Universidad de Alcalá
COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

ACTA DE EVALUACIÓN DE LA TESIS DOCTORAL
(FOR EVALUATION OF THE ACT DOCTORAL THESIS)

Año académico (*academic year*): 2018/19

DOCTORANDO (*candidate PHD*): **CEBRECOS EGUREN, ALBA AMARANTA**
D.N.I./PASAPORTE (*Id.Passport*): ******1014**
PROGRAMA DE DOCTORADO (*Academic Committee of the Programme*): **D444-TECNOLOGÍAS DE LA INFORMACIÓN GEOGRÁFICA**
DPTO. COORDINADOR DEL PROGRAMA (*Department*): **GEOLOGÍA, GEOGRAFÍA Y MEDIO AMBIENTE**
TITULACIÓN DE DOCTOR EN (*Phd title*): **DOCTOR/A POR LA UNIVERSIDAD DE ALCALÁ**

En el día de hoy 05/06/19, reunido el tribunal de evaluación, constituido por los miembros que suscriben el presente Acta, el aspirante defendió su Tesis Doctoral **con Mención Internacional** (*In today assessment met the court, consisting of the members who signed this Act, the candidate defended his doctoral thesis with mention as International Doctorate*)**,** elaborada bajo la dirección de (*prepared under the direction of*) MANUEL FRANCO TEJERO // FRANCISCO ESCOBAR MARTINEZ.

Sobre el siguiente tema (*Title of the doctoral thesis*): *MEASURING HEART-HEALTHY URBAN ENVIRONMENTS: A GEOSPATIAL APPROACH FOR STUDYING THE CONTEXTUAL DETERMINANTS OF CARDIOVASCULAR DISEASE*

Finalizada la defensa y discusión de la tesis, el tribunal acordó otorgar la CALIFICACIÓN GLOBAL[1] de (**no apto, aprobado, notable y sobresaliente**) (*After the defense and defense of the thesis, the court agreed to grant the GLOBAL RATING (fail, pass, good and excellent)*): SOBRESALIENTE

Alcalá de Henares, a 5 de Junio de 2019

Fdo. (*Signed*): Francisco Bolumar Fdo. (*Signed*): Hanna Badland Fdo. (*Signed*): Diana Gomez

FIRMA DEL ALUMNO (candidate's signature),

Fdo. (*Signed*): Alba Cebrecos

Con fecha 24 de junio de 2019 la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado, a la vista de los votos emitidos de manera anónima por el tribunal que ha juzgado la tesis, resuelve:

- [x] Conceder la Mención de "Cum Laude"
- [ ] No conceder la Mención de "Cum Laude"

La Secretaria de la Comisión Delegada

[1] La calificación podrá ser "**no apto**" "**aprobado**" "**notable**" y "**sobresaliente**". El tribunal podrá otorgar la mención de "cum laude" si la calificación global es de sobresaliente y se emite en tal sentido el voto secreto positivo por unanimidad. (*The grade may be "**fail**" "**pass**" "**good**" or "**excellent**". The panel may confer the distinction of "cum laude" if the overall grade is "Excellent" and has been awarded unanimously as such after secret voting.*).

INCIDENCIAS / OBSERVACIONES:
(Incidents / Comments)

En aplicación del art. 14.7 del RD. 99/2011 y el art. 14 del Reglamento de Elaboración, Autorización y Defensa de la Tesis Doctoral, la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado y Doctorado, en sesión pública de fecha 24 de junio, procedió al escrutinio de los votos emitidos por los miembros del tribunal de la tesis defendida por ***CEBRECOS EGUREN, ALBA AMARANTA***, el día 5 de junio de 2019, titulada, *MEASURING HEART-HEALTHY URBAN ENVIRONMENTS: A GEOSPATIAL APPROACH FOR STUDYING THE CONTEXTUAL DETERMINANTS OF CARDIOVASCULAR DISEASE* para determinar, si a la misma, se le concede la mención "cum laude", arrojando como resultado el voto favorable de todos los miembros del tribunal.

Por lo tanto, la Comisión de Estudios Oficiales de Posgrado y Doctorado **resuelve otorgar** a dicha tesis la

MENCIÓN "CUM LAUDE"

Alcalá de Henares, 24 de junio de 2019
EL VICERRECTOR DE INVESTIGACIÓN Y TRANSFERENCIA
F. Javier de la Mata de la Mata

Copia por e-mail a:
Doctorando: CEBRECOS EGUREN, ALBA AMARANTA
Secretario del Tribunal: DIANA GOMEZ BARROSO
Directores de Tesis: MANUEL FRANCO TEJERO // FRANCISCO ESCOBAR MARTINEZ

Universidad de Alcalá

ESCUELA DE DOCTORADO
Servicio de Estudios Oficiales de Posgrado

DILIGENCIA DE DEPÓSITO DE TESIS.

Comprobado que el expediente académico de D./Dª ALBA AMARANTA CEBRECOS EGUREN reúne los requisitos exigidos para la presentación de la Tesis, de acuerdo a la normativa vigente, y habiendo presentado la misma en formato: ☒ soporte electrónico ☐ impreso en papel, para el depósito de la misma, en el Servicio de Estudios Oficiales de Posgrado, con el nº de páginas: 156 se procede, con fecha de hoy a registrar el depósito de la tesis.

Alcalá de Henares a 25 de MARZO de 2019

PILAR DE LA VEGA

Fdo. El Funcionario

Programa de Doctorado en
Tecnologías de la Información Geográfica

Measuring heart-healthy urban environments: a geospatial approach for studying the contextual determinants of cardiovascular disease

Tesis Doctoral presentada por

ALBA CEBRECOS

2019

D. Francisco Escobar Martínez, Coordinador de la Comisión Académica del Programa de Doctorado en Tecnologías de la Información Geográfica

HAGO CONSTAR que la Tesis Doctoral titulada *Measuring heart-healthy urban environments: a geospatial approach for studying the contextual determinants of cardiovascular disease*, presentada por Dª Alba Cebrecos Eguren, bajo la dirección del Dr. Manuel Franco Tejero y del Dr. Francisco Escobar Martínez, ha sido realizada por compendio de artículos, reuniendo los requisitos exigidos a este tipo de tesis, así como los requisitos científicos de originalidad y rigor metodológicos para ser defendida ante un tribunal. Esta Comisión ha tenido también en cuenta la evaluación positiva anual del doctorando, habiendo obtenido las correspondientes competencias establecidas en el Programa.

Para que así conste a los efectos del depósito de la tesis, se firma en Alcalá de Henares a 15 de marzo de 2019

Fdo.: Francisco Escobar Martínez

Universidad de Alcalá

El Dr. Francisco Escobar Martínez y el Dr. Manuel Franco Tejero, ambos profesores titulares de la Universidad de Alcalá, en el Departamento de Geología, Geografía y Medio Ambiente y en el Departamento de Cirugía, Ciencias Médicas y Sociales respectivamente, y directores de la Tesis Doctoral *Measuring heart-healthy urban environments: a geospatial approach for studying the contextual determinants of cardiovascular disease* desarrollada por Alba Cebrecos Eguren,

HACEN CONSTAR

que el trabajo presentado reúne los requisitos científicos exigidos de originalidad y rigor metodológicos suficientes para constituir una Tesis Doctoral con Mención Internacional en la Universidad de Alcalá.

Por tanto, consideramos esta tesis merecedora de nuestro VISTO BUENO para proceder a su lectura y defensa pública ante el correspondiente tribunal.

Alcalá de Henares, a 15 de marzo de 2018

Director de la Investigación
Dr. Francisco Escobar

Director de la Investigación
Dr. Manuel Franco

Programa de Doctorado en
Tecnologías de la Información Geográfica

Measuring heart-healthy urban environments: a geospatial approach for studying the contextual determinants of cardiovascular disease

Tesis Doctoral con mención internacional presentada por

ALBA CEBRECOS

Dirigida por:

Dr. Manuel Franco

Dr. Francisco Escobar

2019

"GIS works best when the computer and the brain combine forces, and when GIS is used to augment human intuition by manipulating and displaying data in ways that reveal things that would otherwise be invisible"

Dr. Michael F. Goodchild

"The application of GIS is limited only by the imagination of those who use it"

Jack Dangermond. Esri founder

Prólogo

Como tantos otros, nunca pensé que desarrollaría una carrera investigadora cuando comencé a cursar Ciencias Ambientales en la Universidad de Alcalá. Supongo que fue en segundo curso cuando el análisis espacial captó mi atención durante la asignatura de Tecnologías de la Información Geográfica. Sin embargo, también considero que fue el azar quien quiso que en el último año de carrera me tocase desarrollar el trabajo "*Planificación y gestión del acceso a la naturaleza en ciudades de tamaño medio*", tutorizado por María Jesús Salado y José Vicente Lucio. Durante ese periodo, y con ayuda de mis tutores, aumenté mis conocimientos y capacidades sobre los Sistemas de Información Geográfica (SIG). Pero lo curioso es que sufrí una especie de epifanía. De repente, fui consciente del alcance de estas tecnologías ¿Sabía el resto del mundo el potencial que podían llegar a tener? Parecía que se abriese ante mí un mundo de posibilidades, por lo que decidí seguir formándome profesionalmente.

Mientras preparaba el máster de Tecnologías de la Información Geográfica conocí al médico John Snow (1813-1858). John sostenía que la epidemia de cólera que asoló Londres a mediados del s. XIX se trasmitía mediante la ingestión de una "materia mórbida" que producía un síndrome diarreico con deshidratación severa. Esta "materia mórbida" se eliminaba a través de las deposiciones que finalmente terminaban en las aguas del Támesis. El círculo de contagio se cerraba cuando la gente extraía para consumo el agua contaminada del río. Para demostrar su hipótesis, durante un brote en 1854 en Golden Square, pintó sobre un mapa de la zona los puntos correspondientes a las muertes por cólera y las distintas bombas de agua de la zona. De esta forma demostró gráficamente la relación espacial entre las defunciones y la bomba de Broad Street. Cuando levantaron la bomba se dieron cuenta que se producían filtraciones de una tubería de alcantarillado próxima.

Así fue como John Snow desarrolló una de las primeras cartografías analíticas, la cual podría considerarse como una especie de proto-SIG.

Me encantó esta historia ya que, en cierto modo, me pareció mágica. El mapa descifró una realidad que hasta el momento no existía, o era desconocida. Me parece un bonito ejemplo de aplicación de los SIG, sobre todo en un tema como la salud que, a simple vista, puede parecer bastante alejado del ámbito de los SIG.

Cerca del final del máster, conocí al equipo de Epidemiología Social y Cardiovascular y su proyecto de Barrios CardioSaludables (o HHH por sus siglas en inglés). Un equipo joven y multidisciplinar, cuyo interés era conocer como la ciudad de Madrid podría influir en la salud cardiovascular de sus residentes. Me pareció sumamente interesante el equipo y su investigación, así que comencé a leer sobre epidemiología. Y para mi sorpresa, volvió a aparecer Jonh Snow. Porque Jonh, además de ser considerado el primero en aplicar un SIG, también es considerado uno de los padres de la epidemiología moderna. En ese momento me pareció una verdadera carambola. Así que cuando se me presentó la oportunidad de desarrollar un doctorado dentro del HHH, no lo pensé dos veces. Y aunque no fuese de origen vocacional, decidí jugármela con Fortuna. Y a día de hoy, ha sido una de las mejores decisiones de mi vida que culmina con este documento.

Agradecimientos

Los agradecimientos son sin duda un gesto pequeño comparado con todo lo que he recibido durante estos cinco años. Es demasiado tiempo para poder acordarme de todas las personas a las que no di las gracias y han influido, directa o indirectamente, en este documento. Pero como la gratitud en silencio no sirve de nada y no hay lugar más oportuno que este, quiero deciros que gracias, muchas gracias a todos.

Venturoso aquel a quien el cielo dio un pedazo de pan sin que le quede obligación de agradecérselo a otro que al mismo cielo. (Miguel De Cervantes)

En el ámbito profesional y académico tengo que dar las gracias principalmente a mis directores, el Doctor Manuel Franco y el Doctor Francisco (Patxi) Escobar. Gracias a ellos, esta tesis de Geografía de la Salud ha sido posible. A Manuel por ver en mi un "je ne sais quoi" que le llevó a ofrecerme la oportunidad de desarrollar mi carrera en el equipo tan extraordinario que ha montado. Su entusiasmo por la investigación es contagiosa. Pero sobre todo, gracias por hacerme ser menos cartesiana y enseñarme que uno de los objetivos finales de la ciencia es mejorar el bienestar de las personas.

A Patxi, por poder decir con orgullo que soy su pupila. Espero haber sabido reflejar en este trabajo su orientación firme por el hacer bien de las cosas; profesional, ética y estéticamente. Gracias por la crítica constructiva acompañada de un consejo. Aunque él no lo recuerde, lleva enseñándome geografía desde hace 10 años. Primero en la carrera de Ciencias Ambientales, pasando por el máster en TIG y finalmente como director de mi tesis. Demasiado tiempo para que mis agradecimientos sean solo profesionales o académicos. Que sienta orgullo por mí y por esta tesis es uno de mis principales propósitos.

A todo el equipo del proyecto Heart Healthy Hoods y el grupo EpiSoc, pero especialmente a Julia, Xisca y Pedro. Estos jóvenes investigadores serán los próximos referentes de sus respectivos campos, y esto se deberá a su excelencia, su constancia y su profundo compromiso social. Mi reconocimiento, admiración y gratitud por ellos es infinita. Me han enseñado desde conceptos básicos de epidemiología y salud pública, hasta cómo gestionar un artículo, un revisor o un director de tesis. Agradezco también su absoluta disponibilidad para visitar a Jordi cuando la semana había sido intensa, y cuando no lo había sido también. Al resto del equipo, Paloma, María, Angélica, Luis, Rober, Andrea, Maca y Natalia, todos piezas imprescindibles que hacen que el proyecto siga funcionando, eso sí, a nuestra manera.

Y a Usama, siempre a miles de kilómetros pero siempre dispuesto a poner su cerebro en pro de tu investigación.

To Olivier Klein PhD, for hosting me during my stay at the Luxembourg Institute of Socio-Economic Research (o LISER). I feel very lucky to have spent 3 months in a prestigious institute such as the LISER and to be able to do it with someone so generous as he is. His door always remained open for any question and was in this period that I began to write the last paper of this dissertation. Thanks for everything Olivier.

To all the researchers who have gone through the HHH, from a variety of different places like Philadelphia, London, Australia, New York, Madrid or Baltimore. Thanks to all of them. They have left in the team and in me great pills of knowledge that we will not forget. But I would like to make a special mention to PhD Luisa Borrell and PhD Hannah Badland, true references for me. Outstanding women in their respective fields, who stimulate to continue developing a young research career. But it is not only their academic prestige worthy of admiration. Their close and humble treatment, as well as their personal interest in PhD students like me, makes them great women to whom I am very grateful.

Al proyecto MEDEA por permitirme trabajar con ellos y formar parte de un equipo coordinado con experiencia del que sigo aprendiendo tanto.

A los directores de mi trabajo fin de carrera, María Jesús Salado y Vicente de Lucio. Porque siempre he tenido la sensación de que con ellos comenzó todo.

Pero si a alguien tengo que darles las gracias es a mi familia. Su educación basada en el esfuerzo y en el trabajo hay que tener por seguro que han influenciado en el logro de esta tesis. Podría decirse que, en esa casa, siempre ha primado el "*si vas a hacer algo, hazlo bien*". Mis padres, Ángel y Maite, son los pilares de lo que soy. Les doy las gracias por obligarme a cuestionarme el porqué de las cosas, a ser crítica y a entender que si quiero algo, tengo que trabajar dando lo mejor de mí para conseguirlo. A mi hermana Andrea, por preocuparse siempre por mí. Su fuerza de voluntad y constancia me ha inspirado a lo largo de mi vida. Además durante este periplo, ella y Monar me han hecho el mejor de los regalos, Bruna. Tampoco quiero olvidarme de mi tía Aurora, siempre atenta por cómo estoy.

Por otro lado también quiero darle las gracias al grupo Fuckadvisor. Borja, Javi, Pablo, Blito, Julia y Xisca, sois lo más divertido que me ha podido dar Madrid. Mi familia de aquí. Que siempre sigamos encontrando las excusas más absurdas para vernos -y comer-.

A mi tierra, Santander la marinera, sigo echándote de menos. Y a sus Tierrucos, que siempre serán mi hogar. Aunque a veces han obstaculizado mis intenciones de trabajar, son un balón de oxígeno necesario para volver con las pilas cargadas y la cabeza despejada. Tampoco puedo olvidarme de otro de mis lugares, Montejo de San Miguel, mi paraíso Burgalés.

Y a mi familia zaragozana. Su preocupación, su cariño y sus tuppers hacen que los tiempos de estrés lo sean menos. Y por supuesto a su hijo Josi. Reafirmo cada día mi amor, orgullo y admiración por ti. Tu comprensión y paciencia durante este periodo me ha enseñado que eres la fortaleza de mi vida. Por eso, por todo lo que me das y por tu amor incondicional, muchas pero muchas gracias. Este va a ser el mejor año de nuestras vidas.

Table of Contents

ABSTRACT

Introduction

Currently, 80.1% of the Spanish population lives in cities and it is estimated that by 2050 this percentage will increase to 84.6%. With this projection, urban environments are contexts with a relevant role on populations' health by having physical and social characteristics (contextual determinants) that may limit or promote specific health related behaviors. Geospatial approaches allow improving the knowledge of these determinants when considering the geographical properties of data and that may impact health results. These approaches will help to improve the understanding of the mechanisms by which contextual urban environments affect health.

Objective

The main objective of this PhD dissertation is to develop a GIS-based methodological proposal to characterize urban contextual determinants of cardiovascular disease (CVD). In order to achieve this objective, three secondary objectives were proposed and developed in three studies.

Study 1

1.1 To design a multi-component tool based on Geographic Information System (GIS) integrating information from food and physical activity environments to characterize the obesogenic environment of urban areas.

Study 2

2.1 To build a socioeconomic deprivation index and to study its statistical and geographic stability through three spatial scales (census section, neighborhoods and districts).

2.2 To study the scale effect on the relationship between socioeconomic deprivation and prevalence of CVD.

Study 3

3.1 To develop and implement the Heart Healthy Hoods (HHH) index integrating characteristics of heart-healthy urban environments (food, physical activity, tobacco, and alcohol) for small areas.

