Data Article

Environmental data and methods from the Accumulating Data to Optimally Predict Obesity Treatment (ADOPT) core measures environmental working group

Beth A. Slotman[a], David G. Stinchcomb[a], Tiffany M. Powell-Wiley[b], Danielle M. Ostendorf[c], Brian E. Saelens[d], Amy A. Gorin[e], Shannon N. Zenk[f], David Berrigan[g,*]

[a] Westat Inc, Rockville, MD 20850, USA

[b] Division of Intramural Research, Cardiovascular Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

[c] Division of Endocrinology, Metabolism, and Diabetes, Department of Medicine, USA and Anschutz Health and Wellness Center, University of Colorado Anschutz Medical Campus, Aurora CO, USA

[d] Department of Pediatrics, Seattle Children's Research Institute, University of Washington, Seattle, WA, USA

[e] Institute for the Collaboration on Health, Intervention, and Policy, University of Connecticut, Storrs, CT, USA

[f] Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health, National Institute of Nursing Research, Bethesda, MD, USA

[g] Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

A R T I C L E   I N F O

A B S T R A C T

This article describes geospatial datasets and exemplary data across five environmental domains (walkability, socioeconomic deprivation, urbanicity, personal safety, and food outlet accessibility). The environmental domain is one of four domains (behavioral, biological, environmental and psychosocial) in which the Accumulating Data to Optimally Predict obesity Treatment (ADOPT) Core Measures Project suggested measures to help explain variation in responses to weight loss interventions. These data are intended to

* Corresponding author.
  E-mail address: berrigad@mail.nih.gov (D. Berrigan).

facilitate additional research on potential environmental moderators of responses to weight loss, physical activity, or diet related interventions. These data represent a mix of publicly and commercially available pre-existing data that were downloaded, cleaned, restructured and analyzed to create datasets at the United States (U.S.) block group and/or census tract level for the five domains. Additionally, the resource includes detailed methods for obtaining, cleaning and summarizing two datasets concerning safety and the food environment that are only available commercially. Across the five domains considered, we include component as well as derived variables for three of the five domains. There are two versions of the National Walkability Index Dataset (one based on 2013 data and one on 2019 data) consisting of 15 variables. The Neighborhood Deprivation Index dataset contains 18 variables and is based on the US Census Bureau's 5-year American Community Survey (ACS) data for 2013–2017. The urbanicity dataset contains 11 variables and is based on USDA rural-urban commuting (RUCA) codes and Census Bureau urban/rural population data from 2010. Personal safety and food outlet accessibility data were purchased through commercial vendors and are not in the public domain. Thus, only exemplary figures and detailed instructions are provided. The website housing these datasets and examples should serve as a valuable resource for researchers who wish to examine potential environmental moderators of responses to weight loss and related interventions in the U.S.

Published by Elsevier Inc.

Specifications Table

Subject	Public Health and Health Policy
Specific subject area	Potential environmental and contextual influences on the magnitude of responses to weight loss, physical activity, and diet related interventions in the U.S.
Type of data	Tables Images (maps)
How data were acquired	Public use and commercially available area level geospatial data sets were acquired by downloading from US Government websites or by purchasing data from private companies.
Data format	Raw and analyzed: Available at: https://gis.cancer.gov/research/adopt.html
Parameters for data collection	The ADOPT Core Measures Project recommended environmental measures concerning walkability, socioeconomic deprivation, urbanicity, personal safety, and food outlet accessibility be included in future research on weight loss and maintenance, physical activity, and diet related interventions. Free nationwide data or examples of commercial data were obtained and processed to facilitate analyses of environmental influences on weight loss and maintenance.
Description of data collection	These data represent a mix of publicly and commercially available pre-existing data that were downloaded, cleaned, restructured, and analyzed to create datasets at the census tract and block group level to create measures of walkability, socioeconomic deprivation, urbanicity, personal safety, and food outlet accessibility. Detailed methods are available at the website (https://gis.cancer.gov/research/adopt.html) and below.

Data source location	U.S.
	• https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/
	• https://nces.ed.gov/surveys/ruraled/definitions.asp
	• http://depts.washington.edu/uwruca/ruca-uses.php
	• https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html
Data accessibility	Repository name: U.S. National Cancer Institute. GIS Portal for Cancer Research
	Data identification number: n/a
	Direct URL to data: https://gis.cancer.gov/research/adopt.html
Related research article	B.E. Saelens, S.S. Arteaga, D. Berrigan, et al. Accumulating Data to Optimally
	Predict Obesity Treatment (ADOPT) Core Measures: Environmental Domain,
	Obesity 26 (2018) S35-S44. 10.1002/oby.22159

