogeneous communities. Preventive Medical Reports. 29 Jun 2019;**15**:100939. DOI: 10.1016/j.pmedr.2019.100939

[52] Zhang Y, Tayarani M, Al'Aref SJ, et al. Using electronic health records for population health sciences: A case study to evaluate the associations between changes in left ventricular ejection fraction and the built environment. JAMIA Open. 2020;**3**:386-394

[53] Sharifi M, Sequist TD, Rifas-Shiman SL, et al. The role of neighborhood characteristics and the built environment in understanding racial/ethnic disparities in childhood obesity. Preventive Medicine. 2016;**91**:103-109

[54] Miles R, Wang Y, Johnson SB. Neighborhood built and social environments and change in weight status over the summer in low-income elementary school children. International Journal of Environmental Research and Public Health. 31 May 2018;**15**(6):1124. DOI: 10.3390/ijerph15061124

[55] Zheng Z, Liu W, Lu Y, Sun N, Chu Y, Chen H. The influence mechanism of community-built environment on the health of older adults: from the perspective of low-income groups.

BMC Geriatrics. 16 Jul 2022;**22**(1):590. DOI: 10.1186/s12877-022-03278-y

[56] Tylavsky FA, Ferrara A, Catellier DJ, et al. Understanding childhood obesity in the US: The NIH environmental influences on child health outcomes (ECHO) program. International Journal of Obesity. 2020;**44**:617-627

[57] Pierce JR Jr, Denison AV, Arif AA, et al. Living near a trail is associated with increased odds of walking among patients using community clinics. Journal of Community Health. 2006;**31**:289-302

[58] Wu ZJ, Song Y, Wang HL, Zhang F, Li FH, Wang ZY. Influence of the built environment of Nanjing's Urban Community on the leisure physical activity of the elderly: An empirical study. BMC Public Health. 6 Nov 2019;**19**(1):1459. DOI: 10.1186/s12889-019-7643-y

[59] Su M, Tan Y-Y, Liu Q-M, et al. Association between perceived urban built environment attributes and leisure-time physical activity among adults in Hangzhou, China. Preventive Medicine. 2014;**66**:60-64

[60] Ali O, Di Nardo F, Harrison A, et al. The link between perceived characteristics of neighbourhood green spaces and adults' physical activity in UK cities: Analysis of the EURO-URHIS 2 study. European Journal of Public Health. 2017;**27**:761-765

[61] Araya R, Montgomery A, Rojas G, et al. Common mental disorders and the built environment in Santiago, Chile. The British Journal of Psychiatry. 2007;**190**:394-401

[62] Camargo DM, Ramírez PC, Fermino RC. Individual and environmental correlates to quality of life in park users in Colombia. International Journal of Environmental Research and Public Health. 19 Oct 2017;**14**(10):1250. DOI: 10.3390/ijerph14101250

[63] Casey R, Chaix B, Weber C, et al. Spatial accessibility to physical activity facilities and to food outlets and overweight in French youth. International Journal of Obesity. 2012;**36**:914-919

[64] Tang S, Lee HF, Feng J. Social capital, built environment and mental health: A comparison between the local elderly people and the 'laopiao' in urban China. Ageing and Society. 2022;**42**:179-203

[65] Ding D, Adams MA, Sallis JF, Norman GJ, Hovell MF, Chambers CD, et al. Perceived neighborhood environment and physical activity in 11 countries: Do associations differ by country? International Journal of Behavioral Nutrition and Physical Activity. 2013 May 14;**10**:57. DOI: 10.1186/1479-5868-10-57

[66] Owen N, Sugiyama T, Koohsari MJ, et al. Associations of neighborhood environmental attributes with adults' objectively-assessed sedentary time: IPEN adult multi-country study. Preventive Medicine. 2018;**115**:126-133

[67] Hobin E, Leatherdale S, Manske S, et al. A multilevel examination of factors of the school environment and time spent in moderate to vigorous physical activity among a sample of secondary school students in grades 9-12 in Ontario, Canada. International Journal of Public Health. 2012;**57**:699-709

[68] Ohanyan H, Portengen L, Kaplani O, Huss A, Hoek G, Beulens JWJ, et al. Associations between the urban exposome and type 2 diabetes: Results from penalised regression by least absolute shrinkage and selection operator and random forest models.

Environment International. Dec 2022;**170**:107592. DOI: 10.1016/j.envint.2022.107592

[69] Curl A, Ward Thompson C, Aspinall P. The effectiveness of 'shared space' residential street interventions on self-reported activity levels and quality of life for older people. Landscape and Urban Planning. 2015;**139**:117-125

[70] Liu J, Zhang Y. Health status and health disparity in China: A demographic and socioeconomic perspective. China Population and Development Studies. 2019;**2**:301-322

[71] Chen L, Zhang Z, Long Y. Association between leisure-time physical activity and the built environment in China: Empirical evidence from an accelerometer and GPS-based fitness app. PLoS One. 31 Dec 2021;**16**(12):e0260570. DOI: 10.1371/journal.pone.0260570

[72] Harden SR, Schuurman N, Keller P, Lear SA. Neighborhood characteristics associated with running in metro Vancouver: A preliminary analysis. International Journal of Environmental Research and Public Health. 2 Nov 2022;**19**(21):14328. DOI: 10.3390/ijerph192114328