3.2 To examine the association of the HHH index with the prevalence of CVD at the area level.

Methods

Study 1 comprised an urban area of 12 contiguous census sections (≈16,000 residents) representing the median socio-demographic profile of the city of Madrid. Through systematic social observation, the food environment was evaluated by auditing all food stores. To assess the physical activity environment, all street segments were audited. A geospatial analysis was performed to integrate information from both urban environments, allowing to characterize the obesogenic environment. Using the Kernel Density Estimation (KDE), point information (food stores) and linear information (street segments) became two continuous mathematical surfaces. This previous step allowed integrating both surfaces in a single one using map algebra. Finally, a zonal analysis was performed to obtain a unique output value for each census section.

In study 2, a deprivation index was calculated with the data from the 2011 census for the municipality of Madrid. Through a Principal Components Analysis (PCA), an index of socioeconomic deprivation was calculated in three analysis scales (census section, neighborhood, and district). Spatial autocorrelation analyses (global and local) and cartographic representations were performed to evaluate the stability of the index through the three spatial scales. In addition, a correlation analysis was run to study the change in the relationship between deprivation and CVD prevalence at the three spatial scales.

Study 3 was also conducted in the municipality of Madrid. Using a multicomponent method based on GIS, two index models were generated (model 0 unweighted and model 1 weighted) using the percentage of deaths for the main CVD behavioral risk factors (diet, physical activity, alcohol and tobacco environments). We performed a global regression analysis (ordinary least squares -OLS-) and a local regression (Geographically Weighted Regression -GWR-) to evaluate the relationship between both index models and the prevalence of CVD, to finally identify the best index model.

Results

In study 1, a continuous synthetic index was obtained for the study area with a range from 0 to 100 where high scores indicated a healthier environment. Results showed a heterogeneous spatial distribution of the obesogenic contextual determinants. In the study area an average score of 25.5 was obtained. About 75% of the area had scores below 36.8 and 50% scores below 25.5. The study area was also characterized in four categories according to the average score of each census section: low (17.7-21.6), medium-low (21.7-30.8), medium-high (30.9-35.1) and high (35.2- 43.8). Four of the 12 census sections were classified as low, four medium-low, three medium-high and one as high.

Study 2 addressed the implications that the Modifiable Areal Unit Problem (MAUP) has on the unit of analysis and the scale of choice when the data of socioeconomic deprivation are aggregated for different areal units. For different administrative units, the index showed different results and correlations with CVD prevalence. Varying the geographical unit in the calculation of the deprivation index, the indicators integrating the index also varied. At the census section and neighborhood level, indicators were related to occupation/labor market and education domains. At the district level, two immigration-related indicators were part of the first component of the ACP. The index presented significant correlations with the prevalence of CVD for all scales, being greater as the size of the administrative unit increased.

In the third study, the weighted model for the HHH index showed a better fit for the two regression analyses. GWR had a better fit and evidenced a non-stationary relationship between the heart-healthy environment and CVD prevalence. The weighted model best captured the spatial variation of environments with connected heart-healthy areas in the city core and the worst values in the less populated areas around Madrid. The areas with inverse relationships showed that improving the cardiovascular environment would contribute to reducing the prevalence rates of CVD. However, we also identified areas with a positive relationship between the models of the HHH index and the prevalence rates of CVD.

Conclusions

This doctoral research has proposed a new research line focusing on the interwoven nature of multidimensional characteristics for the built environment in relation to CVD. In addition, the impact that geographic properties of data have on health outcomes has been evidenced. The spatial dependence, the MAUP, the scale effect, and the spatial heterogeneity have been considered transversally to the evaluation of the urban context and its relationship with CVD.

Therefore, it is possible to appreciate a double contribution of this doctoral research: 1) it considers the necessary geographic conceptualization in ecological studies; and 2) it applies it to a very relevant global health problem as CVD. This geospatial approach is essential to better understand the role of the place in health promotion and in the reduction of health inequalities.

RESUMEN

Introducción

Actualmente en España el 80,1% de la población vive en ciudades y se estima que para 2050 este porcentaje aumentará a un 84,6%. Con esta proyección, los entornos urbanos se presentan como contextos con un papel relevante en la salud de las poblaciones al poseer características físicas y sociales (determinantes contextuales) que limitan o promueven comportamientos saludables. El enfoque geoespacial permite mejorar el conocimiento de estos determinantes al considerar las propiedades geográficas que tienen los datos y que impactan sobre los resultados de salud en los estudio ecológico. Dicho enfoque ayudará a mejorar la comprensión del mecanismo por el cual los determinantes contextuales de las ciudades afectan a la salud.

Objetivos

El objetivo principal de la presente investigación de doctorado es desarrollar una propuesta metodológica basada en SIG para caracterizar los determinantes contextuales urbanos de la enfermedad cardiovascular (ECV). Para lograr este objetivo, se plantearon objetivos secundarios que fueron desarrollados en tres estudios.

Estudio 1.

1.1 Diseñar una herramienta multicomponente basada en Sistemas de Información Geográfica (SIG) que integre información de los entornos de actividad física y de alimentación para caracterizar entorno obesogénico de un área urbana.

Estudio 2.

2.1 Construir un índice de privación socioeconómica y estudiar su estabilidad estadística y geográfica a través de tres escalas espaciales (sección censal, barrio y distrito).

2.2 Estudiar el efecto de la escala en la relación entre la privación socioeconómica y la prevalencia de ECV.

Estudio 3.

3.1 Desarrollar e implementar el índice HHH integrando características de los entornos urbanos cardiosaludables (alimentos, actividad física, tabaco y alcohol) para áreas pequeña.

3.2 Examinar la asociación del índice de HHH con la prevalencia de ECV a nivel de área.

Métodos

El estudio 1 comprendía un área urbana compuesta por 12 secciones censales contiguas (≈16,000 residentes) que representaba el perfil sociodemográfico mediano de la ciudad de Madrid. Mediante observación social sistemática, se evaluó el entorno alimentario auditando todas las tiendas de alimentos. Para evaluar el entorno de actividad física, se auditaron todos los segmentos de calle. El análisis geoespacial se utilizó para integrar la información de ambos entornos lo que permitió caracterizar el entorno obesogénico. Usando la Estimación de Densidad de Kernel (EDK), la información puntual (tiendas de alimentos) y lineal (segmentos de calle) se convirtió en dos superficies matemáticas continuas. Este paso previo permitió integrar ambas superficies en una sola mediante álgebra de mapas. Finalmente, se ejecutó un análisis zonal para obtener un valor de salida único de cada sección censal.

En el estudio 2 se calculó un índice de privación con los datos del censo de 2011 para el municipio de Madrid. Mediante un Análisis de Componentes Principales (ACP), se calculó un índice de privación socioeconómica en tres escalas de análisis (sección censal, barrio y distrito). Se realizaron análisis de autocorrelación espacial (global y local) y representaciones cartográficas para evaluar la estabilidad del índice a través de las tres escalas espaciales. Realizamos un análisis de correlación para estudiar el cambio en la relación entre la privación y la prevalencia de ECV en las tres escalas espaciales.

El estudio 3 también se realizó en el municipio madrileño. Usando un método multicomponente basado en SIG, se generaron dos modelos de índice (modelo 0 no ponderado y modelo 1 ponderado) utilizando el porcentaje de muertes de los principales factores de riesgo comportamentales de ECV (entornos de dieta, actividad física, alcohol y tabaco). Realizamos un análisis de regresión global (mínimos cuadrados –OLS-) y otro local (regresión ponderada geográficamente –GWR-) para evaluar la relación entre ambos modelos de índice y la prevalencia de ECV, e identificar el mejor modelo de índice.

Resultados

En el estudio 1 se obtuvo un índice sintético continúo para el área de estudio con un rango de 0 a 100 donde las puntuaciones altas indicaban un entorno más saludable. Los resultados mostraron una distribución espacial heterogénea de los determinantes contextuales obesogénicos. El área de estudio presentó una puntuación media de 25,5. Alrededor del 75% del área tenía puntuaciones por debajo de 36.8 y el 50% puntuaciones por debajo de 25.5. También se caracterizó el área de estudio en cuatro categorías de acuerdo con el puntaje promedio de cada sección censal: baja (17.7–21.6), media-baja (21.7–30.8), media-alta (30.9–35.1) y alta (35.2–43.8). Cuatro de las 12 secciones del censo están clasificadas como bajas, cuatro como medias-bajas, tres como medias-altas y una como alta.

El estudio 2 abordó las implicaciones que el Problema de la Unidad Espacial Modificable (PUEM) tiene en la elección de la unidad de análisis y la escala cuando los datos de privación socioeconómica se agregan a diferentes unidades de área. Para las diferentes unidades administrativas, el índice produjo diferentes resultados y correlaciones con la prevalencia de ECV. Al variar la unidad geográfica en el cálculo del índice de privación, los indicadores que lo componen también variaron. A nivel de sección censal y de barrio, los indicadores se relacionaron con los dominios de ocupación/mercado laboral y educación. A nivel de distrito, dos indicadores relacionados con la inmigración ingresaron al primer componente del ACP. El índice presentó unas correlaciones significativas con la prevalencia de ECV para todas las escalas, siendo mayor a medida que aumentaba el tamaño de la unidad administrativa.

En el tercer estudio, el modelo ponderado para el índice HHH presentó un mejor ajuste en los dos análisis de regresión. El GWR mostró un mejor ajuste y reveló una relación no estacionaria entre el entorno cardiosaludable y la prevalencia de ECV. El modelo ponderado capturó mejor la variación espacial de los entornos con áreas cardiosaludables conectadas en el núcleo de la ciudad y los peores valores en las áreas menos pobladas de los alrededores de Madrid. Las áreas con relaciones inversas mostraron que mejorar el entorno cardiovascular contribuiría a reducir las tasas de prevalencia de ECV. Sin embargo, también se identificaron áreas con una relación positiva entre los modelos del índice HHH y las tasas de prevalencia de ECV.

Conclusiones

La presente investigación doctoral ha planteado una línea de investigación novedosa centrada en la naturaleza entrelazada de las características multidimensionales del entorno construido relacionado con la ECV. Además, se ha demostrado el impacto que las propiedades geográficas de los datos tienen sobre los resultados de salud. La dependencia espacial, el MAUP, el efecto de escala y la heterogeneidad espacial se han evaluado transversalmente a la evaluación del contexto urbano y su relación con la ECV. Por lo tanto, es posible apreciar una doble contribución de esta investigación: 1) considera la conceptualización geográfica necesaria en estudios ecológicos1; y 2) la aplica a un problema de salud global como la ECV. Este enfoque es crucial para comprender mejor el papel del lugar en la promoción de la salud y la reducción de las desigualdades en la salud.

SECTION I.
INTRODUCTION

1. HEALTH GEOGRAPHY

Health Geography is a subdiscipline of Human Geography. Health Geography focuses on the interaction between people and the environment (Dummer, 2008). Therefore, the association between social and physical environments and health is an area of great interest in Health Geography. The Department of Geography at the University of Alcalá, in which this dissertation has been produced, holds a long history of interest in health research under the framework of Geographic Information Technologies (GIT) (Escobar *et al.*, 2004; Aránguez *et al.*, 2005; Eagleson *et al.*, 2005; Laffly and Handschumacher, 2005). GIT are a special class of information technologies which keep track not only of events, activities and things, but also of *where* these events, activities, and things happen or exist (Longley *et al.*, 2005). Later on we will define in-depth GIT, precisely the Geographic Information Systems (GIS), their functions and applications for Public Health research.

One of the strengths of considering health as a place-based phenomenon is to move beyond a classical biomedical interpretation of the disease (interaction between a causative agent and patient); by thinking about the broader context in which the condition has arisen (Anthamatten and Hazen, 2011). When thinking spatially, geographers emphasize the distinction between *space*, which deals with locating where things are; and *place*, which refers to the cultural meaning of a particular environment (Anthamatten and Hazen, 2011). *Place* refers to both a unique location and to the sense of place associated with it, referring to its social meaning and values imbued by individuals and communities (Curtis, 2004). Both aspects of geography inform the work of health geographers.

Greek philosophers were among the first to interrelate health and place. Some scholars argued that geographical differences were responsible for patterns in disease and that the relationship of people with their environment could be manipulated to influence health (Anthamatten and Hazen, 2011). The seminal text of this perspective is included in the Hippocratic Corpus (Macintyre and Ellaway, 2003; Anthamatten and Hazen, 2011; Brown *et al.*, 2018). This work discussed medical ethics, holistic medicine, and environmental health influences. One of the first mentions of *place* within a health context is to be found in the section entitled *In the rivers, waters and places,* which in simple terms, states that health and diseases are a product of their environment (Macintyre and Ellaway, 2003; Brown *et al.*, 2018).

Since the early 1990s, this social and cultural theory has been progressively adopted in Health Geography studies (Brown *et al.*, 2018). The discipline adopts a holistic perspective encompassing context and society, as well as conceptualizing the role of place and location in well-being and illness (Dummer, 2008). Concepts such as politics, economics, race, gender or underserved populations have become increasingly important in these "post-medical" approaches to health and medical care (Anthamatten and Hazen, 2011; Brown *et al.*, 2018). These approaches reject the positivist and reductionist points of views focused on explaining health and disease simply as the presence or absence of a pathogen. Instead, they consider the broader role of the social and psychological life of a person (Anthamatten and Hazen, 2011). *Place* in this context includes not only the landscape attributes or the built environment but also how the people who live in it interact and respond to their environment.

Currently, Health Geography seeks to explore the social, cultural and political contexts of health within a spatial organization frame (Dummer, 2008). This discipline is closely aligned with epidemiology. However, Health Geography focuses mainly on relationships and spatial patterns; while epidemiology focuses on the study of the distribution of health determinants within populations. In addition, general definitions reveal that the concept of "distribution" within the epidemiological field is not used from a strictly geographic point of view (Krieger, 2011). Health Geography studies cover several lines of research, some of them closely related to epidemiology, and support the development of public health policy (Table 1).

Another important facet of Health Geography comprises exploring and questioning daily practices and their complex interrelationships with both space and places which we co-inhabit. Crucial to our understanding of the every is not only that we account for those processes that (materially) structure people's experiences, but that we also recognize that these experiences depend on the spaces and times in which people live (Brown *et al.*, 2018). In current societies, many human activities must be performed in spaces that are functionally segregated and distant from one another during established periods of time (work schedules, access to services)(Lenntorp, 1976, 1999, Hägerstrand, 1985, 1989) . People go through different urban contexts exposing themselves to different environments during their daily activities. These movements depend on the identities that people assume and those that are attributed to them socially, either for reasons of race/ethnicity, class condition, sex or sexual orientation, capacity/disability, etc. ; being important to differentiate these experiences related to health (Anthamatten and Hazen, 2011). The study of individual spatio-temporal patterns is key to understand the exposure of different social groups to urban environments and their influence on health. There is a whole field of research on space-time mobility and individual perception of urban environments (Hägerstrand, 1989; Kwan, 2012; Perchoux *et al.*, 2013; Horner and Wood, 2014; Chen and Kwan, 2015), however, these works are beyond the scope of this PhD dissertation.

Table 1: Examples of Health Geography research relevant to public health policy. Based on Dummer, 2008.

Research area	**Examples**
Services, infrastructures and land-use planning	Regulating the local availability of tobacco retailing in Madrid, Spain: a GIS study to evaluate compliance (Valiente *et al.*, 2018) Availability of healthy foods and dietary patterns: The Multi-Ethnic study of atherosclerosis (Franco *et al.*, 2009)
Disease surveillance, modelling and mapping	Census tract socioeconomic and physical environment and cardiovascular mortality in the Region of Madrid (Spain) (Domínguez-Berjón *et al.*, 2010) Geographic analysis of motor neuron disease mortality and heavy metals released to rivers in Spain (Sánchez-Díaz *et al.*, 2018)
Disease etiology and determinants of health	Neighborhood social and economic change and retail food environment change in Madrid (Spain): The heart healthy hoods study (Bilal *et al.*, 2018) Associations of alcohol availability and neighborhood socioeconomic characteristics with drinking: Cross-sectional results from the Multi-Ethnic Study of Atherosclerosis (MESA) (Brenner *et al.*, 2015)
Environmental health risk factors assessment	Study of non-Hodgkin's lymphoma mortality associated with industrial pollution in Spain, using Poisson models (Ramis *et al.*, 2009) Mapping environmental injustices: Pitfalls and potential of geographic information systems in assessing environmental health and equity (Maantay, 2002)
Health service use	Poverty and access to health care in developing countries (Peters *et al.*, 2008)
Inequalities in health outcomes	Socioeconomic inequalities in mortality in 16 European cities (Borrell *et al.*, 2014) Increasing mortality differentials by residential area level of poverty: Britain 1981–1997 (Shaw *et al.*, 2000)
Therapeutic and healthy landscapes	Urban residential environments and senior citizens longevity in megacity areas: The importance of walkable green spaces (Takano, Nakamura and Watanabe, 2002)

1.1 SPATIAL INEQUALITIES OF DISEASE

Spatial differences in mortality, morbidity and health behaviors between neighborhoods, regions, and countries have been studied for decades (Macintyre and Ellaway, 2003; Mackenbach *et al.*, 2003). It has been demonstrated that vulnerable and socially disadvantaged populations have a higher prevalence of unhealthy behaviors and a lower life expectancy than those with higher social status (Marmot *et al.*, 1997; Marmot, 2005). Principally, this is due to an increased exposure to harmful products, such as contaminated water, tobacco, unhealthy foods, or limited access to health services (Marmot, 2006). Disturbingly, recent evidence has shown that these differences have not been reduced but in turn, have increased (Pearce *et al.*, 2010; Richardson *et al.*, 2014; Chetty *et al.*, 2016; Stanaway *et al.*, 2018). According to estimates, in 2040 Japan, Singapore, Spain, and Switzerland will have a life expectancy of more than 85 years (Foreman *et al.*, 2018). At the same time, Central African Republic, Lesotho, Somalia, and Zimbabwe will have a life expectancy lower than 65 years which indicates global inequalities in survival are likely to persist if current trends are maintained.