Value of the Data

- Obtaining and using consistent environmental data layers is important because substantial unexplained variation is found in response to weight loss, physical activity and diet related interventions. However, little is known about whether environmental factors are related to the individual variability seen in adults' intentional weight loss, maintenance, or related behavioral outcomes.
- These geospatial data layers and detailed methods will benefit researchers who wish to examine whether environmental factors are associated with individual variation in responses to weight loss and maintenance, physical activity or dietary interventions. *The intent of the ADOPT resource is to outline a path for incorporation of environmental variables that is relatively straightforward and that does not require a high level of geospatial and data science expertise.*
- We anticipate retrospective and prospective analyses of weight loss and weight maintenance, physical activity, and diet related interventions using these GIS data layers in the U.S. to better characterize contextual co-variates to explain responses to weight loss interventions.
- The ADOPT Core Measures Project, consisting of four working groups addressing key domains (behavioral, biological, environmental, and psychological), identified a set of core measures that can be analyzed across studies to better understand the variation in response to weight loss treatment. However, few details on how to obtain and process the data to operationalize the constructs were provided.
- This resource may also contribute to an improved understating of the influence of environmental factors on weight loss and maintenance, physical activity or dietary interventions.

1. Data Description

Our understanding of how environmental and contextual factors influence weight loss and maintenance, physical activity, and/or diet related interventions is limited and researchers have pointed to a need for further research in this area [1]. These datasets are intended to facilitate research around the prevention and treatment of obesity. The datasets are housed at https://gis.cancer.gov/research/adopt.html and consist of files containing data for three environmental domains thought to potentially impact an individual's response to weight loss or behavioral interventions: walkability, socioeconomic deprivation, and urbanicity. National public use data sets were not available for the other two domains: personal safety and food outlet accessibility. Therefore, data for these two domains were purchased through commercial vendors. The ADOPT resource contains exemplary figures and detailed instructions for obtaining the data, cleaning purchased data sets, and processing the resulting datasets into desired measures. Furthermore, this paper and the ADOPT resource provide references to detailed technical reports and publications containing further details of the variables and how they were generated. This material could inform researchers in other countries who wish to adopt a similar approach.

National Walkability Index data tables are presented for both the previous walkability index (2013) and a recently updated (2019) walkability index (https://www.epa.gov/smartgrowth/smart-location-mapping#walkability; Accessed 1/3/2022) in Excel and comma delimited text (CSV) file formats for block group and census tract geographies. Block group and tract are U.S. census units that typically contain between 600 and 3000 people. Census tracts contain at least one block group and contain ~1200 to ~8000 people. They vary widely in size (https://www.census.gov/programs-surveys/geography/about/glossary.html; Accessed 1/3/2022). The block group data from the Environmental Protection Agency (EPA) have also been aggregated to provide the walkability index at the census-tract level. This allows use of the same geographic unit for different environmental variables. Tract ranks and the resulting index are based on population weighted averages of block group values. There is a row for each census tract or census block group in the 50 states, the District of Columbia, and Puerto Rico with data from 2013 and 2019 and the block group and tract level given in separate files. In addition to the National Walkability Index value, we include all the variables provided in the original EPA walkability dataset including values and ranks for each of the four dimensions of walkability used to generate the original index. Tables 1 and 2 contain descriptions of the data included in the 2013 and 2019 National Walkability Index datasets respectively. Detailed information and methods for these variables are available (https://www.epa.gov/sites/default/files/2021-06/documents/epa_sld_3.0_technicaldocumentationuserguide_may2021.pdf; Accessed 1/3/2022).

Neighborhood Deprivation Index (NDI) data tables are available in Excel and comma delimited text (CSV) file formats. There is a row for each census tract in the 50 states and the District of Columbia. The dataset contains the NDI value and the original 13 socioeconomic status (SES) variables (described below) obtained from the Census Bureau's 5-year American Community Survey (ACS) for years 2013–2017 that were used to create the NDI value. Table 3 contains a description of the data included in the NDI datasets.

We also provide two tract-level urbanicity measures. Urbanicity data tables are available in both Excel and comma delimited text (CSV) file formats. There is a row for each census tract in the 50 states and the District of Columbia. The USDA rural-urban commuting area (RUCA) codes are based on data from the 2010 decennial census and the 2006–10 American Community Survey and the National Center for Education Statistics (NCES) urban/rural locale definitions were applied to 2010 Census urban/rural data. Table 4 contains a description of the data included in the urbanicity datasets.

2. Experimental Design, Materials and Methods

The National Heart Lung and Blood Institute (NHLBI)-led Accumulating Data to Optimally Predict obesity Treatment (ADOPT) Core Measures Project identified a standard set of about 50 Core Measures, or factors, that can be analyzed across studies to better understand the variation in response to obesity treatments [2]. The ADOPT Project encourages and facilitates the consistent use of these Core Measures in future clinical trials to treat obesity in adults to identify predictors of successful weight loss and maintenance for use in developing more targeted and effective interventions. Within the ADOPT Core Measures Project, the ADOPT Environmental domain subgroup recommended measuring the following GIS-based environmental constructs: walkability, socioeconomic deprivation, urbanicity, personal safety and food outlet accessibility [3]. Note this data set and associated procedures are intended to provide straightforward set of covariates that could be used without a high level of geospatial expertise. More experienced users could process these data layers in different ways, for examples, values of some variables could be calculated for buffers around respondent homes or work places using weighted averages of overlapping census delineations or distance metrics to food outlets could be extracted rather than using area averages as presented.