Within cities, such inequalities also occur. For example, in the city of London, women living in Camden have approximately 5 more years of life expectancy than women living in Barking or Dagenham (Office for National Statistics, 2017). In the case of Madrid, on a 20-minute subway trip between the Districts of Chamberí and Puente de Vallecas, men will travel a 2.5-year gap on life expectancy (Estadísticas Madrid, 2017). Within cities, populations with similar socioeconomic characteristics tend to group in the same areas, generating spatial segregation. There is no doubt that health statistics in these areas are influenced by the socio-economic composition of the population living in them. Age, education, sex, and other individual factors influence people's health. But are these compositional factors enough to explain the geographical variations? Is there a contextual effect, some kind of local influence that makes the health of an area better (or worse) than what it would be expected from the composition of the population? Geography becomes a key element to understand how these socio-spatial inequalities are generated and maintained and to discern if they are explained by compositional, contextual, or a combination of both factors.

1.1.1 Geographic differences in health: contextual or compositional?

Often two possible explanations for the geographical variations in health have been identified: the compositional and the contextual (Macintyre and Ellaway, 2003). The compositional explanation argues that health differences between places are due to the type of people residing in these places, while the contextual explanation posits that it is due to the characteristics of the places themselves. For example, the compositional explanation argues that the poorest die before the rich, so the areas with a higher number of poor people will have higher mortality rates and lower life expectancy (Macintyre and Ellaway, 2003; Brown *et al.*, 2018). That is, the poor will die sooner and have worse health than the rich, no matter where they live. This compositional explanation implies that geographical differences in health are simply the result of the geographical distribution pattern of the poor and the rich. The compositional factors of geographical disparities in health would include individual socio-economic characteristics such as age, sex, poverty or social class.

Instead, the contextual explanation suggests that there are characteristics of the social and physical environment of a place that directly or indirectly influence the health of the people living there (Macintyre, Ellaway and Cummins, 2002; Diez-Roux and Mair, 2010); or even other individuals who may be exposed to that environment in the performance of their daily tasks (work, go to school, etc.) (Brown *et al.*, 2018). Are these differences between places what drive the socio-spatial pattern of health? People can live healthier/un-healthier lives if they reside in areas with characteristics that promote/restrict healthy habits. Such characteristics could include access to parks and green spaces (to promote physical activity and mental well-being), greater exposure to a poor quality food environment (such as fast food establishments that promote the consumption of nutrition-poor foods) or the design of an active transport infrastructure (to promote active travel such as walking and cycling).

It is important to understand the differences between the two models when exploring socio-spatial inequalities in health. However, compositional and contextual explanations have tended to be seen as mutually exclusive, competitive, and culturally and historically universal (Macintyre, Ellaway and Cummins, 2002). However, and as we will see in the next section, this distinction may be more apparent than real.

1.1.2 Relationship between composition and context: deprivation amplification

It is very likely that the distinction between people and places, composition, and context, is somehow artificial. People modify places and places modify people (Macintyre and Ellaway, 2003). Context and composition do not necessarily have to exert independent effects but can act synergistically through the deprivation amplification process (Macintyre, 2007). This refers to a process by which individual disadvantage is compounded by the disadvantaged area or neighborhood level. Let's imagine a person with low educational and socioeconomic status. It is likely that this person is doomed to live in a neighborhood with lower housing quality or with fewer community resources (green spaces, healthy food opportunities or local facilities). Living in this neighborhood may be less auspicious to maintaining good health, further amplifying the individual level risk. Therefore, the underlying social and spatial process that determine individual-level risks can also be, in part, a product of the environment to which he or she is exposed. Geographical space becomes both a product and a producer of health inequalities, reflecting a certain social, economic and political organization, materialized in a spatial segregation within cities (Barcellos and Buzai, 2006).

Therefore, residents of unfavorable neighborhoods experience a double disadvantage whereby they are not only personally disadvantaged but probably also lack the necessary infrastructure to carry on a healthy life in their neighborhood. This could translate into the development of unhealthy habits (smoking, poor diet or insufficient physical activity) that would become part of the causality chain of health outcomes (Macintyre, Ellaway and Cummins, 2002) and would not act alone as confounding factors in the analysis.

1.2 METHODOLOGICAL CHALLENGES OF HEALTH GEOGRAPHY

When analyzing the relationship between urban environments and health from a geographical perspective, it is necessary to take into account the challenges arising from this type of research. In this work, these challenges are mainly focused on the point of view of ecological studies. Ecological studies are those that examine health outcomes (or risk factors) using aggregated data, either under geographical or under temporal units. They are commonly used in Health Geography because they link health outcomes with a particular space. This kind of studies are comparatively un-expensive and easier to perform than cohort or case-control studies. However, health policies derived from this type of studies can be limited by three classic methodological problems: the ecological fallacy, the Modifiable Areal Unit Problem, and the spatial dependence and the spatial heterogeneity.

1.2.1 The ecological fallacy

The ecological fallacy consists of the incorrect inference of the relationships observed at the aggregate level (for example, between the low socioeconomic status of a community and the prevalence of obesity) in the same direction and magnitude at the individual level (for example, the probability of individuals with low socioeconomic status being obese) (Robinson, 1950; Macintyre, Ellaway and Cummins, 2002; Macintyre and Ellaway, 2003). Therefore, it is important to distinguish between the misuse of aggregated data in area units as a proxy for individual data (ecological fallacy), and the analysis of social and physical environment effects on the health of individuals or populations (ecological perspective) (Macintyre and Ellaway, 2003).

It should be noted that ecological analysis in epidemiology have been avoided for a while due to concerns about the ecological fallacy (Pearce, 2000; Macintyre and Ellaway, 2003). However, ecological studies are back, as shown by the number of studies published in the last decade using this study design (Coutts and Hahn, 2015; Golden *et al.*, 2015; Sallis and Owen, 2015; Kiraly *et al.*, 2017). This is because it is increasingly recognized that, even when studying risk factors at the individual level, studies at the population level play an essential role in the definition of most important public health issues and in the generation of hypothesis about their possible causes (Diez-Roux, 1998; Pearce, 2000). Ecological studies are back due to the fact that it is increasingly recognized that some disease risk factors actually operate at the population level (Rose, 1985; Pearce, 2000; Cummins and Macintyre, 2006).

Going back to the implications of the ecological fallacy in geography and health studies, it is important to take this "problem" into account when conducting multi-scale studies. When working through scales, we must avoid the erroneous assumption that a relationship observed on a scale will necessarily be maintained at other scales. For example, in an ecological study on a national scale, it may be concluded that higher population density levels (such as within cities) are more likely to have asthma as compared to populations with lower density levels (e.g., rural). However, this does not mean that at the city level, areas with higher population density are more likely to suffer from asthma than areas with low-population density.

1.2.2 The Modifiable Areal Unit Problem

When we attempt to collect contextual data, one of the main methodological issues to be taken into account is the selection of the aggregation unit (areal unit) and its relationship with the most appropriate spatial scale. Spatial resolution is related to the size and definition of the geographic area we hypothesize influences a health outcome (Diez-Roux, 2001).

Although many epidemiological and Health Geography studies use administrative areal units (census tracts, districts, municipalities, provinces and so on), it is often recognized this may not be the most appropriate spatial resolution (Macintyre, Ellaway and Cummins, 2002). Some of its influences can operate at an extremely local level (pedestrian accessibility to health food stores for instance), while others can operate at the regional, state or provincial level (the regulation of tobacco consumption in public places) (Macintyre and Ellaway, 2003).

This methodological challenge is known as the Modifiable Areal Unit Problem (MAUP) and is the geographical manifestation of the ecological fallacy (Openshaw, 1983; Eagleson, Escobar and Williamson, 2003; Anthamatten and Hazen, 2011). MAUP is an inherent problem in the use of aggregated data in areal units. It implies a potential error of measurement due to the data aggregation in units that have been delimited for historical, political or statistical reasons; and not for homogeneity reasons of social, economic or health population conditions (Fotheringham and Wong, 1991; Eagleson, Escobar and Williamson, 2002; Schuurman *et al.*, 2007). Although MAUP is an old friend of geosciences, it has also been identified as a methodological challenge in epidemiology and public health (Diez-Roux, 2001).

The way the MAUP reveals itself is two-fold: the *zoning effect* and the *scale effect* (Openshaw, 1984; Fotheringham and Wong, 1991). The *zoning effect* refers to the different configurations of the same size areas generating different results (Houston, 2014). An example of this effect would be when the results obtained in a continuous grid system of 100 m, differ from the results using a grid system of 100 m oriented differently (figure 1). The second MAUP effect

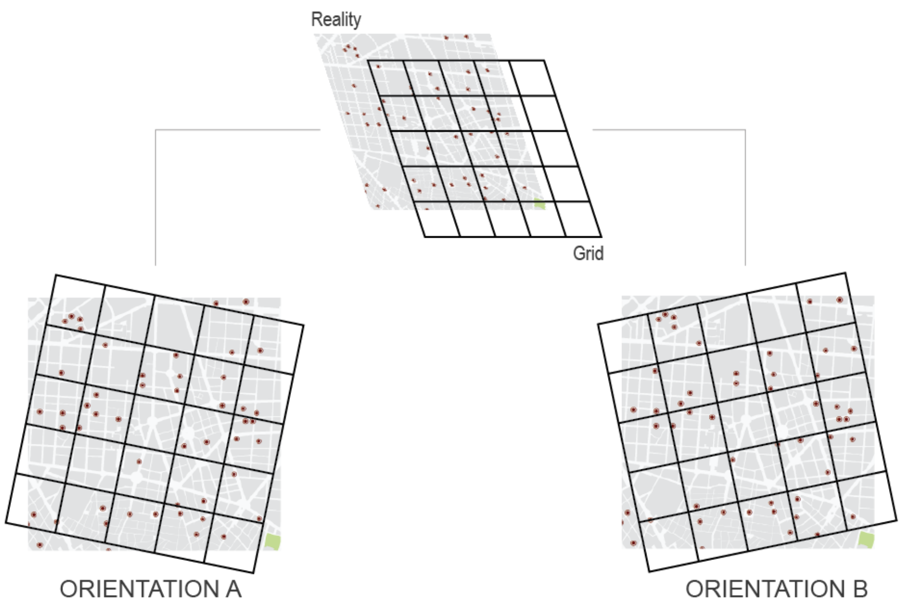


Figure 1. Representation of the zoning effect when changing the orientation of a grid.

is the *scale effect,* the one that commonly affects health studies. This effect implies the variation results using a grid system of 100 m oriented differently (figure 1). The second MAUP effect is the *scale effect,* the one that commonly affects health studies. This effect implies the variation that can be obtained in the health results when the data for a set of areal units are added progressively in smaller and larger analysis units (Openshaw, 1984). For example, when census data are aggregated in census sections, neighborhoods or districts, different results are obtained (figure 2).

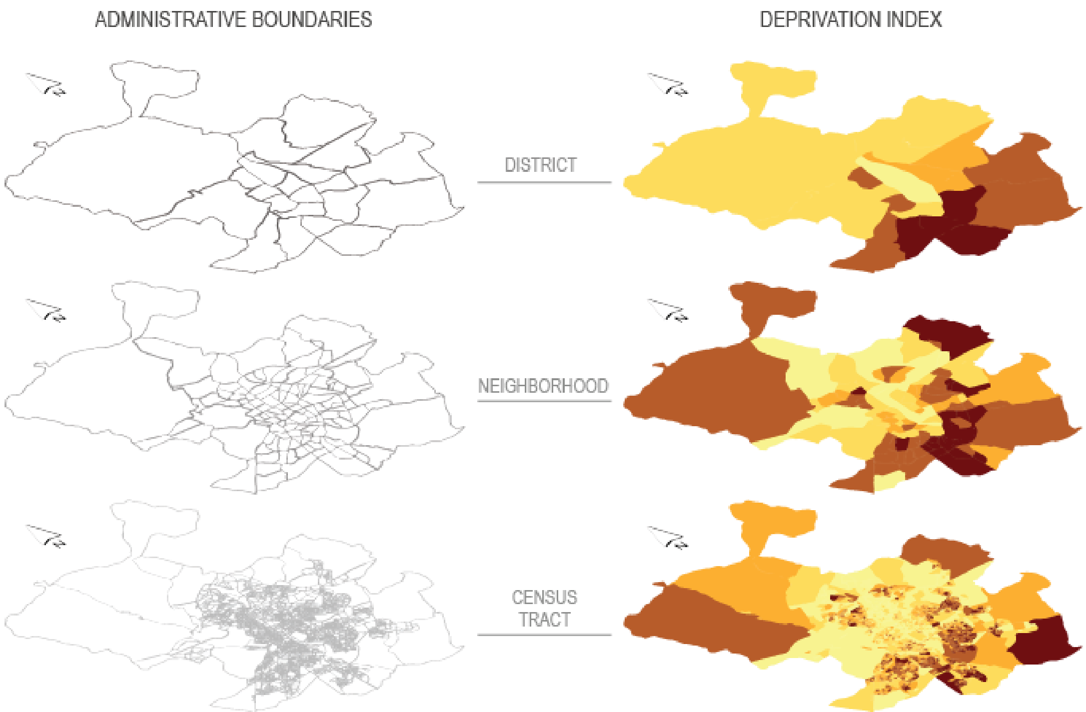


Figure 2. Representation of the scale effect on a deprivation index by increasing the size of the unit of aggregation (Census tract, neighborhood and District) of Madrid municipality.

Selecting the appropriate spatial scale involves establishing the spatial resolution of the study, that is, the ability to distinguish objects on the surface of the earth. On a larger scale, statistical aggregation of data on smaller surface units (i.e., census tracts, figure 2) will be required, but these areas contain fewer cases and, therefore, less stable rates. On the contrary, at smaller scales, aggregation occurs in larger areas (i.e., districts, figure 2), blurring significant variability and, sometimes, leading to interpretations that are contradictory to those derived from finer resolution data. (Hanigan, Cochrane and Davey, 2017; Nelson and Brewer, 2017). Therefore, it is necessary to take into account how MAUP affects most of the statistics and its impact on the variance, standard deviation, correlation, regression analysis and other statistical results (Fotheringham and Wong, 1991; Flowerdew, Manley and Sabel, 2008).

Any health phenomenon being analyzed is linked inseparably to the scale, given that its scale provides its meaning. However, many times there is no choice and we are constrained by the spatial scale or the level of aggregation to which the information is available. We have to be aware, when interpreting our findings, that this problem may be impacting our results.

1.2.3 Spatial dependence and spatial heterogeneity

Two problems arise when data have a spatial component: 1) the spatial dependence between the observations and 2) the spatial heterogeneity in the relationships we are modeling. These additional spatial "problems" or "effects" are related to the challenge of applying classical statistical methods to spatially aggregated data. Both of them are important in the statistical analysis since they can invalidate certain methodological results, demand adaptations to others, and in some contexts, they require the development of specialized techniques (Anselin, 1988).

Most parametric statistics assume that the variables are independent of the region where they happen. This assumption is directly in conflict with the first law of geography *"everything is related to everything else, but nearby things are more related than distant things"* (Tobler, 1970). This law expresses the process of spatial dependence or spatial autocorrelation. Spatial dependence means that in a geographical area if an event occurs in a particular location, another event is more likely to occur near the initial event than in locations farther away, and therefore, events cannot be considered independent (Griffith and Chun, 2014). In addition, spatial dependence also takes into account the influence that observation units have on each other.

As shown in figure 3, spatial patterns that would show the data can be dispersed, random or clustered. If the presence of an observation in a spatial location makes such observation less likely to occur in its environment, we speak of dispersion (negative autocorrelation). When the presence of an observation in a location increases the probability of finding similar observations in nearby places, we speak of a clustered pattern (positive autocorrelation). Finally, when none of the previous cases occurs, we say of a random pattern or spatial independence (autocorrelation close to zero). Several studies have shown that health outcomes and the contextual characteristics are not randomly distributed through space (Gebreab, 2018). On the contrary, they tend to show a certain degree of spatial clustering, depending on the scale at which they are collected and investigated (Jerrett, Gale and Kontgis, 2010; Root, 2012).

However, far from being simply a statistical nuisance that must be corrected, spatial autocorrelation can provide a substantive view of contextual determinants that affect health outcomes and their relevant spatial scales (Chaix, Merlo and Chauvin, 2005; Root, 2012). For example, the presence of spatial dependence on a national scale may suggest the influence of large-scale contextual factors, such as health infrastructure or policies to regulate tobacco use. Alternatively, spatial dependence between nearby neighborhoods may suggest the influence of local contextual factors, such as access to healthy food stores or sports facilities. Spatial autocorrelation analysis of the data can help determine the best scale for measurements and predictions in the investigation of context determinants that affect health.

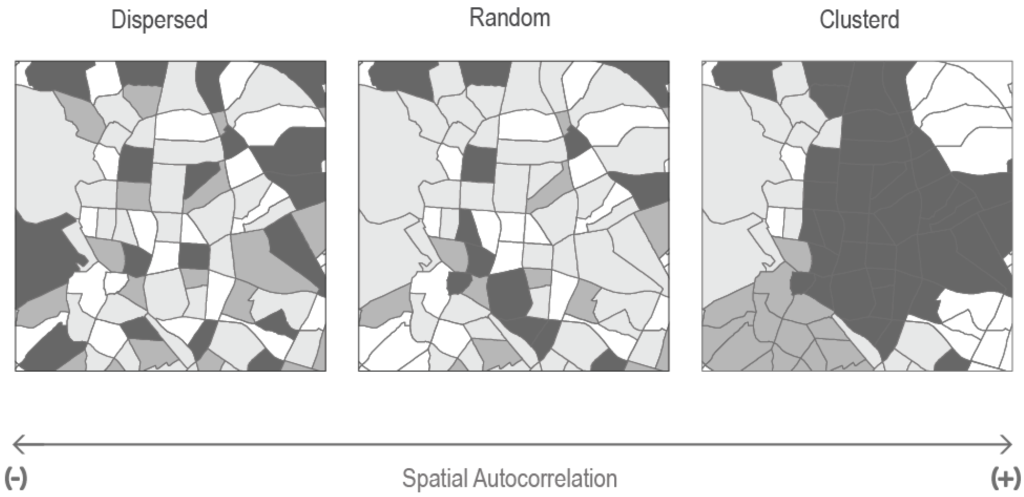


Figure 3: Types of spatial autocorrelation and their spatial pattern

The second spatial effect, spatial heterogeneity, is a concept defined by the absence of structural stability through geographical space (Anselin, 1988; Goodchild, 2004). Due to spatial heterogeneity, it is impossible to conceive an average location on the surface of the Earth and its full range cannot be understood by simply extrapolating a local conditions (Goodchild, 2004, 2009) (Goodchild, 2009, 2004). In words of Goodchild, *"Like the weather, geographic phenomena do not oscillate around a mean, but drift from one locally average condition to another. If this were not so, European explorers would not have found "new" worlds, but new samples of old ones. "* (Goodchild, 2009). Human activities are the main forces behind spatial heterogeneity in built environments. Geographical information relating to urban geography captures essentially spatial variations of the built environments, which also shows heterogeneity (Jiang, 2015).

Implications of spatial heterogeneity, or nonstationary in the statistical meaning of the term, lie in the fact that there is a variation in the relationship under study through the geographical space. Previously we mentioned that spatial dependence considers the observations to be dependent on the location where they are collected and on the neighboring observations. Following this assumption, spatial heterogeneity refers to the local variance of spatial dependence (Jiang, 2015). While spatial dependence violates the independent observations assumption, spatial heterogeneity violates the existence of a single linear relationship assumption with constant variance through sample data observations. (LeSage, 1999). This creates problems for traditional regression methods assuming a single constant relationship for the entire data sample. If the relationship varies, or if the variance changes as we move forward in the spatial data sample, alternative estimation procedures are needed to successfully model this variation and to extract appropriate inferences. For this reason, a series of methods have been developed to consider these variations through space (with random coefficients and variable parameters) (Anselin, 1988). However, these efforts have been slowly incorporated into the public health field due to the widespread lack of quality data, the training in spatial thinking and the software tools for spatial statistics. (Auchincloss *et al.*, 2012).

2. THE EPIDEMIC OF CARDIOVASCULAR DISEASE

Cardiovascular Disease (CVD)[1] is the leading cause of death in Europe. It presents the highest social and economic burden among all diseases, given the costs of treatment and its complications (Barton *et al.*, 2011). CVD covers a wide range of disorders, including diseases of the heart muscle and vascular system supplying heart, brain, and other vital organs. The most common manifestations of CVD are ischemic the heart disease, congestive heart failure, and stroke (Lopez *et al.*, 2006). CVD and associated risk factors such as diabetes, obesity, and hypertension continue to rise epidemically in contemporary populations (Franco *et al.*, 2011; Stanaway *et al.*, 2018). According to projections, its burden will increase in the next decades due to the increase in prevalence (Franco *et al.*, 2011; Foreman *et al.*, 2018). The prevention of cardiovascular diseases, as well as other non-communicable diseases such as cancer, diabetes and lung diseases, continues to be a current issue in public health, not only in Europe but worldwide (Foreman *et al.*, 2018; Stanaway *et al.*, 2018).

A general classification of the main risk factors of CVD divides them into non-modifiable (age, sex, genetic factors/family history, among others); and modifiable (hypertension, smoking, diabetes, unhealthy diet, obesity, sedentary lifestyle, dyslipidemia and alcohol consumption, mainly). In turn, modifiable risk factors can be classified as behavioral or biological. Behavioral risk factors include unhealthy habits, such as smoking, unhealthy diets, sedentary lifestyle and excessive alcohol consumption (Ezzati and Riboli, 2012; Eckel *et al.*, 2014). Behavioral factors are directly associated with biological risk factors such as hypertension, dyslipidemia, and diabetes; representing a large proportion of the excess risk of CVD in populations (Bilal, Díez, *et al.*, 2016).

2.1 BEHAVIORAL RISK FACTORS OF CARDIOVASCULAR DISEASE

2.1.1 Diet

Four of the fundamental behavioral risk factors of cardiovascular diseases (hyperlipidemia, hypertension, obesity, and diabetes) are mainly of dietary origin (Franco *et al.*, 2011). Unhealthy dietary practices comprise a high consumption of saturated fats, refined carbohydrates, and salt, as well as a low consumption of fruits and vegetables.

[1] CVD is used as an abbreviation of cardiovascular disease, not cerebrovascular disease

According to the "Global Burden Disease" research program that assesses the burden of disease and mortality at the global and regional level (http://www.healthdata.org/gbd), dietary risk factors account for 51.48% of all deaths by CVD in the world, or 9,085,088 deaths per year. Of the individual dietary risk factors, the highest attributable burden in 2010 was associated with low-fruits diets (4.9 million annual deaths), followed by diets high in sodium (4.0 million deaths), low in dry fruits and seeds (2.5 million), low in cereals (1.7 million), low in vegetables (1.8 million) and low in omega-3 fatty acids of marine origin (1.4 million) (Lim *et al.*, 2012). Spain is below these values, but they are no less worrisome. 35.17% of CVD deaths in Spain are associated with dietary risk factors, representing 44,617 deaths per year.

The Spanish diet is generally identified with the traditional Mediterranean diet. It is characterized by a high intake of olive oil, fruits, nuts, vegetables and cereals; a moderate intake of fish and poultry; a low intake of dairy products, red meats, processed meats and sweets; and a moderate consumption of wine, consumed with meals (Willett *et al.*, 1995). The Mediterranean diet is the most likely dietary model to provide protection against coronary heart disease (Gómez-Gracia *et al.*, 2013). Nowadays, Spaniards consume a Mediterranean diet modified by the food changes associated with the economic development of the last decades. Our diet includes an acceptable consumption of fruits and vegetables, relatively rich in cereals (in the form of bread), with a high amount of added fats in the form of vegetable oils (mainly olive oil) and high consumption of fish. However, the consumption of meat, prepared dishes, non-alcoholic beverages (juices and soft drinks, mostly sugary) and sugar is high, while the consumption of legumes is low. Although the consumption of dairy products is adequate, that of the varieties richest in fat is excessive (Villar *et al.*, 2007).

2.1.2 Physical activity.

It is estimated that insufficient physical activity is a risk factor that causes, globally, 3.2 million deaths per year (Lim *et al.*, 2012). In summary, the scientific evidence of the benefits of physical activity on individual health considers that it reduces the risk of cardiovascular disease, hypertension, diabetes, certain forms of cancer, and that ii has an important role in the management of certain chronic diseases. On mental health, physical activity has positive effects by reducing stress, anxiety, depression and, possibly, delays the effects of Alzheimer's disease and other forms of dementia (Lee *et al.*, 2012; WHO, 2015). It is estimated that not walking vigorously 15-30 minutes a day (WHO, 2010) is responsible for between 6% and

10% of the main non-contagious diseases burden (Lee *et al.*, 2012), with a global burden of disease similar to smoking (Wen and Wu, 2012). In particular, the burden of disease attributable to lack of physical activity is 5.8% for coronary heart disease, 7.2% for type II diabetes, 10.1% for breast cancer, 10.4% for colon cancer and 9.4% for general mortality (Lee *et al.*, 2012). In Spain, these attributable burdens are higher; therefore, the lack of compliance with WHO recommendations is responsible for 8.3% of coronary heart disease, 10.3% of type 2 diabetes, 13.8% of breast cancer, 14.9% of colon cancer and 13.4% of mortality general (Lee *et al.*, 2012). In addition, complying with WHO recommendations for physical activity would increase life expectancy by 0.68 years worldwide (Lee *et al.*, 2012).

2.1.3 Alcohol Consumption

In 2016, 5.3% (3 million) of all deaths worldwide were attributed to harmful alcohol consumption, being higher in men (2.3 million) than in women (0.7 million) (WHO, 2018a). Of all global deaths attributable to alcohol consumption in 2016, 19% were due to CVDs (WHO, 2018a). The global strategy to reduce the harmful use of alcohol defines "harmful use" as a consumption pattern that causes harmful health and social consequences for the drinker, the people around the drinker, the society in general, and is associated with an increased risk of adverse health effects (WHO, 2014). Harmful alcohol consumption carries a heavy burden. However, the relationship between alcohol consumption and CVD is complex: the volume of alcohol consumed and the pattern of consumption over time affect health in a different way. In particular, the pattern of alcohol consumption has been related to injuries, both involuntary and intentional (Macdonald et al. 2013), and the risk of CVD, mainly ischemic heart disease and stroke (Roerecke and Rehm, 2012). On the other hand, alcohol consumption has detrimental effects on hypertension, atrial fibrillation and stroke, regardless of the pattern of consumption (Roerecke and Rehm, 2012).

Surveys and mortality studies, particularly from the developed world, suggest that there are more drinkers, more drinking occasions and more drinkers with low-risk consumption patterns in the highest socioeconomic groups (WHO, 2014). This results in a paradox according to which disadvantaged populations with a higher number of abstainers and a lower or equal level of alcohol consumption (Van Oers *et al.*, 1999), seem to suffer greater harm and be more vulnerable to the consequences of alcohol consumption than the most favored populations (Jones *et al.*, 2015).

2.1.4 Smoking

One of the main risk factors for CVD and other chronic diseases such as cancer or respiratory diseases is smoking. The WHO considers smoking an epidemic that kills more than 7 million people a year and almost 80% of the more than one billion smokers in the world, residing in low or middle-income countries, where the burden of morbidity and associated mortality is greater(WHO, 2017).

In Spain, the smoking habit resembles that of the rest of the countries of the European Union with a smoking prevalence of 31.4% for men and 27.4% for women (WHO, 2018b). The historical pattern followed by the epidemic in Spain marked historical highs in men in the 70s and 80s, with prevalence rates between 54.0% and 61.3%. In the case of women, maximums were reached in the 90s and beginning of 2000 with a prevalence between 26.0% and 26.7% (Bilal *et al.*, 2014). The intersection between gender and class helps to understand these gender differences in smoking habits. Men from upper social classes first started smoking, followed by men from the middle classes and upper-class women (Bilal, Beltrán, *et al.*, 2016). As men of the most favored classes were abandoning the habit, it was adopted by men of a lower social class and women of that same class. Today, people from the most disadvantaged social classes have the highest prevalence of smoking (Hiscock *et al.*, 2012; Bilal, Beltrán, *et al.*, 2016).

2.2 STRATEGIES FOR PREVENTION OF CARDIOVASCULAR DISEASE

Given the economic and public health burden of CVD and their associated risk factors, it is necessary to adopt new research approaches and methodologies to obtain reliable results for future prevention policies. But before, it is essential to understand the etiology frames of the disease. According to the classic epidemiological work of Geoffrey Rose (Rose, 1985), two preventive approaches to cardiovascular disease are differentiated: the individual and the population.

The individual approach is focused on the sick individuals and seeks answers to questions such as, why does this person have hypertension? Meanwhile, the population approach seeks answers to questions such as: Why are there so many people with hypertension in my country, compared to others? The first approach corresponds to the traditional medical approach and is addressed by individual prevention strategies (i.e. smoking cessation programs). The second approach attempts to change the entire distribution of a given disease (or risk factor) in a population group. Therefore, the effect of prevention is not limited to people with high risk but on contextual factors, the “mass forces" defined by Rose (Rose, 1985), affecting the entire population. The limited success of current preventive strategies based on the individual level demonstrates the need for new approaches based on the determinants of CVD at the population level (Lake and Townshend, 2006; Arnott *et al.*, 2014).

In epidemiology and public health, this approach is relatively new and was originated as a response to certain postulates of individualism and positivism. It tries to shift the attention to the structural and contextual influences on health (and their behaviors), as well as to inequalities in health (Macintyre, Ellaway and Cummins, 2002). On their side, geographers and sociologists have long argued that place is important to explain the social and spatial pattern of health (Anthamatten and Hazen, 2011; Brown *et al.*, 2018). The cities and neighborhoods where we live represent a unique opportunity to study how the determinants (social and physical) of urban environments contribute to health outcomes (Borrell *et al.*, 2014; Franco, Bilal and Diez-Roux, 2015), and based on the knowledge generated, develop sound population prevention strategies.

3. URBAN CONTEXT AND HEALTH

The United Nations estimates that by 2050, 68.4% of the world population will live in cities, reaching figures of 88.0 % in developed countries (UN DESA, 2018). In Spain, currently 80.1% of the population live in cities and it is estimated that by 2050 this percentage will increase to 84.6 % (UN DESA, 2018). Given this scenario, urban environments are presented as important contexts with a role in the health of the populations, since they possess physical and social characteristics (Diez-Roux and Mair, 2010) that can limit or promote unhealthy behaviors.

Modifying physical and social characteristics of the urban environment may have the potential to improve the health of individuals and may also produce a reduction in social and environmental inequalities in health. For example, an effective prevention strategy for CVD will depend on actions to improve the living environment, facilitate physical activity, reduce tobacco and alcohol consumption, regulate access to dangerous products, and provide information and education services for healthier lifestyles (WHO, 2008). Having a deeper appreciation of the physical and social contextual characteristics of the "place" allows us to better understand the specific causal mechanism between the contextual factors of cities and health, specifically in the case of this PhD dissertation, on cardiovascular health.

3.1 URBAN CONTEXT I: BUILT ENVIRONMENT AND CVD

The health risks associated to the urban environment have been theorized to cover a wide range of environmental risks together with aspects of the built environment which can be potentially positive or negative for health (Brown *et al.*, 2018). The definitions and conceptual frameworks that define the scope of built environment vary throughout the scientific literature. A definition of built environment could be one that encompasses all relatively fixed aspects, such as buildings, spaces, and products people create or modify. It includes urban design, transportation systems, together with planning and policies that define land uses (Rao *et al.*, 2007; Duncan and Kawachi, 2018).

The causal relationships between built environment and health are difficult to clarify, but there is a growing body of evidence showing mechanisms by which the built environment impacts on health. In 2003 Diez-Roux created a conceptual model of possible routes by which certain characteristics of the built environment could impact on CVD risk (Diez-Roux, 2003). In this conceptual framework (figure 4), Diez-Roux shows the characteristics of the built environment (contextual determinants or factors) that can be relevant for cardiovascular health outcomes. The contextual determinants of CVD are complex and multifaceted, considering in this work, only those that influence the behaviors related to physical activity, eating, alcohol consumption, and smoking.

Figure 4. Conceptual model of possible pathways linking the built environment to CVD (Adapted from Diez-Roux 2003).

The contextual determinants of physical activity can be divided into those influenced by active transport and those by physical activity in leisure time. The determinants of active transport are related to walking and to the use of bicycles and include characteristics such as the quality of pavements, the presence of bicycle lanes, safety and aesthetics, land uses, destinations or connectivity of streets (Pikora *et al.*, 2003). The contextual influences of physical activity in leisure time, or recreational, are determined by the accessibility and availability of sports facilities and green spaces (Pikora *et al.*, 2006). Likewise, the availability and accessibility to certain characteristics of the local environment can promote or mitigate behaviors related to diet, alcohol consumption and smoking. The contextual determinants of healthy eating include all aspects of the local food environment that influence eating behavior (Caballero, Finglas and Toldrá, 2016). Food stores and their associated accessibility, as well as the availability of healthy foods, have been shown to affect dietary behaviors

(Franco *et al.*, 2009). The contextual determinants that influence the consumption of alcohol and tobacco are also associated with the accessibility, availability and promotion of alcoholic beverages and tobacco (Popova *et al.*, 2009; Lee *et al.*, 2015; Foster *et al.*, 2017; Sureda *et al.*, 2017).

Researchers have expressed the need to examine how the characteristics of the built environment give rise to the presence of cardiovascular risk factors (Diez-Roux, 2003). However, research on health and urban environment encompasses a series of key challenges, some of which have already been discussed (the scale or spatial unit choice), along with others such as the development of methodologies to examine contextual determinants or the relationship between the characteristics of different built environments. In one of the most exhaustive longitudinal research in this regard, the Multiethnic Study of Atherosclerosis (MESA), features of the built environment, such as access to healthy foods, and physical activity resources, were linked to various CVD risk factors. Along 5 years, adult participants aged 45 to 84, residing in an area with access to healthy foods and physical activity resources, presented a 10% lower risk of obesity (Auchincloss *et al.*, 2013) and 38% lower diabetes (Auchincloss *et al.*, 2008) regardless of individual characteristics and behaviors. However, despite other studies supporting these results (Papas *et al.*, 2007), the diversity of methodologies used and the results obtained reflect the complexity of the chain of causality that links contextual determinants and different chronic diseases, as well as the challenges inherent to the measurement of complex social phenomena (Glass and McAtee, 2006; Feng *et al.*, 2010; Leal and Chaix, 2011; Cobb *et al.*, 2015; Townshend and Lake, 2016).

3.1.1 Interrelation of built environments

Previous studies have generally characterized the built environment narrowly (i.e., considering a single exposure construct) examining its association with specific behavioral risk factors and specific health outcomes (Popova *et al.*, 2009; McCormack and Shiell, 2011; Lee *et al.*, 2015; Lytle and Sokol, 2017). Taking the regulation of body weight as an example. Body weight depends on multiple factors, being physical activity and eating healthy among the most relevant (Bethlehem *et al.*, 2014). Yet, research on contextual determinants and obesity tends to treat both physical activity and food environments independently, obtaining associations with different health outcomes. However, characteristics of the urban environments are not isolated but are the result of social forces shaping neighborhoods (Guthman, 2013). The strong correlation between physical activity and dietary behaviors requires strategies addressing physical activity and diet options simultaneously (Goverment Office for Science, 2003; Economos *et al.*, 2015; Meyer *et al.*, 2015). Interventions may result ineffective if those only focus on promoting physical activity ignoring a food environment which could promote the intake of unhealthy foods (Economos *et al.*, 2015). For these reasons, a new approach emerged to operationalize the contextual determinants of obesity by simultaneously adding contextual factors to both the physical activity and the food environment. It was named the obesogenic environment (Lake and Townshend, 2006; Saelens *et al.*, 2012; Townshend and Lake, 2016). For example, Saelens et al., (2012) found a significant association between physical activity and food environments combined with obesity, suggesting characteristics of neighborhoods can work together to create obesogenic environments (Saelens *et al.*, 2012).