Table 1

Description of variables included in the walkability index datasets (2013 version) for each tract or block group.

Variable	Format	Description
TractID OR BlkGrpID	Char 11 OR Char 12	The fully qualified census tract ID based on tract assignment from the 2010 Census (including changes through 2017). Includes the state Federal Information Processing System (FIPS) code (2 chars), the county FIPS code (3 chars) and the tract ID (6 characters). The tract ID has an implied decimal before the last two characters. For example, "010,102" is referred to in Census tables and descriptions as tract 101.02. OR The fully qualified census block group ID based on tract assignments from the 2010 Census (including changes through 2017). Includes the state FIPS code (2 chars), the county FIPS code (3 chars), the tract ID (6 characters), and the block group ID (1 char).
StCoFIPS	Char 5	The state and county FIPS code which is the first 5 characters of the TractID or BlkGrpID. Useful for selecting data for a particular county or set of counties.
StAbbr	Char 2	The alphabetic state postal abbreviation. Useful for selecting data for a particular state or set of states.
NatWalkInd	Numeric	The National Walkability Index. Values range from 1 (least walkable) to 20 (most walkable).
Pop2010	Numeric	Total population from the 2010 census
HU2010	Numeric	Total number of housing units from the 2010 census
HH2010	Numeric	Total households from the 2010 census
D2A_EPHHM	Numeric	The mix of employment types and occupied housing. A block group with a diverse set of employment types (such as office, retail, and service) plus a large quantity of occupied housing units will have a relatively high value.
D2B_E8MIXA	Numeric	The mix of employment types in a block group (such as retail, office or industrial).
D3B	Numeric	Street intersection density (pedestrian-oriented intersections). This variable was calculated as a weighted sum of different intersection types with zero weight for automobile oriented intersections and lower weights for 3- vs. 4-way intersections.
D4A	Numeric	Distance from population weighted centroid to nearest transit stop (meters)
D2A_Ranked	Numeric	Rank of block groups or tracts for D2A_EPHHM within all block groups or tracts. Range from 1 to 20, higher ranks indicate greater walk trip likelihood.
D2B_Ranked	Numeric	Rank of block groups or tracts for D2B_E8MIXA within all block groups or tracts. Range from 1 to 20, higher ranks indicate greater walk trip likelihood.
D3B_Ranked	Numeric	Rank of block groups or tracts for D3B within all block groups or tracts. Range from 1 to 20, higher ranks indicate greater walk trip likelihood.
D4A_Ranked	Numeric	Rank of block groups or tracts for D4A within all block groups or tracts. Range from 1 to 20, higher ranks indicate greater walk trip likelihood.

2.1. Walkability

The walkability of a neighborhood may impact walking and potentially weight loss and maintenance [1,3]. The National Walkability Index dataset characterizes Census block groups and in this data resource we have also aggregated block level data to give the walkability index for tracts based on their relative walkability. The Walkability Index is based on data from the EPA's Smart Location Mapping project which includes a National Walkability Index for census block groups [4]. Walkability depends upon characteristics of the built environment that influence the likelihood of walking being used as a mode of travel. The index is associated with self-reported transportation walking and to a lesser extent, leisure walking [5]. Quite a few different walkability indices have been proposed over the past 2 to 3 decades [6,7] with no consensus about which is best. We selected the EPA National Walkability Index because it is freely available nationwide and shows evidence of validity [5]. A recent analysis suggests associations between two well-known measures of walkability (the EPA National Walkability Index used here and Walk Score,

Table 2

Description of variables included in the walkability index (2019 version) datasets for each tract or block group. Note that data for block groups and tracts are given in separate files.