Although the evidence about the usefulness of integrated measures of built environment is increasing, the study of the obesogenic environment has mainly focused on the relationship with the overweight or obesity (Nau *et al.*, 2015; DeWeese *et al.*, 2018; Hobbs *et al.*, 2018). Despite this, recent studies have found spatial associations between other built environment domains and health behaviors (Schneider and Gruber, 2013; D'Angelo *et al.*, 2015; Shortt *et al.*, 2015). These studies focused their attention to a new area of research that combines multidimensional measures of the built environment related to health. For example, a study in Germany found tobacco, alcohol and fast food stores were more likely to be grouped in low-income neighborhoods (Schneider and Gruber, 2013). This study concluded the obesogenic environments and addictive environments can have a contextual influence on a person's lifestyle and contribute to health risks. Similarly, a study in Glasgow, Scotland, showed clusters of alcohol, tobacco, fast food, and gambling in the most deprived areas (Macdonald *et al.*, 2018).

Following the evidences provided by these studies, some efforts should focus on the interwoven nature of the multidimensional characteristics of the built environment related to CVD. The cardiovascular environment could be considered as one that integrates multiple factors of the urban context affecting CVD behaviors. As previously noted, there are investigations that relate the food environment to obesity (Lytle and Sokol, 2017) and the physical activity environment with behaviors related to sedentary lifestyle or lack of physical activity (McCormack and Shiell, 2011). Other studies have explored the associations between availability and accessibility to both alcoholic beverages and tobacco stores, with alcohol and tobacco consumption behaviors, respectively (Popova *et al.*, 2009; Lee *et al.*, 2015). Therefore, the need to integrate different domains of the urban environment that influence CVD behaviors is considered imperative (Guthman, 2013; D'Angelo *et al.*, 2015; Meyer *et al.*, 2015; Green *et al.*, 2018).

3.2 URBAN CONTEXT II: SOCIAL ENVIRONMENT

In the investigation of health behaviors, the influence of social factors is widely recognized. Although there is no definition of "social environment" universally accepted, it refers to the configuration of norms, the application of social control patterns, promoting or limiting opportunities to develop certain healthy behaviors and reducing or producing stress (Berkman, Kawachi and Glymour, 2014). In this dissertation, we focus on the social phenomena related to socioeconomic stratification, discrimination or poverty. It is important to emphasize that both the differentiation of social environments and the grouping of people with a similar social and economic position are not accidental, rather, it expresses the social structure through the geographical differentiation of the city (Roberts, 1997).

We know poverty and poor health are linked to each other: most disadvantaged countries generally have poorer health conditions, and within countries, socially and economically disadvantaged people tend to be less healthy (Marmot *et al.*, 1997; Anthamatten and Hazen, 2011). Research has revealed that this "health gap" forms a clear gradient: health status improves with increasing socioeconomic status, whether stratifying individuals or areas (Marmot *et al.*, 1997; Marmot, 2005; Kreatsoulas and Anand, 2010). Consequently, unhealthy behaviors are not distributed randomly in populations, but they present a socio-economic pattern and also tend to group together (Berkman, Kawachi and Glymour, 2014). The research focused on disadvantaged areas has supported the idea that factors at local level might help explain the relationship between health and socioeconomic inequality. Area effects (or neighborhood effects) consider the "net change in the contribution to life opportunities that occur when living in one area rather than in other" (Anthamatten and Hazen, 2011). The central idea is that neighborhood characteristics are important for health despite the characteristics at individual level. For instance, a national study in the United States found that neighborhood socioeconomic status is more closely related to health status than individual or family socioeconomic status (Robert, 1998). Another study in Chicago, with individual-level variables held constant, six neighborhood-level indicators predicted low birth weight, together contributing to a variation in rate of 5.5% (Roberts, 1997).

Marmot et al. (1997) considered three types of explanation for these social inequalities in health: health determines social position (health selection), social position determines health (social causation), and factors operating early in life determine both the social status reached as the state of health (indirect selection) (Marmot *et al.*, 1997). The social causes suggest disadvantaged people are not as healthy as a result of their low income and social position. This explanation focuses attention on the social determinants of health: the social, political and economic conditions that influence health. As Marmot showeds in his work "*Improving the social environment to improve health"* (Marmot, 1998), "*the main determinants of disease are mainly economic and social, and therefore their remedies must also be economic and social*".

Consequently, this work focuses on the socioeconomic determinants of health. Specifically, socioeconomic determinants include the conditions under which people are born, grow, live, work and grow old (WHO, 2008), and are determined by the distribution of money, power and resources globally, nationally and locally (Stronks *et al.*, 2016). Despite advances in primary and secondary prevention of CVD, large inequalities are still maintained across space and time (Roth *et al.*, 2017; Foreman *et al.*, 2018). As mentioned before, the prevalence of some cardiovascular risk factors (for example, obesity or diabetes) is increasing worldwide (Roth *et al.*, 2017; Foreman *et al.*, 2018). Therefore, it is necessary to focus efforts on understanding the role of the "causes of causes" (i.e., the social determinants of health) to help closing the current equality gap.

3.2.1 Social determinants of health: socio-economic deprivation and CVD

The urban environment itself can be considered as a social determinant of health by offering the conditions under which people are born, grow, live, work and age (WHO, 2008). In turn, these urban environments are determined by political, social and economic forces; and they influence a person's opportunity to be healthy, their risk of disease and their life expectancy (Stronks *et al.*, 2016). Social inequalities in health are the unfair and avoidable differences in health status among socially, economically, demographically and geographically defined population groups (WHO, 2008). They are the result of the unequal distribution of social determinants. These inequalities tend to be greater in urban areas, with disadvantaged groups concentrated in marginalized neighborhoods (Borrell *et al.*, 2014).

Several studies have documented associations between local socioeconomic determinants and cardiovascular risk. Diex-Roux et al. (2001), in a longitudinal study, examined socioeconomic disadvantages at the neighborhood level according to socioeconomic characteristics, with the incidence of coronary heart disease (Diez-Roux *et al.*, 2001). With nine years of follow-up, they found that living in the most disadvantaged neighborhoods was associated with a 70% to 90% higher risk of coronary heart disease in whites and between 30% and 40% higher in blacks, regardless of the individual characteristics (Diez-Roux *et al.*, 2001). Other studies have found a similar association between neighborhood socioeconomic status and the prevalence of CVD, incidence, mortality, and a variety of risk factors (Waitzman and Smith, 1998; Chaix *et al.*, 2006; Chaix, Rosvall and Merlo, 2007; Brown *et al.*, 2011), regardless of individual characteristics, suggesting residents' socioeconomic characteristics do not fully explain them.

Frequently, the concept of area deprivation has been used to characterize and to study the impact of socioeconomic determinants on health (Burrows *et al.*, 2011; Laraia *et al.*, 2012; Weng *et al.*, 2017). The term "deprivation" emerged in Britain in the late 1980s as a result of a long tradition in the analysis of social inequalities in health (Townsend, Phillimore and Beattie, 1988). Deprivation can be defined as a state of observable and demonstrable disadvantage in relation to the community, society or nation to which an individual, a family or a group belongs (Townsend, Phillimore and Beattie, 1988). Deprivation indexes are tools that provide a synthetic measure of the different aspects of socioeconomic deprivation. These measures have been widely used because they are easily available from the census data, which allows calculating deprivation indexes to different area units and relating them with georeferenced health data.

4. METHODS AND TOOLS FOR MEASURING URBAN ENVIRONMENTS: GEOGRAPHIC INFORMATION SYSTEMS

Geographic Information Systems (GIS) is a valuable addition to the health geographer's toolkit. Since the maps were used for the first time to analyze and describe spatial patterns in health, no innovation has had as much impact on the study of Health Geography as GIS. There is not a single, widely accepted definition of GIS. The British-American geographer M.F. Goodchild defined GIS based on three characteristics: by its functionality (a system for input, storage, analysis and output of geographically referenced information), its contents (a system containing geographically referenced information) and its purpose (a system for support of spatial decision making)(Goodchild, 1987). However, all three characteristics are broad enough to include a vast array of software products. Another definition of GIS could be a system of hardware, software, data and users that allows to capture, store, display, map, analyze, etc. geographical information and with this help to make decisions (Bosque, 1992). Along with new technologies such as Global Positioning Systems (GPS), terabytes of open spatial data on the Internet, advances in visualization techniques such as 3D or animated or dynamic cartography; the GIS continues to increase the possibilities of health studies (Dummer, 2008; Anthamatten and Hazen, 2011).

All spatial information stored in a GIS is georeferenced. It can be based on a geographical coordinate system (latitude and longitude) or on a projected one (X and Y axes). With the defined coordinate system, GIS allows to store spatial characteristics that are the representation of any geographical entity with a location, being a person, an address, a road or a city. In addition, GIS allows storing additional information about these spatial characteristics in the form of an attribute. For example, if we know the population density of all the neighborhoods of a given city, it would be possible to generate a thematic map representing the population density gradient of that city. This additional information may consist of statistics on almost anything interesting to the researcher, such as smoking habits, the density of green spaces or the prevalence of a disease. When these attributes (variables) are linked to the position data, the result is a powerful means to analyze the geographical distribution of the concerned phenomenon.

One of the main uses of GIS in the study of exposure to urban environments is the development of spatial exposure measures. Regardless of the domain of exposure (food environment, physical activity, alcohol or tobacco), the main spatial exposure measures used in the literature are limited to the spatial dimensions of "availability" and "accessibility" (Popova *et al.*, 2009; Charreire *et al.*, 2010; Bivoltsis *et al.*, 2018; Valiente *et al.*, 2018). Density measures are the most developed in relation to the availability dimension (Bivoltsis *et al.*, 2018). They are based on the presence, proportion, variety, count, relative density or diversity of an entity, within a defined area. In order to assess the spatial dimension of accessibility, proximity measures are usually used between a reference point and one or several destination points. Proximity measures are generally expressed as distances within the street network, straight line distances, travel times or spatial interaction models, including gravity models that quantify the decay ratio of the distance between two locations where the use decreases with increasing distance from a reference point (Bivoltsis *et al.*, 2018). As it has already been pointed out, when speaking about built environment, the different exposure measures used in the literature contribute to the current contradictory evidence in the study of urban environments and health.

Another useful feature of GIS is the ability to combine data from several sources allowing the exploration of spatial patterns of different phenomena. For example, when exploring the mortality patterns due to coronary ischemic disease, GIS allows linking a table with the adjusted mortality rates by census section of Madrid, with a file containing geographic boundaries of the census sections of Madrid. This way, it is possible to produce a map that can visually reveal spatial patterns. When having additional spatial information on income, poverty, or pollution, we may explore if any of these characteristics may be related to the mortality pattern. This is an example of the well-known Exploratory Spatial Data Analysis (ESDA). The ESDA extends the definition of Exploratory Data Analysis (EDA) to spatial data and represents a preliminary process in which data and research results are viewed from many different points of view, one of which is data visualization on maps. Visual analytics is an emerging interdisciplinary approach and the creation of meanings to explore, analyze, communicate and generate hypotheses based on large, complex and heterogeneous data (Nelson and Brewer, 2017) (Nelson & Brewer, 2015). Thus, one of the advantages of ESDA is that it allows generating new hypotheses from a large volume of data instead of validating them a priori.

While visualization of spatial patterns is a key part of the study of Health Geography, more sophisticated techniques are usually required to generate meaningful conclusions. For example, spatial statistics have developed several methods making possible to consider the dependence or spatial heterogeneity of the data in the analysis. These methods have been slow to penetrate the fields of epidemiology and public health, possibly due to the lack of data quality, the lack of training in spatial thinking or the lack of specialized staff in spatial statistics software tools. Today, GIS and spatial statistics tools have become much more accessible. Further below, I will explain in detail the methods and techniques of spatial statistics used in this PhD thesis.

It could be said that these are the most widely used GIS applications in the field of public health research. However, the rise of smart mobile communication devices has opened a new world of possibilities. These devices are able to identify their position with a high degree of precision through Global Positioning Systems (GPS) and through the triangulation of mobile telephone towers (Oliver et al., 2015). In addition, smart mobiles have an internet connection and a multitude of applications that allow their connection with the geographical location of the device. What is interesting and powerful is the fact that mobile phones, when connected, leave behind a digital trail, which can be used to analyze and model human behavior at the individual and aggregate levels (Oliver et al., 2015). These functionalities have led public health research to new scenarios as the following examples show.

Accurate estimation of an individual's exposure to an environmental characteristic implicated in the etiology of a disease has been an inaccurate science so far. Often, the place of residence or work is used as an approximation. However, individuals develop their daily activities in complex patterns. For this reason, they are exposed to a variety of urban contexts. Therefore, researchers are using the new features of remote devices to unravel the relationship between exposure to urban environments and health outcomes (Kirchner *et al.*, 2013). Kirchner et al. (2013) used GPS data captured from mobile phones to study the relationship between the desire to smoke and exposure to tobacco outlets. Mobility data were used to quantify the number of times the participants came into contact with a tobacco outlet, and in turn, they registered the desire to smoke and the state (whether they had smoked or not) in real time through the Ecological Momentary Assessment (EMA) (Kirchner *et al.*, 2013). EMA is a method increasingly used in research that requires the collection of self-reported data in real time on the experiences of people in their daily activity environment (Stone and Shiffman, 1994). EMA usually collects multiple observations per day for periods of several days, weeks or even months. EMA, integrated into an intelligent device, provides the opportunity to study geographic-temporal patterns and the momentary processes that influence behavior within the natural environment (Engel *et al.*, 2016). Another study in New Zealand used accelerometers, GPS and GIS to assess the proportion of physical activity that children performed in public parks with playgrounds. They concluded that the combination of these methodologies is a successful procedure to improve knowledge of the locations where children engage in physical activity (Quigg *et al.*, 2010). GPS devices together with GIS have also been used to track when and where people buy food in relation to their availability and activity space within their food environment (Christian, 2012; Gustafson *et al.*, 2013)

Mobile applications developed for the self-monitoring of the health and physical condition of users can also be very useful for research. Currently, there are several interactive tools for people tracking dietary or physical activity behaviors online. Using GPS on their phones, or other wireless sensors, users might track their physical activity or their food behaviors (Hirsch *et al.*, 2014; Middelweerd *et al.*, 2014; Franco *et al.*, 2016; Schoeppe *et al.*, 2017). This allows researchers to link dietary information or geocoded physical activity with other geographic characteristics at specific dates and times. Tracking facilitates the investigation of large-scale activity patterns and the influence of small-scale factors (such as the socioeconomic status of the neighborhood, characteristics of built environment or the parks and green spaces). The global increase of the use of these tools allows for larger samples of behavior and location that don't have been available so far. In addition, the samples collected throughout the world allow developing international research with the comparison of different countries.

Social networks are also representing an important new data source for researchers from the Geography of Health and Public Health domains. The Ghosh and Guha obesity study stands out in the United States. Using data from the Twitter social network they showed that within a quarter of a mile from the location of a tweet with food-related content, it is easier to find two McDonalds than a supermarket or a large grocery store (Ghosh and Guha, 2013). The same study found a significant correlation between the state with the highest number of policies related to obesity and the number of tweets with content on prevention of obesity. In the same line of research, another study from the United States, using geolocalized tweets, explored the prevalence of healthy and unhealthy foods. In addition, it examined whether tweets about unhealthy foods were more common in underprivileged areas (Widener and Li, 2014). Their results showed most disadvantaged areas tend to have a smaller proportion of tweets about healthy foods with positive sentiment and a higher proportion of unhealthy tweets in general (Widener and Li, 2014). Other studies have used geolocalized tweets within the field of health. Their interest lies on analyzing happiness, food environment and physical activity (Nguyen *et al.*, 2016); depression (Yang and Mu, 2015), tobacco (Kim *et al.*, 2015), the social environment (Shelton, Poorthuis and Zook, 2015; Nguyen *et al.*, 2017) and the most one, mobility (Hawelka *et al.*, 2014; Lenormand *et al.*, 2014; Jurdak *et al.*, 2015; Luo *et al.*, 2016; Yin *et al.*, 2016).

The purpose of this section is not to develop an exhaustive review on positioning technologies and their application in health. However, it is important to show that, thanks to the ubiquity of mobile devices and advances in communication technologies, today we have the ability to monitor human behavior outside clinical environments and without having to rely on self-reported information.

5. THE HEART HEALTHY HOODS PROJECT

This PhD dissertation has been developed along and within the Heart Healthy Hoods (HHH) project. This European Research Council-founded project, is a social epidemiology study whose main objective is to study the association between the social and physical characteristics of the urban environment and cardiovascular health in the municipality of Madrid, Spain (Bilal, Díez, *et al.*, 2016). To achieve this, it proposes an innovative and multidisciplinary methodological framework, where each scientific domain offers the relevant paradigmatic grounds, and in which geographical analysis, with tools such as GIS, plays a fundamental role. The theoretical framework underlying the project assumes that both the social and the physical environments impact on the individual's health-related risk factors (figure 5).

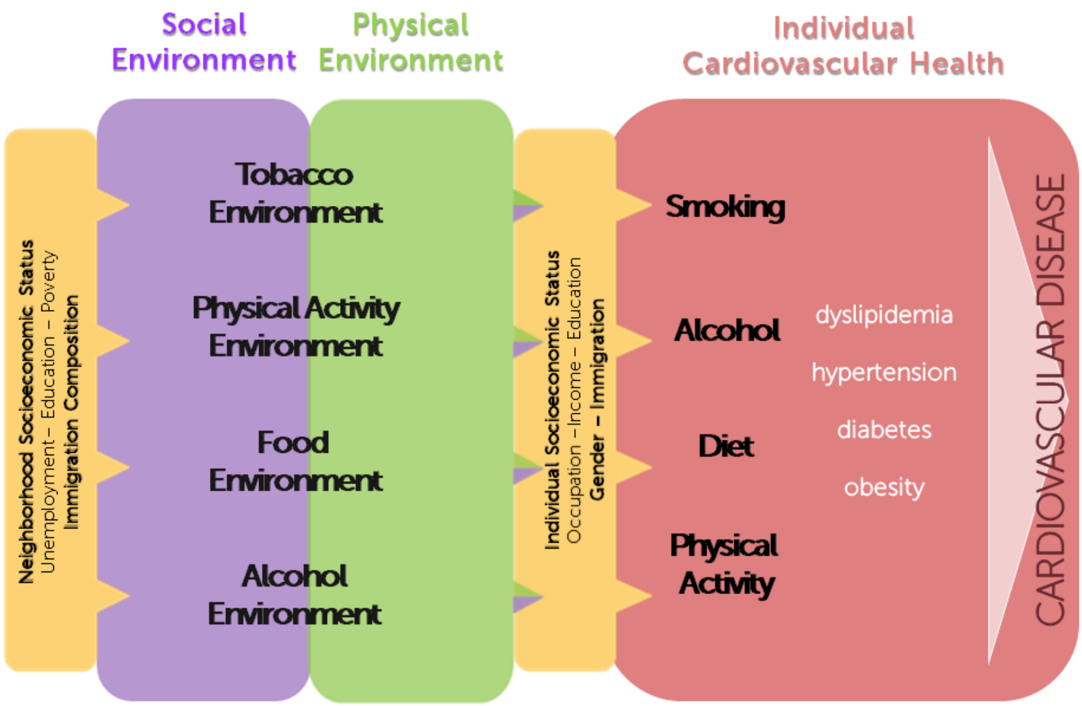


Figure 5. Conceptual framework of the Heart Healthy Hoods project. Social and physical environments affect individual behavioral risk factors (smoking, unhealthy diets, lack of physical activity and consumption of alcohol) directly associated with CVD. Behavioral risk factors increase the risk of hypertension, dyslipidemia, and diabetes (which represent a large proportion of the excess CVD risk at the population-level).

Individual risk factors (smoking, unhealthy diets, insufficient physical activity, and alcohol consumption) are directly associated with CVD. These behavioral risk factors increase the biological risk of hypertension, dyslipidemia, and diabetes represent a large proportion of the excess CVD risk at the population-level (Bilal, Díez, *et al.*, 2016). Within this project, urban environments refers to food, physical activity, tobacco, and alcohol, which are measured and evaluated through three complementary approaches: residents' perceptions, geographic information systems and systematic social observation. These observations are correlated with cardiovascular health data obtained from two different sources. First, a cohort study based on primary care that includes 2200 people recruited from 31 primary health care centers in Madrid. Second, a population-based study that includes all residents of Madrid aged 40 to 75 using electronic primary care records. This study also combines econometrics, geography, sociology, and anthropology to obtain a complete description of the environments in which residents live and work in Madrid. In addition, the cohort study includes direct measurements of cardiovascular health indicators, which constitute a solid and multifaceted data source. The study of the entire population offers a complete picture of the cardiovascular health of the ~ 1.5 million residents of Madrid aged 40 to 75. This PhD dissertation began in the early stages of the HHH project, with the aim of designing novel methods to better characterize the physical and social environment of Madrid.

SECTION II.
RESEARCH

As previously discussed, the geospatial approach has become essential to help unravel the relationships between urban factors influencing people's cardiovascular health. It is becoming clearer that the social and physical characteristics of geographical areas contribute to health outcomes (contextual effect) independently of the profile of the individuals who live in those areas (compositional effect). Health Geography recognizes the importance of the context, the environment, and the spatial scale, from the global to the local, in the determination of health outcomes.

This dissertation aims at addressing the issues associated with the measurement and analysis of the urban contextual determinants of cardiovascular disease discussed in the previous sections. .

There is still potential for the full uptake of GIS technologies and geographical theory in ecological studies researching urban environments related to CVD. The widespread use of functions such as proximity analysis or analysis in defined areas (buffers or administrative areas) is underutilizing other more robust GIS functions, such as interpolations (Charreire *et al.*, 2010; Feng *et al.*, 2010; Buck *et al.*, 2011; Thornton *et al.*, 2012). An example would be the Kernel Density Estimate (KDE) interpolation. The KDE adjusts to a mathematical surface (composed of pixels) with a normal distribution at each observation point. Essentially, the value of each point is smoothed over the study area producing a density value that is higher at the location of each point and decays when increasing the distance using a defined bandwidth (Carlos *et al.*, 2010). KDE overcomes the limitations in the binary definitions of the analysis based on fixed geographical limits (for example, number of stores per census section). Smooth transitions across boundaries (defined administratively) better represent the reality of urban environments (Chaix *et al.*, 2009). The resulting KDE surface can be used as an independent variable in statistical models (Carlos *et al.*, 2010).

Another example would be the use of traditional linear models which assuming the independence of observations. This assumption will probably produce inaccurate estimates in the presence of spatial autocorrelation (Kirby, Delmelle and Eberth, 2017). In addition, the traditional models also assume a unique linear relationship with constant variance through observations of sample data (LeSage, 1999). This implies not considering the spatial heterogeneity of data by assuming that a single constant relationship is maintained for the entire data sample.

The health data generally used in this type of studies have been based on self-reported or mortality results. Self-reported data are limited by the fact that they can rarely be independently verified and usually contain several sources of bias (response, memory, attribution or social desirability) (Rothman, Greenland and Associate, 2014). Spain has a universal health system, integrated, based on primary care and is free. It has an electronic health registration system that provides support for administrative functions (appointments, schedules, activity records) and management (to obtain indicators), providing a complete record of the medical care process. Thanks to the HHH project, this research has cardiovascular disease data of ~ 1.5 million inhabitants of Madrid City obtained from these electronic records. Having this information makes a difference when studying the impact of both built and social environment characteristics on the behavior of the inhabitants and their results in cardiovascular health.

Another justification for this research is the need to develop studies linking health and the urban environment in Europe. At present, several reviews have summarized the available evidence on the relationship between the environment and different health outcomes (Feng *et al.*, 2010; Giskes *et al.*, 2011; Lee and Maheswaran, 2011; Caspi *et al.*, 2012; Carroll-Scott *et al.*, 2013). However, most studies were conducted in North America and Australia, while the proportion of studies conducted in Europe is more limited. It is necessary to generate evidence about the European urban environment since they differ markedly from the Australian or American context (Kasanko *et al.*, 2006), and particularly in Mediterranean cities such as Madrid.

1. HYPOTHESIS

The relationship of contextual determinants (physical and social) with health outcomes has been evidenced. In addition, it is clear that a geospatial approach improves the measurement of contextual determinants and improves the analyses in relation with CVD. Specifically, GIS allows developing indexes integrating different measures of the urban environment. With all this in mind, this PhD dissertation raises the following hypothesis:

GIS-based multicomponent indexes may capture the intertwined nature of the contextual determinants of CVD in a more complete fashion than single indexes.

2. OBJECTIVES

The main objective of this PhD research is to develop a GIS-based methodological proposal to characterize urban contextual determinants of CVD.

To achieve the main objective, it is necessary to achieve the secondary objectives listed below.

2.1 SECONDARY OBJECTIVES

SO1. To design a GIS-based multicomponent index to integrate information from food and physical activity environments to characterize the obesogenic environment of an urban area.

SO2. To build a socioeconomic deprivation index and to study its statistical and geographic stability through three different spatial observation units (census section, neighborhoods and districts).

SO3. To study the scale effect on the relationship between socioeconomic deprivation and prevalence of CVD.

SO4. To develop and implement a multicomponent index, the Heart Healthy Hoods index (HHH index), integrating characteristics of healthy urban environments for the heart (food, physical activity, tobacco, and alcohol) using GIS tools.

SO5. To examine the association of the HHH index with the prevalence of CVD at the area level.

To achieve the proposed general and secondary objectives, we have undertaken three separate and complementary studies that have been published in three different international and relevant health geography scientific journals. In the next section the data and methods used in these three studies are discussed.

3. SUMMARY OF DATA AND METHODS

Study 1. Characterizing physical activity and food urban environments: A GIS-based multicomponent proposal.

The study area comprised a socio-demographically average urban area of 12 contiguous census sections (≈16.000 residents), in Madrid. To assess food environment, 40 food stores were audited through the Nutrition Environment Measures Survey in Stores (NEMS-S). To assess the physical activity environment, 145 street segments were audited through the tool Systematic Pedestrian and Cycling Environment Scan (SPACES-M). Both tools were adapted to our research context.

Geospatial analysis was used to integrate the information of both environments and characterize the obesogenic environment. Using a Kernel Density Estimation (KDE), information punctual (food stores) and linear (street segments) became two continuous mathematical surfaces. This previous step allows integrating both surfaces in a single surface performing map algebra. The operation adopted for map algebra was a local unweighted average that computes an average of pixels at the same location in both surfaces. Finally, we used zonal analysis that calculates a single output value for each census section averaging all pixels that fall within each area.

Study 2. Geographic and statistic stability of deprivation aggregated measures at different spatial units in health research

For Madrid municipality a deprivation index was calculated with the 2011 census data provided by the Spanish Statistical Office. By Principal Components Analysis (PCA), socioeconomic and demographic indicators for the urban area were added in an index in three scales of analysis (census sections, neighborhoods and districts).

Spatial autocorrelation analysis (global and local) and cartographic representations were performed to evaluate the stability of the index through the three spatial scales. The global analysis was a Moran's Index and the local a bivariate Local Indicators of Spatial Association (or bivariate LISA). We conducted a Pearson correlation analyses to study the change in the relationship between deprivation and the CVD prevalence across the three spatial scales.

Study 3. A multicomponent method assessing healthy cardiovascular urban environments: The Heart Healthy Hoods Index

This study was also conducted in Madrid municipality. Using GIS methods, we generated two index models (model 0 unweighted and model 1 weighted) using the percentage of deaths for the main behavioral risk factors for CVD (diet, physical activity, alcohol, and tobacco environments). We performed global (Ordinal Least Square, OLS) and local (Geographically Weighed Regression, GWR) regression analyses to assess the relationship between both index models and CVD prevalence, and to identify the best index model.

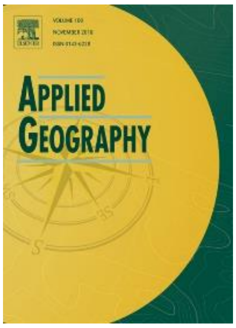
SECCTION III.
RESEARCH DEVELOPMENT

1. SCIENTIFIC PUBLICATIONS

Cebrecos, A., Díez, J., Gullón, P., Bilal, U., Franco, M., Escobar, F., 2016. Characterizing physical activity and food urban environments: A GIS-based multicomponent proposal. Int. J. Health Geogr. 15

Journal	**International Journal of Health Geographics**
Impact last 5 years (2016)	**3,199**
Quartile and position (2016)	**Q1 (10/157)**

Cebrecos, A., Domínguez-Berjón, M.F., Duque, I., Franco, M., Escobar, F., 2018. Geographic and statistic stability of deprivation aggregated measures at different spatial units in health research. Appl. Geogr. 95, 9–18.



Journal	**Applied Geography**
Impact last 5 years (2016)	**3,401**
Quartile and position (2016)	**Q1 (11/79)**

Cebrecos, A., Escobar, F., Borrell, LN., Díez, J., Gullón, P., Sureda X., Klein, O., and Franco, M., 2018. A multicomponent method assessing healthy cardiovascular urban environments: The Heart Healthy Hoods Index. Health and Place.

Journal	**Health & Place**
Impact last 5 years (2016)	**3,381**
Quartile and position (2016)	**Q1 (24/157)**

Study 1. Characterizing physical activity and food urban environments: A GIS-based multicomponent proposal

Cebrecos *et al. Int J Health Geogr (2016) 15:35*
DOI 10.1186/s12942-016-0065-5

International Journal of Health Geographics

METHODOLOGY

Open Access

Characterizing physical activity and food urban environments: a GIS-based multicomponent proposal

Alba Cebrecos[1,2], Julia Díez[1], Pedro Gullón[1,3], Usama Bilal[1,4], Manuel Franco[1,4] and Francisco Escobar[1,2*]

Abstract

Background: Healthier urban environments influence the distribution of cardiovascular risk factors. Our aim was to design and implement a multicomponent method based on Geographic Information Systems to characterize and evaluate environmental correlates of obesity: the food and the physical activity urban environments.

Methods: Study location comprised a socio-demographically average urban area of 12 contiguous census sections (≈16,000 residents), in Madrid, Spain. We conducted on-field audits on all food stores and street segments. We designed a synthetic index integrating continuous measures of both environments, by kernel density analyses. Index ranges from 0 to 100 (least-most healthy).

Results: We found a heterogeneous distribution with 75 and 50 % of the area scoring less than 36.8 and 25.5, respectively. Census sections of study area were categorized by Jenks intervals as high, medium–high, medium–low and low. 41.0 % of residents lived in an area with a low score, 23.6 % medium–low and 31.1 % medium–high and 4.2 % in a high.

Conclusion: The proposed synthetic index may be a relevant tool to inform urban health interventions, providing a feasible way to integrate different measures of barriers and facilitators of healthy urban environments in terms of food and physical activity.

Keywords: Synthetic index, Geographic Information Systems, Healthy food availability, Physical activity, Obesogenic environments

Background

The obesity epidemic is one of the main public health concerns of the present century [1]. Prevalence of overweight and obesity in European countries ranges from 45 to 67 %. Spain presents some of the highest levels of overweight (60.9 %) and obesity (23.7 %) in Europe [1].

The limited success of current individual-level based strategies shows the need for new approaches based on population-level determinants of obesity [2]. These approaches focus on affecting the fundamental causes [3] of the distribution of risk factors in the whole population [4]. These fundamental causes were called *mass influences* by Rose [4] and are mostly environmental or social factors at several levels. There is a large and renewed interest in these fundamental causes, especially at urban contexts, and particularly at neighborhood level [4–8].

Much of this renewed interest on neighborhood research in chronic diseases focuses on cardiovascular diseases, diabetes mellitus, and obesity [8–10], given that poor access to healthy foods and limited opportunities for physical activity are related to potentially health-relevant neighborhood physical and social environments. As the place of residence is associated with socioeconomic status, neighborhood characteristics can contribute significantly to health inequalities [9, 10]. Unlike other studies focusing on access inequalities to healthy areas as green spaces [11], healthy food environment [8] or health facilities [10],

*Correspondence: francisco.escobar@uah.es
Department of Geology, Geography and Environmental Sciences, University of Alcalá, Calle Colegios 2, 28801 Alcalá de Henares, Madrid, Spain
Full list of author information is available at the end of the article

our focus is based on neighborhood characterization as the next step within a wider research.

Body weight regulation depends on multiple factors, such as physical activity and healthy eating [12]. The contextual determinants of physical activity are complex and multifaceted, but can be roughly classified into transport-related physical activity and leisure-time physical activity influences. The determinants of active transportation relate to walking and biking and include features such as quality of pavements, safety, mix land use, destinations or connectivity [13]. Contextual influences of leisure-time or recreational physical activity include sports facilities and green spaces [14]. For this work we took into account the contextual determinants related to walkability. Con-textual determinants of healthy eating include all aspects of the local food environment that influence dietary behaviors [15]. Food stores and their associated accessibility and availability of healthy foods have been previously shown to affect dietary behaviors [16].

The literature on the associations between contextual determinants of physical activity and healthy eating is mixed. The diversity of methodologies used and the results obtained [17–19] highlight the complexity of the chain of causation linking contextual factors and different chronic diseases, as well as the challenges inherent on measuring complex social phenomena [20]. Among these challenges there is the intertwining of environmental features: physical activity environments and food environments are not isolated but rather the result of social forces that affect neighborhoods [21].

Much of the previous research has focused solely on one factor in isolation, such as walkability [22] or healthy food availability [23]. Moreover, the strong correlation between physical activity and dietary behaviors calls for strategies that tackle sedentary and unhealthy choices concurrently [24–26]. Interventions may be ineffective if only focused on promoting physical activity, ignoring a food environment which may promote unhealthy foods [24]. Thereby, there is a need of an integrated approach to understand contextual factors of both environments.

A potential promising avenue to operationalize the contextual determinants of obesity is to aggregate measures of both physical activity and diet determinants. Previous studies have aggregated urban context indicators in a synthetic index, finding significant correlations with health outcomes [18, 27, 28]. Kelly-Schwartz et al. [28], found a significant association between a composite index county sprawl index) and obesity, but not between their components and health outcomes [28].

Geographic Information Systems (GIS) are rap-idly becoming a relevant part of the panoply of methods adopted in Public Health research [29]. GIS is a well-suited tool to define healthy urban environments allowing to integrate data from different sources and scales, both spatial and non-spatial. Our objective is to design a multivariable tool based on GIS to integrate information from the physical activity and food environments to better characterize obesogenic environments in urban areas.

Methods

This study was conducted within the multidisciplinary Heart Healthy Hoods project [30]. The main objective of this European project is to analyse the impact of the physical and social urban environment in relation to residents' cardiovascular health in Madrid, Spain.

Study area

Madrid is the capital city of Spain, located in the central area of the country with a population of 3,186,595 habitants [31]. Madrid Metropolitan Region has around 6.5 million residents, the third-largest in Europe, after Lon-don and Paris. The City of Madrid is administratively divided into 21 districts which, in turn, are divided into 2412 census sections, the smallest administrative area for the Spanish Census (population~ = 1000–1500 per census section).

In order to conduct our study in an area that was not extreme in sociodemographic or urban form terms, we selected these 12 census sections using the Median Neighborhood Index (MNI) [32]. This method selects clusters of census sections that are on average closest to the median neighborhood in four variables: % above 65 years of age or older, % with low education, % foreign-born and population density. More details on this method can be found in Bilal et al. [32] supplementary material.

Study area is located in the southern part of the district of Ciudad Lineal, adjacent to the city ring road (M-30) (see Fig. 1). This area has an extension of approximately 42 hectares with a total population of 14,980 residents [31]. The study area was developed in the early twentieth century. At that time it was part of the municipality of Canillas and was not incorporated to the City of Madrid until 1955, when the main city incorporated all its sur-rounding municipalities [33]. The area has received a considerable influx of migration from other rural areas from Spain, especially coincident with the rural exodus of the 60s. Both urban morphology and building structure is relatively homogeneous throughout the area. Most buildings are residential and rank from three to nine stories being the vast majority five stories tall. Given the small size of each census section (~1000–1500 people), there is some random variability in socio-demographic as well as population density within the area, mostly related to the height of the buildings (and hence differential residential density).