Variable	Format	Description
TractID2019* OR BlkGrpID2019*	Char 11 OR Char 12	The fully qualified census tract ID based on the American Community Survey 5 year estimates (2014–2018). Includes the state FIPS code (2 chars), the county FIPS code (3 chars) and the tract ID (6 characters). The tract ID has an implied decimal before the last two characters. For example "010102" is referred to in Census tables and descriptions as tract 101.02. OR The fully qualified census block group ID based on the American Community Survey 5 year estimates (2014–2018). Includes the state FIPS code (2 chars), the county FIPS code (3 chars), the tract ID (6 characters), and the block group ID (1 char).
TractID2010* OR BlkGrpID2010*	Char 11 OR Char 12	The fully qualified census tract ID based on the 2010 Census. Includes the state FIPS code (2 chars), the county FIPS code (3 chars) and the tract ID (6 characters). The tract ID has an implied decimal before the last two characters. For example, "010,102" is referred to in Census tables and descriptions as tract 101.02. OR The fully qualified census block group ID based on the 2010 Census. Includes the state FIPS code (2 chars), the county FIPS code (3 chars), the tract ID (6 characters), and the block group ID (1 char).
StCoFIPS2019	Char 5	The state and county FIPS code based on the 2019 Block Group ID or Tract ID. Useful for selecting data for a particular county or set of counties.
StAbbr	Char 2	The alphabetic state postal abbreviation. Useful for selecting data for a particular state or set of states.
NatWalkInd	Numeric	The National Walkability Index. Values range from 1 (least walkable) to 20 (most walkable).
Pop2018	Numeric	Total population from the 2014 to 2018 Census American Community Survey (ACS) (5-Year Estimate)
HU2018	Numeric	Total housing units from the 2014 to 2018 Census ACS (5-Year Estimate)
HH2018	Numeric	Total households (occupied housing units) from the 2014 to 2018 Census ACS (5-Year Estimate)
D2A_EPHHM	Numeric	The mix of employment types and occupied housing. A block group with a diverse set of employment types (such as office, retail, and service) plus a large quantity of occupied housing units will have a relatively high value.
D2B_E8MIXA	Numeric	The mix of employment types in a block group (such as retail, office or industrial).
D3B	Numeric	Street intersection density (pedestrian-oriented intersections calculated in the same way as the 2013 variable (see Table 1).
D4A	Numeric	Distance from population weighted centroid to nearest transit stop (meters)
D2A_Ranked	Numeric	Rank of block groups or tracts for D2A_EPHHM within all block groups or tracts. Range from 1 to 20, higher ranks indicate greater walk trip likelihood.
D2B_Ranked	Numeric	Rank of block groups or tracts for D2B_E8MIXA within all block groups or tracts. Range from 1 to 20, higher ranks indicate greater walk trip likelihood.
D3B_Ranked	Numeric	Rank of block groups or tracts for D3B within all block groups or tracts. Range from 1 to 20, higher ranks indicate greater walk trip likelihood.
D4A_Ranked	Numeric	Rank of block groups or tracts for D4A within all block groups or tracts. Range from 1 to 20, higher ranks indicate greater walk trip likelihood.

Table 3

Variables included in the national deprivation index data set for each tract.

Variable	Format	Description
TractID	Char 11	The fully qualified census tract ID based on tract assignment from the 2010 Census (including changes through 2017). Includes the state FIPS code (2 chars), the county FIPS code (3 chars) and the tract ID (6 characters). The tract ID has an implied decimal before the last two characters. For example, "010,102" is referred to in Census tables and descriptions as tract 101.02.
StCoFIPS	Char 5	The state and county FIPS code which is the first 5 characters of the TractID. Useful for selecting data for a particular county or set of counties.
StAbbr	Char 2	The alphabetic state postal abbreviation. Useful for selecting data for a particular state or set of states.
NDI	Numeric	The Neighborhood Deprivation Index computed using data from all U.S. census tracts. Values range from −2.5 to +1.9. Higher values indicate greater neighborhood deprivation (lower socioeconomic status)
NDIQuint	Char 24	Quintiles for the Neighborhood Deprivation Index. Possible values are: "1-Least deprivation": the tract is in the first NDI quintile "2-BelowAvg deprivation": the tract is in the second NDI quintile "3-Average deprivation": the tract is in the third NDI quintile "4-AboveAvg deprivation": the tract is in the fourth NDI quintile "5-Most deprivation": the tract is in the highest NDI quintile "9-NDI not avail": the NDI value is missing for this tract
MedHHInc	Numeric	Median household income (dollars) SES dimension: wealth and income ACS table source: B19013
PctRecvIDR	Numeric	Percent of households receiving dividends, interest, or rental income SES dimension: wealth and income ACS table source: B19054
PctPubAsst	Numeric	Percent of households receiving public assistance SES dimension: wealth and income ACS table source: B19058
MedHomeVal	Numeric	Median home value (dollars) SES dimension: wealth and income ACS table source: B25077
PctMgmtBusScArti	Numeric	Percent in a management, business, science, or arts occupation SES dimension: occupation ACS table source: C24060
PctFemHeadKids	Numeric	Percent of households that are female headed with any children under 18 years SES dimension: housing conditions ACS table source: B11005
PctOwnerOcc	Numeric	Percent of housing units that are owner occupied SES dimension: housing ACS table source: DP04
PctNoPhone	Numeric	Percent of households without a telephone SES dimension: housing conditions ACS table source: DP04
PctNComPlmb	Numeric	Percent of households without complete plumbing facilities SES dimension: housing conditions ACS table source: DP04
PctEducHSPlus	Numeric	Percent with a high school degree or higher SES dimension: education ACS table source: S1501
PctEducBchPlus	Numeric	Percent with a college degree or higher SES dimension: education ACS table source: S1501
PctFamBelowPov	Numeric	Percent of families with incomes below the poverty level SES dimension: wealth and income ACS table source: S1702
PctUnempl	Numeric	Percent unemployed SES dimension: occupation ACS table source: S2301

Table 4
Variables included in the urbanicity dataset for each tract.