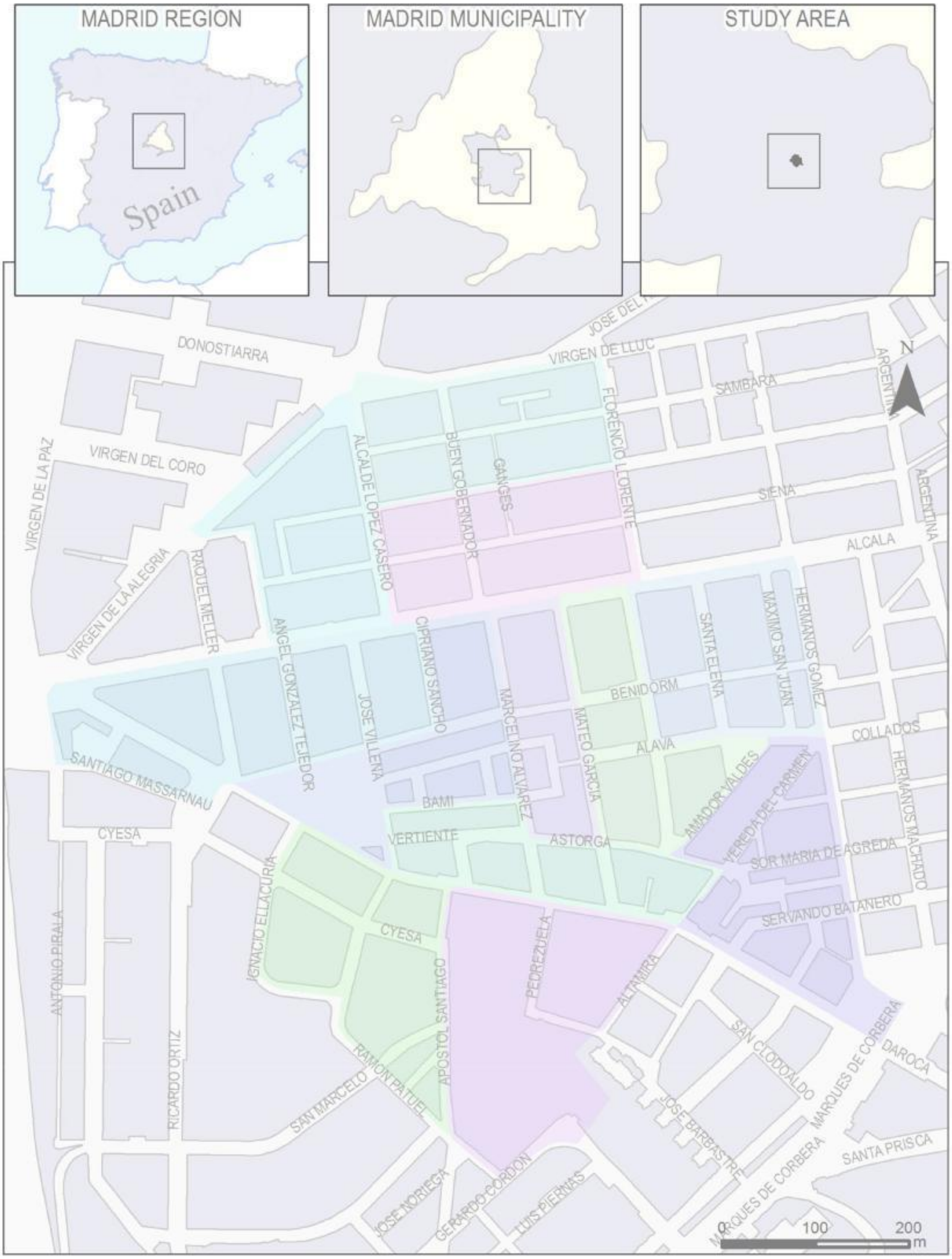


Fig. 1 Map of study area. *Colored areas* represent census sections within the pilot study area

According the 2014 Health Report of Madrid City Council [34], 41.2 % of the population presents over weight or obesity, 30.1 and 11.2 % respectively. Although these values are below national measures, 38.4 of overweight 18.2 % of obesity conforming the last National Health Survey, they values remain alarmingly high. Regarding risk factors, the *Risk Factors Surveillance System Associated with Noncommunicable Diseases in*

Adult Population of 2013 [35], indicated that 49.2 % of Madrid citizens affirm to be sitting most of their working time and that 73 % were inactive at leisure time. Regarding diet, Madrid citizens eat in average 1.2 rations of fresh fruit and 1.1 rations of vegetables per day. These rates are far from the recommendations of 5 rations per day [36].

Characterizing the urban environment

Food environment

We measured healthy food availability by conducting in-store audits within all food stores present in the study area. We found 40 retail food stores within the selected census sections, which were classified by store type as: corner stores (small stores with a low variety of items and generally no fresh products); grocery stores (mid-sized stores with higher variety and presence of fresh products); supermarkets (large stores with highest variety and presence of fresh products); specialty stores (greengrocers, butcheries, fishmongers and bakeries); gas stations; and convenience stores (long opening hours and no presence of fresh products). More details of these measurements properties have been published in Bilal et al. [32].

Trained data collectors used an abbreviated version of the Nutrition Environment Measures Survey in Stores (NEMS-S), developed and validated by Glanz et al. [23]. The NEMS -S has been used in several contexts, including the US [37] and Brazil [38]. The abbreviated version was developed by the Johns Hopkins Center for a Liv-able Future for an assessment of Baltimore's Food Environment [39]. This instrument examines the availability of healthy options versus less-healthier options over 12 types of foods, such as skim/low-fat milk (vs whole milk), 100 % fruit juice (vs juice drinks), lean ground beef (vs regular), skinless chicken (vs regular), whole grain bread (vs refined bread), or low-regular cereals, as examples [23]. From these surveys, we produced a Healthy Food Availability Index (HFAI) for each food store. There-fore, we looked at 12 food groups: milk, juice, fruits, vegetables, meats, chicken, seafood, canned goods, frozen foods, packaged foods, bread and cereals. The HFAI score in this study could range from 0 to 27.5 points, with a higher score indicating a greater availability of healthy foods [37].

Physical activity environment

The Systematic Pedestrian and Cycling Environment Scan (SPACES) [40] is an observational audit of features of the built environment that can influence walking and cycling along a street network. We adapted this audit tool to the Madrid (M-SPACES) environment and conducted a validity and reliability study before [41]. For the purposes of this study, and due to the residual use of bikes in the area, only walkability measures were considered.

A trained researcher audited all street segments of the study area (n = 145 segments) by foot. A street segment is defined as one section of a street that runs between two intersections. It is often used as the basic observation unit in neighborhood or community analysis. Items of the M-SPACES tool are then added up to four domains: functionality, safety, aesthetics and destinations. These, in turn, can be added to compute a walkability score for each street segment (ranging from 0, least walkable, to 1, most walkable). Main audited characteristics were functionality, safety, aesthetics and destinations. More details on this audit tool and its measurements properties have been published before in Gullón et al. [41], and Bilal et al. [32].

Spatial datasets

Contextual information on the study area was collected from the Spanish National Mapping Agency and Spanish National Spatial Data Infrastructure, allowing us to generate a georeferenced database to integrate and map the results from the food and physical activity environment assessment. Administrative boundaries (district and census sections) and street networks were collected in vector polygon and line formats, respectively. We also used orthophotograpy of the study area obtained from the Orthophotography Air National Plan.

ArcGIS 10.1 software was used to integrate, standardize and manage these datasets. First, all information was projected to a common system (ETRS89 UTM 30N). The physical activity environment data (collected with the M-SPACES tool) was associated with the street network layer by a relational join. The food environment data (collected with the abbreviated NEMS -S tool) was integrated in the system using a point-based layer with a relational join. All other layers (administrative boundaries, blocks and orthophotos) were introduced to the final maps as reference information.

Geospatial analysis

The aim of this study was to integrate data on the physical activity and food environment in characterizing the urban environment by using GIS. Figure 2 summarizes our approach. In summary, we converted line and point data, linked to physical activity and food respectively, into surface-based data on the whole study area, as a means of facilitating integration of both environments into a single surface. Only after the measures of both environments were known at pixel level (the minimum spatial unit of the

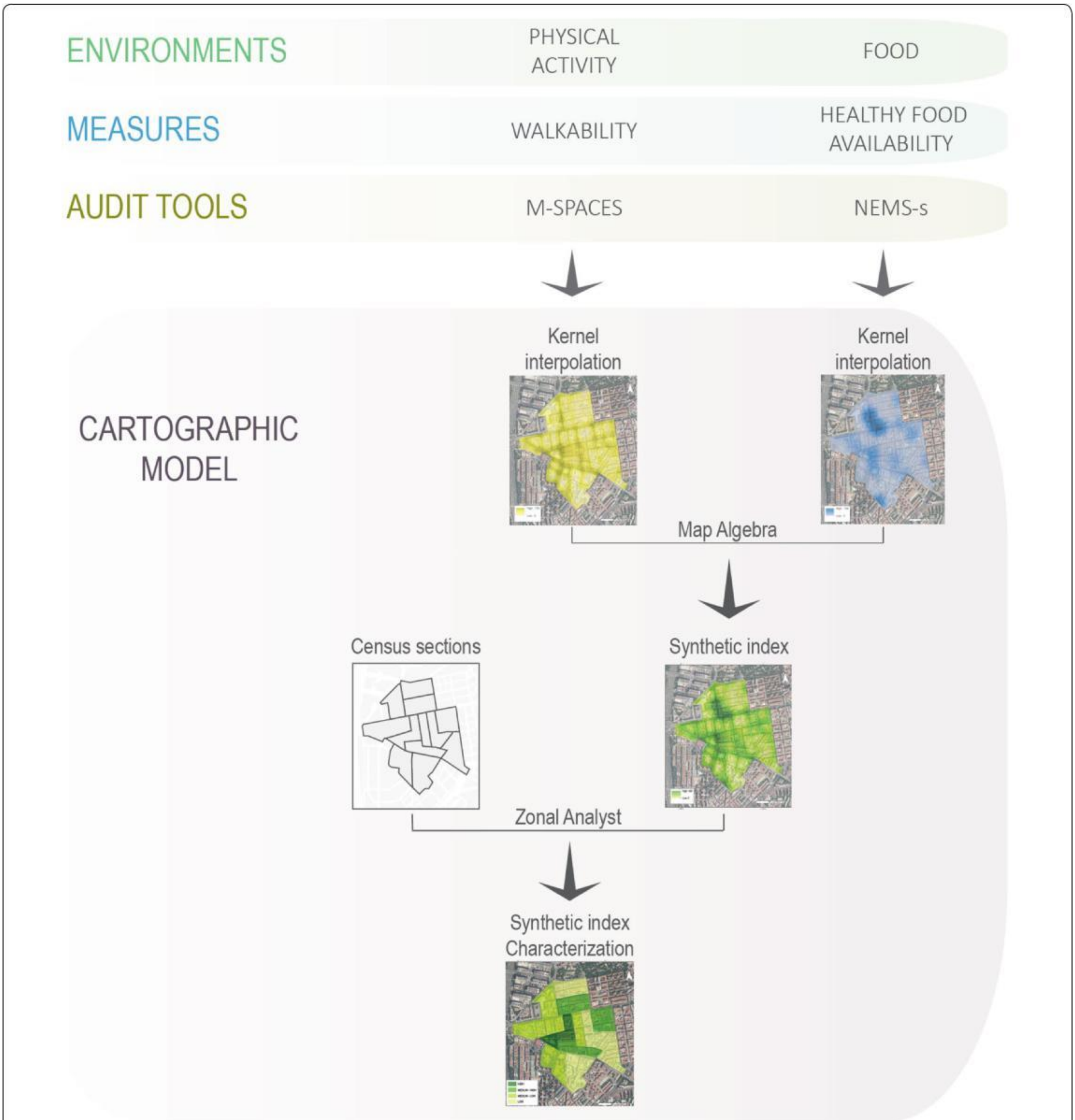


Fig. 2 GIS model for the construction of a spatial integrated index on walkability and food environment. Kernel interpolation of the Healthy Food Availability index (HFAI) was created with data from direct observation of food stores measured with the Nutrition Environment Measurement Surveys in Stores (NEMS-S) audit tool. Kernel interpolation of walkability was created with data from direct observation of each street segment measured with the Madrid Pedestrian and Cycling Environment Scan (M-SPACES) audit tool

newly created surface), a map algebra-based arithmetic operation to combine both measurements was possible. Finally, a categorization of the combining results was applied in order to ease interpretation.

These steps are detailed as follows: first, line data (walkability index for each street segment) and point data (healthy food availability index for each food store) was extended to the whole study area by applying kernel density estimation (KDE), resulting on a pixel-based surface. Both walkability and health food availability indices could have been kept under their original geometric form, line and point respectively. However, a surface-based approach was adopted in order to facilitate future data integration with additional data. Being most statistics aggregated under administrative boundaries, the integration of line or point-based data presents more inconveniences than the assumed ones produced by the extension of the information to the whole surface. In addition, an eventual proposal of a combined index including information on green areas reinforced the surface-based solution.

KDE fits a mathematical surface (composed of pixels) with a normal distribution over each point based on (a) the value empirically collected for each point, and (b) the distance from each location in the surface to all points in the area within defined radius or bandwidth. Essentially, the value of each point is smoothed over the study area producing a density value that will be the highest at the location of every point, and decaying from there with distance using a defined bandwidth [42]. KDE is widely adopted in spatial analysis where input data present different geometric forms and it is of interest the integration of such data with other variables collected on the same territory.

We use de KDE integrated in ArcGis 10.1 software which employs the quadratic Kernel function of Silverman [43]:

$$\hat{f}(x) = \frac{1}{nh} + \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

where K is the quadratic Kernel function defined by $K(x) = \frac{3}{4}\,(1 - x^2), x \leq 1$, "x" is the point at which density is estimated, "x_i" is the value of the variable in the case "i", "n" is the number of cases and "h" is the bandwidth. The basic idea consists calculated for specific points, the averaged sum (hence the estimator involves summing over "n" and then divide by this value) of Kernels centered on the observations.

This spatial analysis allows weighting each component by their associated attributes, in our case the HFAI and M-SPACES scores. For example, if the component has associated value attribute equal to 3, the case counts as 3 cases. Thus, density value in each pixel of the output image is calculated summing the values of all overlapping kernel surfaces. All surfaces were generated with a pixel size of 3 m. We used a bandwidth of 100 m, given that the average distance from one food store to the closest food store was around half that length (improving smoothing). A static bandwidth was used because of the small study area and the homogeneous population density distribution [42].

The cartographic model presented in Fig. 2 shows the development of both continuous density surfaces: one from the food stores layer weighted by the value of HFAI; and the other from the street segments layer weighted by the scores obtained from the M-SPACES audit.

After generating both surfaces, we performed a map algebra analysis. First, we homogenized data in a range from 0 to 100, to make them comparable with each other. The operation adopted for map algebra was a local unweighted average that computes an average of pixels at the same location in both the physical activity and food environment surfaces, treating both environments with equal weight, generating the synthetic index. To fully integrate the synthetic index into the geographic con-text of the area, we assigned each census section an obesogenic (synthetic index) value. For this, we used zonal analysis that calculates a single output value for each census section averaging all pixels that fall within each area. To improve the interpretability of our results, we categorized census sections into four classes according to their value in the synthetic index (high, medium–high, medium–low and low). For this, we used the Jenks intervals (or natural breaks) approach that reduce the variance within classes, while maximizing the variance between them.

Results

Figure 3 shows the calculated KDE surface obtained for the food and physical activity environment. Regarding the food environment, the figure shows a concentration of food sales scored with high HFAI values in the North and South ends of the study area, with patches of medium–high density of HFAI distributed throughout the area. Most stores with high HFAI were quite close to each other and mostly located along important roads, creating "islands" of healthy foods. Stores with low HFAI were distributed more evenly creating "healthy food deserts". Regarding the physical activity environment, the surface resulted from the M-SPACES showed highest values at streets intersections, on streets with wide side-walks, and in the surroundings of squares and parks. In consequence, the greater the number of intersections, the greater the walkability of the area.

The synthetic index surface resulted from averaging the food and physical activity environment is depicted in Fig. 4. Figure 5 shows the distribution of values of the synthetic index. This distribution is right skewed, with a median score of 25.4 (IQR 15.4–36.9) and a mean score of 27.7. Around 75 % of the area is below 36.8 of the index score and half of the area below 25.5.