Variable	Format	Description
TractID	Char 11	The fully qualified census tract ID based on tract assignment from the 2010 Census (including changes through 2017). Includes the state FIPS code (2 chars), the county FIPS code (3 chars) and the tract ID (6 characters). The tract ID has an implied decimal before the last two characters. For example, "010102" is referred to in Census tables and descriptions as tract 101.02.
StCoFIPS	Char 5	The state and county FIPS code which is the first 5 characters of the TractID. Useful for selecting data for a particular county or set of counties.
StAbbr	Char 2	The alphabetic state postal abbreviation. Useful for selecting data for a particular state or set of states.
RUCA_UrbCat	Char 12	RUCA-based urbanicity category. Possible values are: "1-UrbanFocus": the tract is either in an urban center or in an area where a significant portion of the population commute to an urban center "2-RuralFocus": the tract is in a small town or rural area without significant urban commuting "9-NotCoded": the tract does not have a RUCA code assigned
RUCA_1	Numeric	The original level 1 RUCA code. Useful for creating an alternative RUCA-based categorical variable.
RUCA_2	Numeric	The original level 2 RUCA code. Useful for creating an alternative RUCA-based categorical variable.
NCES_UrbCat	Char 8	Urbanicity category using NCES urban/rural locale definitions applied to Census urban/rural population data. Possible values are: "1-City": 90% or more of the tract population is living in a large urban area and a principal city "2-Suburb": 90% or more of the tract population is living in a large urban area and not in a principal city "3-Town": 90% or more of the tract population is living in a small urban cluster "4-Rural": 90% or more of the tract population is not living in an urban area or urban cluster "5-Mixed": None of the above – the tract population is living in a mix of urbanicity types "9-NoPop": the tract had a population of zero in the 2010 Census
NCES_PctCity	Numeric	Percent of the tract population is living in a large urban area and a principal city. Useful for creating an alternative NCES-based categorical variable.
NCES_PctSuburb	Numeric	Percent of the tract population is living in a large urban area and not in a principal city. Useful for creating an alternative NCES-based categorical variable.
NCES_PctTown	Numeric	Percent of the tract population is living in a small urban cluster. Useful for creating an alternative NCES-based categorical variable.
NCES_PctRural	Numeric	Percent of the tract population is not living in an urban area or urban cluster. Useful for creating an alternative NCES-based categorical variable.

a commercially available measure) and showed that several components of walkability and the likelihood of walking were similar [8]. This further supports use of this index to explore whether or not walkability moderates responses to weight loss interventions in the US.

We have aggregated the block group data from the EPA to provide a similar walkability index at the census-tract level as described below. There are four original variables used to compute the index across three dimensions of walkability:

1. Employment type dimension (D2):
   a. Employment type and occupied housing (D2A_EPHHM)
   b. Mix of employment types (D2B_E8MIXA)
2. Connectivity dimension
   a. Street intersection density (D3)
3. Mode Choice dimension
   a. Transit accessibility (D4)

For the tract-level index, we generated averages of these four variables among all block groups in a census tract, weighted by block group populations. We then ranked the tracts from 1 to 20 for each of the four variables, where a higher number indicates greater walkability. Note, the assignment of the value of 1 to 20 does not indicate a continuous score, but rather an ordinal categorization. In this ranking, if more than 1/20th of the tracts had a particular value (zero or missing), a reduced number of ranks were assigned for the remaining tracts. Finally, the ranks were combined to calculate the tract level National Walkability Index, giving equal weight to each of the three dimensions:

$$\text{NatWalkIndex} = [(\text{D2A\_Rank}/6) + (\text{D2B\_Rank}/6)] + (\text{D3B\_Rank}/3) + (\text{D4A\_Rank}/3)$$

We confirmed with EPA staff that these equal weights by dimension were used to calculate the overall walkability index. The National Walkability Index is based on ranks of tracts and block groups at the national level.

2.2. Neighborhood deprivation

Economically disadvantaged neighborhoods have fewer and poorer quality resources for healthy diets and physical activity and living in a neighborhood with higher levels of socio-economic deprivation is associated with greater weight gain [9]. Further, disadvantaged neighborhoods are known to have poorer health outcomes with respect to obesity-related conditions such as cardiovascular disease [10,11]. Thus, individuals living in neighborhoods with lower SES may respond differently to weight loss interventions because of differential access to resources for health and because of the differing distribution of health status and health behaviors in other neighborhood residents. A Neighborhood Deprivation Index (NDI) for each Census tract in the U.S. was created using factor analysis to identify key variables from 13 measures in the following dimensions of SES: wealth and income, education, occupation, and housing conditions [12]. The specific 13 measures are:

- Wealth and income
  1. Median household income (dollars)
  2. Percent of households receiving dividends, interest, or rental income
  3. Percent of households receiving public assistance
  4. Median home value (dollars)
  5. Percent of families with incomes below the poverty level
- Education
  6. Percent with a high school degree or higher
  7. Percent with a college degree or higher
- Occupation
  8. Percent in a management, business, science, or arts occupation
  9. Percent unemployed
- Housing conditions
  10. Percent of households that are female headed with any children under 18
  11. Percent of housing units that are owner occupied
  12. Percent of households without a telephone (Includes landline, cell phone and other phone devices)
  13. Percent of households without complete plumbing facilities

These 13 variables were obtained from the Census Bureau's 5-year American Community Survey (ACS) data for 2013–2017. Factor analysis was then used to generate the NDI. This involved the following steps:

- Log transform median household income and median home value
- Reverse code percentages so that higher values represent more deprivation. For example, the percent of housing units that are owner occupied was converted to the percent of housing units that are not owner occupied.

Table 5

Categorization of RUCA codes for creation of dichotomous urbanicity variable.

Category	RUCA codes
Urban focused	1.0, 1.1, 2.0, 2.1, 3.0, 4.1, 5.1, 7.1, 8.1, and 10.1
Rural city/town focused	4.0, 4.2, 5.0, 5.2, 6.0, 6.1, 7.0, 7.2, 7.3, 7.4, 8.0, 8.2, 8.3, 8.4, 9.0, 9.1, 9.2, 10.0, 10.2, 10.3, 10.4, 10.5, and 10.6

- Z-standardize the percentages
- Run a factor analysis using Promax (oblique) rotation and a minimum Eigenvalue of 1
- Calculate the factors using only variables with a loading score > 0.4 for the first factor (this removed three variables: the percent of housing units that are owner occupied, the percent of households without a telephone, and the percent of households without complete plumbing facilities)
- Calculate Cronbach's alpha correlation coefficient among the factors and verify values are above 0.7.
- Use the resulting calculation of the first factor as the Neighborhood Deprivation Index (NDI).

2.3. Urbanicity

Saelens et al. [3] did not include a recommendation for incorporation of an urbanicity measure. However, further conversation with the expert panel and a growing interest in rural-urban differences in how urban and rural environments might differentially influence health behaviors and health outcomes led us to include two measures of urbanicity in this data resource. The two urbanicity measures are based on the latest USDA rural-urban commuting area (RUCA) codes (2010) [13] and the high level National Center for Education Statistics (NCES) urban/rural locale definitions [14] applied to 2010 Census urban/rural data [15]. The RUCA-based measure describes the types of nearby cities and towns based on commuting patterns and characterizes general access to services typically found in urban areas. The NCES-based measure, based on the NCES urban/rural locale definitions, more closely corresponds to the rural/urban nature of the immediate environment [16,17]. For the RUCA-based measure, we generated a dichotomous urbanicity variable based on the original RUCA codes using the University of Washington's "Categorization C" coding scheme that assigns various RUCA codes to Urban focused versus Rural city/town focused groupings [18]. These are shown in Table 5.

For the NCES-based measure, we calculated the percentage of the tract population in each of the four top-level NCES urban/rural locale categories:

1. City: in an urbanized area and a principal city
2. Suburb: in an urbanized area but not in a principal city
3. Town: in an urban cluster
4. Rural: not in an urbanized Area or an urban area

Urbanized area and urban cluster are defined by the Census: urban areas with populations of 50,000 or more are designated as urbanized area; those with populations between 2500 and 50,000 are designated as urban cluster. Census tracts are then assigned one of the four NCES categories if 90% of their population is in the area and tracts are assigned to a "Mixed" category otherwise. The original population percentages are provided to allow researchers to create other categorical variables if desired.

2.4. Personal safety

To look at personal safety as an environmental construct for assessing differential response to weight loss and behavioral interventions, we purchased crime data from Applied Geographic So-

lutions (AGS), one of several resources that compile and sell data from national and local sources to provide an example that illustrates tract-level crime rates. One year of data (2019) for one state (Colorado) cost approximately $900. The data set included county names, population, total crime, personal crime (murder, rape, robbery, assault) and property crime (burglary, larceny, motor vehicle theft) by census tract. One approach to incorporating these data into an analysis of a weight loss intervention is to calculate a crime level for each census tract in your dataset. Then study participants could be assigned a crime quintile and analysts can produce a map or regression model. We acknowledge that crime data in the US are widely believed to be incomplete and focus on violent and property crimes. Therefore, this variable could be considered exploratory (https://www.pewresearch.org/fact-tank/2020/11/20/facts-about-crime-in-the-u-s/ accessed 1/4/2022).

In our data, quintiles are based on total crime rate. Several tracts had the same value, so a second sort by personal crime rate was performed.

We used the following method to create population-weighted quintiles:

• Sort by total crime rate (low to high) and then personal crime rate (low to high)
• Added a column for the cumulative population. Calculate it as:
  ○ CumPop = this tract's population + the previous tract's CumPop
• Added another column for the ratio of the CumPop to one fifth of the total pop
• Added another column for the crime quintile category:
  ○ "1:Low" for tracts where the ratio is <= 1.00
  ○ "2:MedLow" for tracts where the ratio is > 1 and <= 2.0
  ○ "3:Medium" for tracts where the ratio is > 2 and <= 3.0
  ○ "4:MedHigh" for tracts where the ratio is > 3 and <= 4.0
  ○ "5:High" for tracts where the ratio is > 4 (and <= 5.0)

2.5. Food environment

Several studies have shown that greater accessibility to supermarkets is associated with healthier dietary intake and lower body weight, whereas greater accessibility to fast-food restaurants and convenience stores is associated with less healthy dietary intake and higher body weight [3]. To understand the impact of food outlet accessibility on weight loss trials, the ADOPT working group recommended measuring the density of outlets within a given area around the participant's home and distance to the closest food outlet.

We followed the process outlined in Jones, et al. [19], which describes data improvement methods for commercially obtained food outlet data, developed based on existing validation studies. Their process includes purchasing records from commercial business lists based on store/restaurant names as well as standard industrial classification (SIC) codes, reclassifying records by store type, improving geographic accuracy of records, and de-duplicating records. As recommended [19], we purchased historical listings for supermarkets and grocery stores, pharmacies, convenience stores, general merchandise stores, and liquor stores from Data Axle USA® for the year 2019 for one state (Colorado) at the cost of approximately $1900. Similarly, we purchased historical listings for the same state and time period for limited-service restaurants from Dun & Bradstreet at the cost of approximately $1600.

Availability of open access data concerning retail food environments is growing. For example, the USDA Food Environment Atlas (https://www.ers.usda.gov/data-products/food-environment-atlas/ accessed 1/4/2022) contains data concerning access and proximity to grocery stores. However, these data are only available at the county level. The related USDA Food Access Research Atlas (https://www.ers.usda.gov/data-products/food-access-research-atlas/ accessed 1/4/2022) provides tract level supermarket access data, but it lacks information on many other healthy and unhealthy food outlets. Online and map data are available concerning the food environment across much of the US, but we are unaware of a nationwide validated

compilation of such data at a finer spatial scale than county. Together these considerations led the ADOPT environmental working group to select these commercial databases for this resource.

The datasets purchased included a variety of variables including company name and address, contact information, employee size, sales volume, SIC code(s), FIPS code, Latitude, Longitude, Census block, Census tract, state and county codes, employee size, sales volume, DUNS number and others. After receiving the data, we conducted an additional batch and manual geocoding on the datasets to improve the overall geocoding of the address data. The census tract ID of the food outlet was added to each resulting dataset based on the geocodes. Data were cleaned using a de-duplication and re-classification process [19,20]. Care should be taken to inspect the resulting datasets and check for multiple listings at single addresses which may reflect actual presence of different outlets or errors in the data set. Lastly, analysis of the resulting food outlet density data may benefit from aggregating tract level data into larger groupings such as including the participant's tract, along with the neighboring tracts (i.e. tracts that are adjacent to the participant's tract), since people may shop in adjacent tracts.

To create food establishment density variables by census tract, we obtained the land area of the tract as well as the number of specific food outlet types per tract.

We obtained land area data from the Census Bureau website for Colorado counties and tracts using the following steps:

- Navigate to the Census Geography TIGER Line shapefiles web page
  https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
- Pick the year that you want – we used 2019 to match the food outlet data
- Click "Download" using the "Web Interface"
- Download the Tract zip file for desired location
- Download the County zip file (only available for the whole US)
- Unzip both files
- Save a copy of the DBF file from within each of the unzipped folders as an Excel file
- For the county file, delete the rows for states that you do not need
- Reformat as needed
- Keep just the basic IDs, the name, and the ALAND (land area in square meters) and AWATER (water area in square meters) columns.
- Add columns to calculate the land area in square kilometers and square miles
  - LandArea_SqKm = ALAND / 1000,000
  - LandArea_SqMi = ALAND / 2589,988

For each dataset food outlet type, calculate the number per tract.

- Create a frequency table for each dataset using the primary SIC code and Tract ID variable.
- Merge the total number of food outlets per tract (the combined number of all Primary SIC codes) with the land area (square miles) per tract into one file.
- Calculate density variables by taking the total number of food outlets per tract and dividing by the land area in square miles per tract.

Ethics Statements

The data contained in this paper describe area-level geospatial data and do not involve identifiable human subjects or animal experiments.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

CRediT Author Statement

Acknowledgments

References

[1] S.N. Zenk, E. Tarlov, A.N. Kraft, Environmental Modification of Adult Weight Loss, Physical Activity, and Diet Intervention Effects, in: Geospatial Approaches to Energy Balance and Breast Cancer, Springer, Cham, 2019, pp. 255–281, doi:10.1007/978-3-030-18408-7_11.

[2] P.S. MacLean, A.J. Rothman, H.L. Nicastro, S.M. Czajkowski, T. Agurs-Collins, E.L. Rice, A.P. Courcoulas, D.H. Ryan, D.H. Bessesen, C.M. Loria, The Accumulating Data to Optimally Predict Obesity Treatment (ADOPT) core measures project: rationale and approach, Obesity 26 (2018) S6–S15, doi:10.1002/oby.22154.

[3] B.E. Saelens, S.S. Arteaga, D. Berrigan, R.M. Ballard, A.A. Gorin, T.M. Powell-Wiley, C. Pratt, J. Reedy, S.N. Zenk, Accumulating Data to Optimally Predict Obesity Treatment (ADOPT) core measures: environmental domain, Obesity 26 (2018) S35–S44, doi:10.1002/oby.22159.

[4] United States Environmental Protection Agency, National walkability index, (2021). https://www.epa.gov/smartgrowth/smart-location-mapping#walkability. Accessed June 10, 2021.

[5] K.B. Watson, G.P. Whitfield, J.V. Thomas, D. Berrigan, J.E. Fulton, S.A. Carlson, Associations between the national walkability index and walking among US adults — National Health Interview Survey, 2015, Prev. Med. 137 (2020) 106122 (Baltim), doi:10.1016/j.ypmed.2020.106122.

[6] P. Maghelal, C.J. Capp, Walkability: a review of existing pedestrian indices, URISA J. 23 (2011) 5–19.

[7] L.V. Giles, M.S. Koehle, B.E. Saelens, H. Sbihi, C. Carlsten, When physical activity meets the physical environment: precision health insights from the intersection, Environ. Health Prev. Med. 26 (2021) 1–10, doi:10.1186/S12199-021-00990-W/TABLES/1.

[8] R.R. Dalmat, S.J. Mooney, P.M. Hurvitz, C. Zhou, A.V. Moudon, B.E. Saelens, Walkability measures to predict the likelihood of walking in a place: a classification and regression tree analysis, Health Place 72 (2021) 102700–102700, doi:10.1016/J.HEALTHPLACE.2021.102700.

[9] T.M. Powell-Wiley, C. Ayers, P. Agyemang, T. Leonard, D. Berrigan, R. Ballard-Barbash, M. Lian, S.R. Das, C.M. Hoehner, Neighborhood-level socioeconomic deprivation predicts weight gain in a multi-ethnic population: longitudinal data from the Dallas Heart Study, Prev. Med. 66 (2014) 22–27 (Baltim), doi:10.1016/J.YPMED.2014.05.011.

[10] A.V.D. Roux, S.S. Merkin, D. Arnett, L. Chambless, M. Massing, F.J. Nieto, P. Sorlie, M. Szklo, H.A. Tyroler, R.L. Watson, Neighborhood of residence and incidence of coronary heart disease, N. Engl. J. Med. 345 (2001) 99–106, doi:10.1056/nejm200107123450205.

[11] A.V. Diez Roux, M.S. Mujahid, J.A. Hirsch, K. Moore, L.V. Moore, The impact of neighborhoods on CV risk, Glob. Heart 11 (2016) 353–363, doi:10.1016/j.gheart.2016.08.002.

[12] A.V. Diez Roux, C. Mair, Neighborhoods and health, Ann. N. Y. Acad. Sci. 1186 (2010) 125–145, doi:10.1111/j.1749-6632.2009.05333.x.

[13] USDA Economic Research Service, Rural-urban commuting area codes, (2020). https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes. Accessed June 10, 2021.

[14] National Center for Education Statistics, Rural education in America - definitions, (n.d.). https://nces.ed.gov/surveys/ruraled/definitions.asp. Accessed June 10, 2021.

[15] U.C. Bureau, Urban and rural, (n.d.). https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html. Accessed June 14, 2021.

[16] L.G. Hart, E.H. Larson, D.M. Lishner, Rural definitions for health policy and research, Am. J. Public Health 95 (2005) 1149–1155, doi:10.2105/AJPH.2004.042432.

[17] J.L. Moss, D.G. Stinchcomb, M. Yu, Providing Higher Resolution Indicators of Rurality in the Surveillance, Epidemiology, and End Results (SEER) Database: Implications for Patient Privacy and Research, Cancer Epidemiol Biomarkers Prev. 28 (9) (2019) 1409–1416, doi:10.1158/1055-9965.EPI-19-0021.

[18] Rural Health Research Center, Rural urban commuting area codes data, (n.d.). http://depts.washington.edu/uwruca/ruca-uses.php. Accessed June 10, 2021.

[19] K.K. Jones, S.N. Zenk, E. Tarlov, L.M. Powell, S.A. Matthews, I. Horoi, A step-by-step approach to improve data quality when using commercial business lists to characterize retail food environments, BMC Res. Notes 10 (2017) 35, doi:10.1186/s13104-016-2355-1.

[20] W. Xiang, J. Kk, M. Sa, Z. Sn, K. Jones, Distance to nearest environmental setting: weight and veterans' environments study GIS protocol, (2018). http://waves.uic.edu/. Accessed January 4, 2022.