Another result is obtained from zonal analysis, where we mapped the 12 census sections of the area in Fig. 6 with single output value for each census section averaging all pixels that fall within each area. Characterization created by using natural Jenks grouped the census sections into four categories about themselves according to the average score: low (17.7–21.6), medium–low (21.7–30.8), medium–high (30.9–35.1) and high (35.2–43.8). Four out

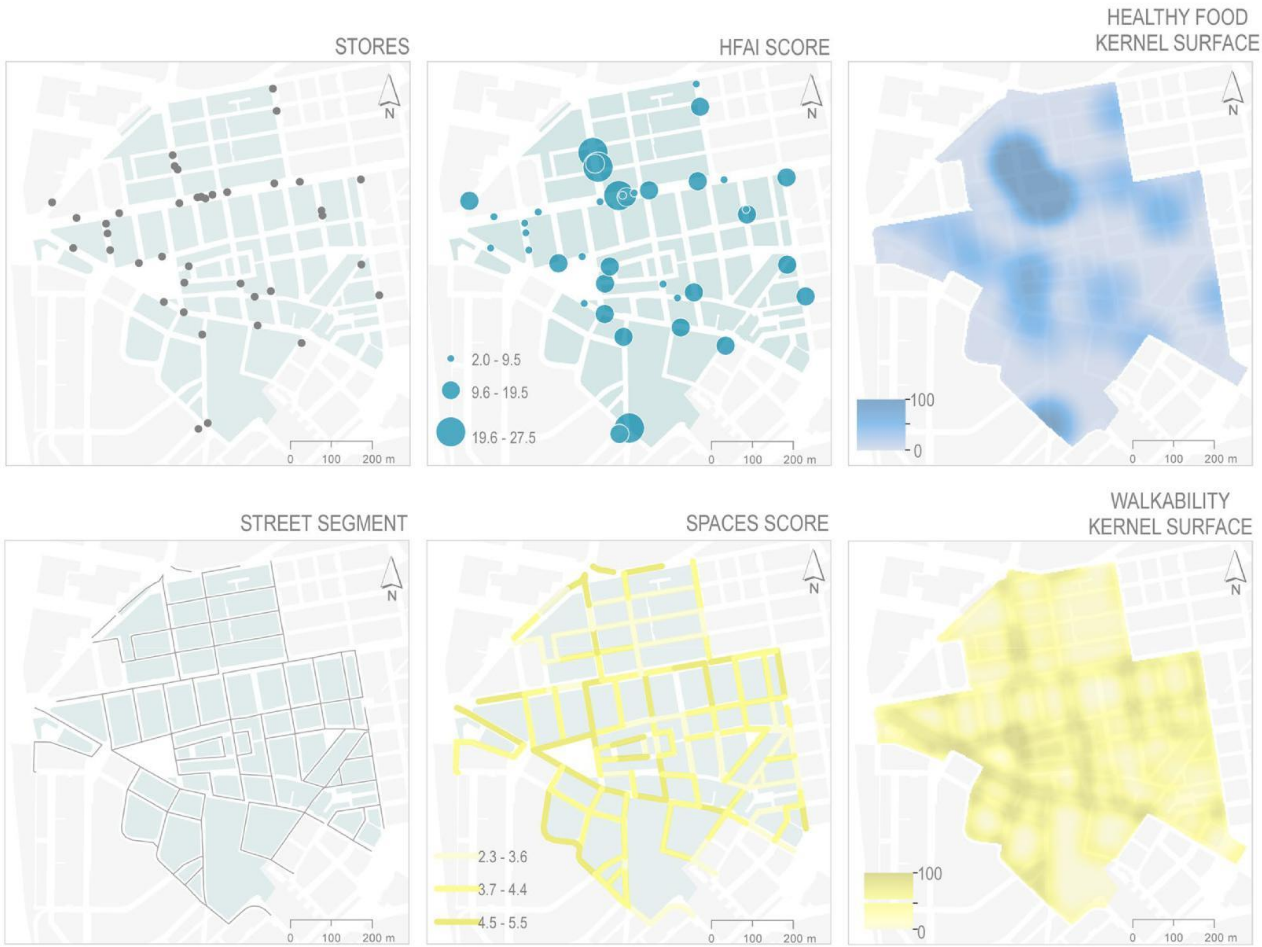


Fig. 3 Development of continuous KDE variables from direct observation data. On the *left*, is depicts the calculating for the food environment starting with the location of all stores, continues with the Healthy Food Availability Index (HFAI) score for each one, and the next is the KDE surface weighted by HFAI score. On the *right*, the development for the physical activity environment. Above the location of all the street segments, continues with Pedestrian and Cycling Environment Scan (SPACES) score by each one. And in front the KDE weighted by the SPACES

of the 12 census sections are classified as low, 4 medium–low, 3 medium–high and 1 as high.

Table 1 shows basic sociodemographic characteristics of the twelve census sections and the entire study area. Around 4.2 % of the population live in areas characterized as "healthy" environments (defined as a "high" synthetic index), while 41.0 % of the residents live in an area with the lowest rating. Analysing the results by sex, 40.8 % of women have lower scores than men (41.3 %). 4.6 % of women and 3.7 % of men live in a section with high score. In the case of foreign-born residents, 49.3 % of them live in the unhealthiest areas and 3.0 % in the healthier. If results are studied by age, the majority of young people live in a census section with low score (51.5 %) as well as adult people (44.4 %) but in the case of elderly people, they live in a high healthy space (27.7 %). Only the 1.9, 3.5 and 7.0 % of young, adult and elderly people respectively live in a healthy section.

Discussion

This paper documents the development of an innovative method to assess the obesogenic environment by using a synthetic index that integrates continuous measures of both food and physical activity environments generated by KDE. The results show a heterogeneous distribution of obesogenic determinants in the study area. 36.5 % of the census sections have a low synthetic index value, followed by medium–low healthy (33.7 %), and medium– hig (22.4 %). Only one census section falls under the category of high value in the synthetic index, representing only 4 % of the study area.

This healthy census section is delimited by the main streets of the neighborhood, where healthiest food stores were present. Main streets are also designed to be more walkable and have more intersections. Moreover, while we did not measure parks, the only park located in the study

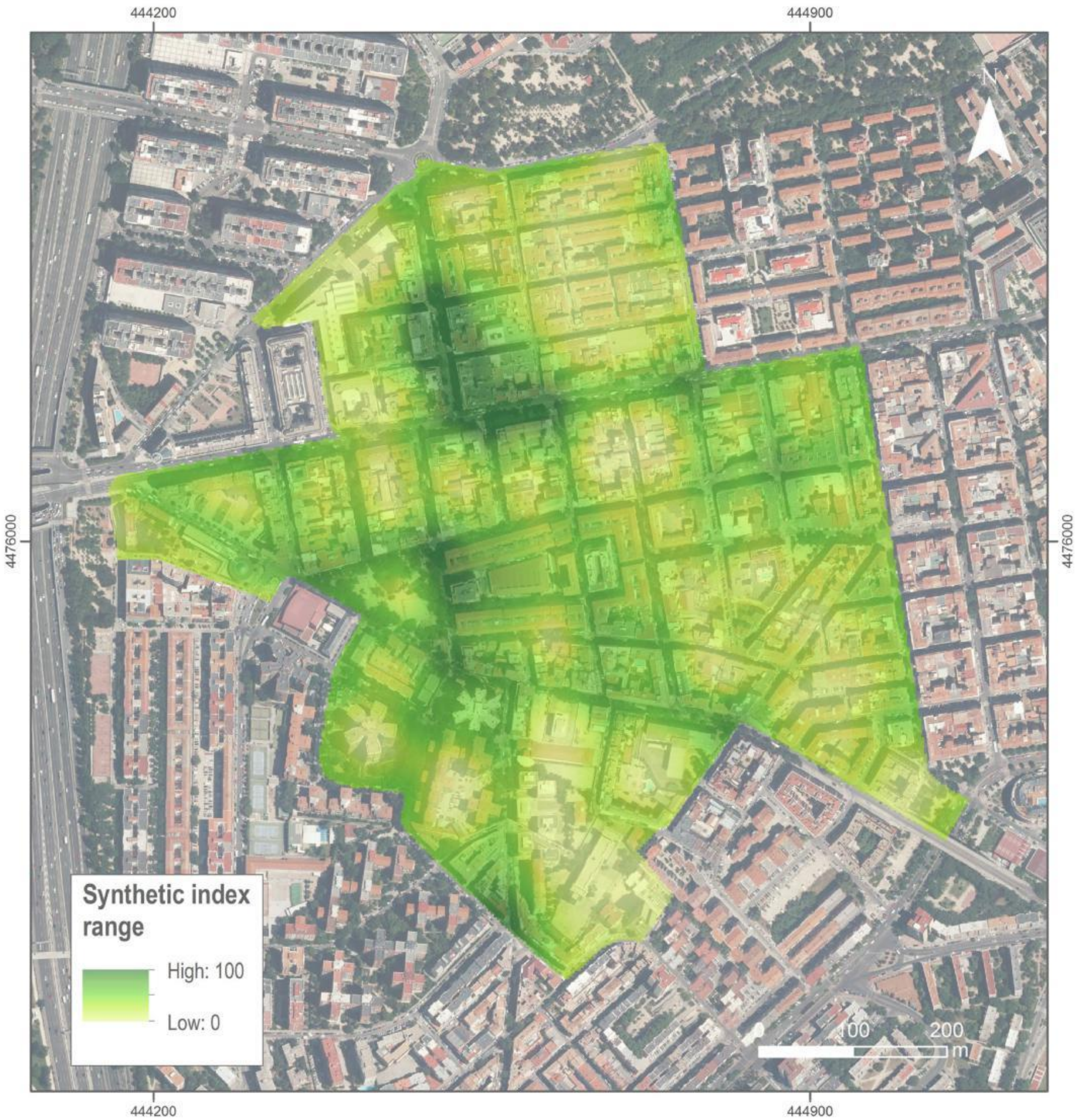


Fig. 4 Continuous synthetic index surface. This surface is the local average of the pixels of walkability KDE and the pixels of food availability KDE. The size of pixel is 3 × 3 m and the bandwidth selected to the smoothing was 100 meters

area, was also located within this census section. The census sections with the lowest synthetic index value were located in the east of the study area. Food store density is smaller with 3 corner stores, 1 bakery and 1 small supermarket, and we found narrow streets residential inter-block.

In order to understand the obesogenic environment it is necessary to consider the interrelations between the food and physical activity environments, as built environment metrics are correlated with each other [44]. The use of composite indices reduces collinearity and over-adjustment, confers ease of interpretation, and may reduce measurement errors [18]. Besides, integrating different indicators within an index can detect associations not previously found [28]. In our case, a systematic observation of a built environment, using validated tools,

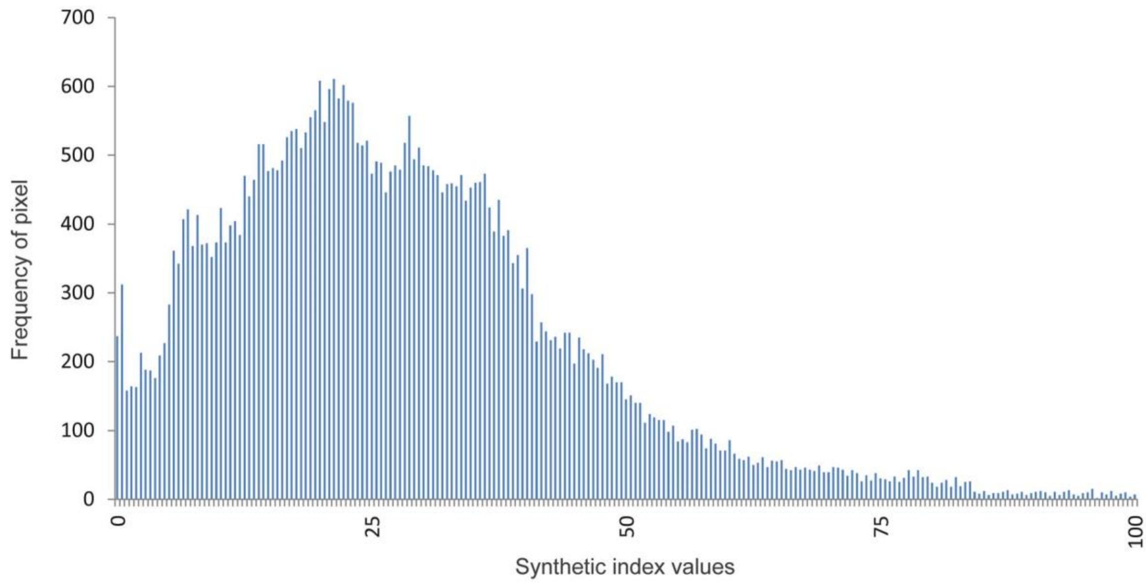


Fig. 5 Histogram of synthetic index surface. It depicts the frequency of pixel values of the study area with a range from 0 to 100 with higher scores indicating a healthier environment

provided highly detailed spatial data. This also ensured variability on measures of both constructs and statistical power. The use of an extensive sampling strategy to maximize the variation between environmental factors reduced the sample sized needed to assess associations between built environments and obesity outcomes [45].
Previous studies have considered both physical activity and food environments to characterize environ-mental obesogenity, but have not obtained a composite score [25, 27, 45]. These urban environment measures have been used to evaluate their relation with diabetes incidence [46] or cardiometabolic risk factors [47]. The method described here considers previously studied variables, such as food store density, food store type, street intersection density, parks or street aesthetic, among others. The majority of these studies used GIS to integrate all information from diverse sources, mostly from secondary databases. On top of these, our study adds other variables as availability of healthy foods captured by NEMS-s or the aesthetic or safety domains measured by M-SPACES tool, which are very difficult to assess from secondary administrative databases.

KDE remains underutilized when compared to proximity analysis or to analysis over defined statistical areas [48–51], although the number of examples using KDE technique to study the obesogenic environment has increased in recent years [48–51]. KDE overcomes the limitations of binary definitions present in analysis based in fixed geographic boundaries (for example, number of stores per census section). Smooth transitions across (administratively defined) boundaries represent the reality of urban environments better [52]. The resulting KDE surface can then be used as an independent variable on statistical models [42].

Our study was conducted in the City of Madrid, Spain. In Spain, the smallest administrative level where data is publicly available is the census section, composed of ≈1500 people. Our study area is made up of 12 census sections, although the estimation of our synthetic index creates a smoothed surface over the entire study area, regardless of census section boundaries. This method is therefore replicable in other settings where the administrative spatial hierarchies are different, as long as data is collected at the appropriate level. This method is also replicable at larger units, like municipalities or countries, taking always into account the effort associated with data collection at any level.

The proposed method has several limitations. First, it requires primary data collection through systematic observation, which is a resource and time intensive process. Thanks to advances in Geographic Information Technologies, these costs can be drastically reduced, by using available secondary databases with spatial information and new geographic remote devices to collect geocoded primary data [29]. Second, this work has not considered the relative importance of the two domains with respect to each other, treating both environments with equal weight. The controversy regarding the quantification of the proportion of food or physical activity responsible for the obesity epidemic is still very much alive [53, 54]. We decided to adopt a local unweighted average, but any study using this method to estimate the associations

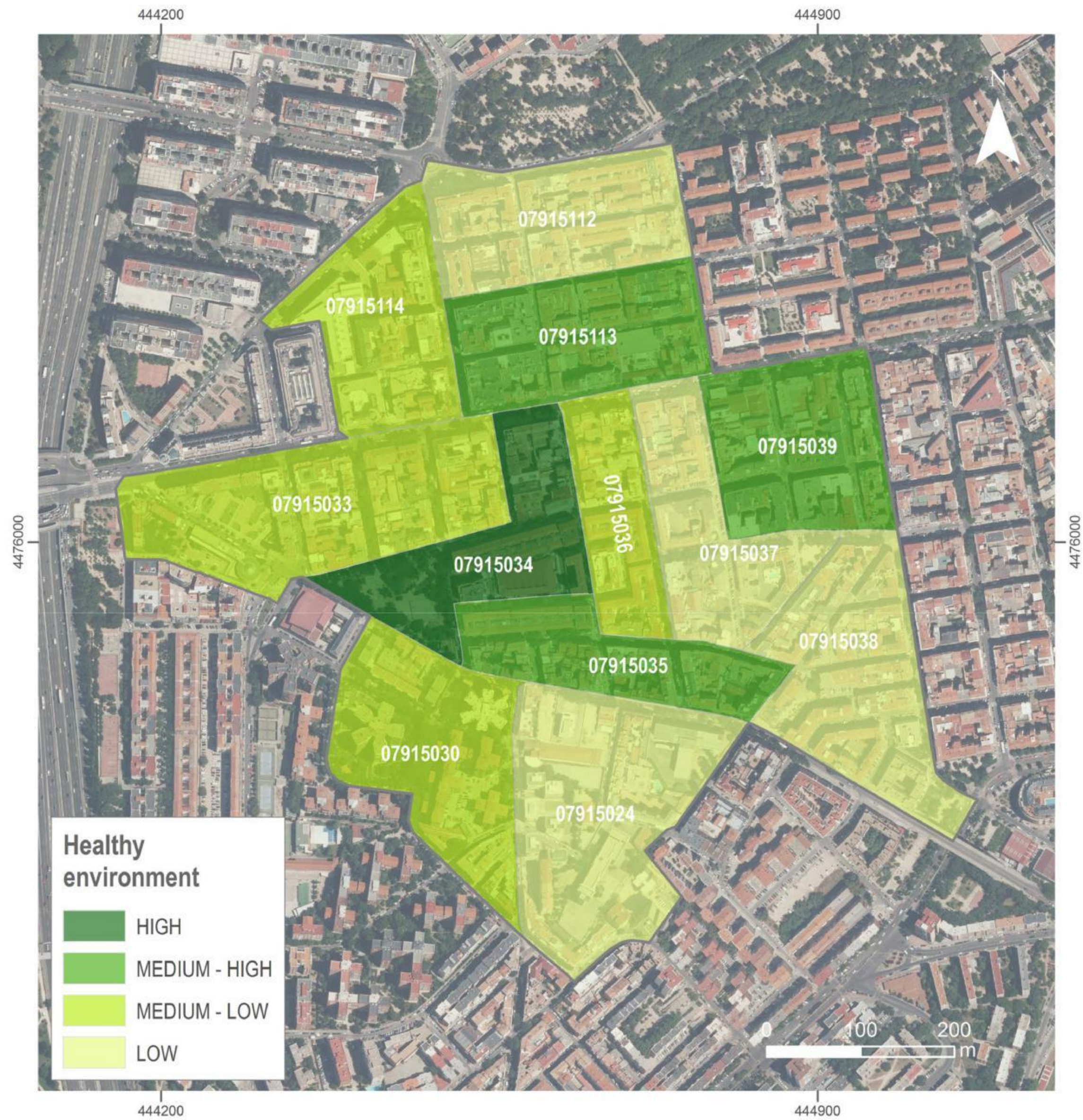


Fig. 6 Study area characterization at census section level. Is the result of zonal analysis of each administrative area having in account all the pixels of the local analysis within each area

between obesogenic environment and health outcomes should consider sensitivity analysis that alter these weighting decisions.

Conclusion

The proposed synthetic index provides a feasible way to integrate different measures of physical barriers and promoters of healthy urban environments. This method opens new ways to capture inter-relations between physical activity and health food availability urban environment domains that did not emerge when they were studied in an isolated way. Thus, applying this index is a preliminary step to promote healthier urban environments and bridge the health inequalities present in large cities like Madrid.

The proposed index, and the cartography associated with it, may be useful tools to inform future research and urban health recommendations.

Table 1 Description of study area population by census section and group. Source: 2011 census data

Census section	7915024	7915030	7915033	7915034	7915035	7915036	7915037	7915038	7915039	7915112	7915113	7915114	Total
Population (%)	980 (6.5)	1110 (7.4)	1265 (8.4)	635 (4.2)	1205 (8.0)	540 (3.6)	133 (8.9)	2145 (14.3)	1980 (13.2)	1680 (11.2)	1480 (9.9)	625 (4.2)	14,980
Women (%)	555(6.5)	600 (7.0)	780 (9.1)	395 (4.6)	670 (7.8)	320 (3.7)	750 (8.7)	1210 (14.1)	1135 (13.2)	980 (11.4)	865 (10.1)	315 (3.7)	8575(57.2)
Foreign born (%)	130(7.1)	150 (8.2)	70 (3.8)	65 (3.6)	60 (3.3)	65 (3.6)	125 (6.8)	395 (21.6)	440 (24.1)	405 (22.2)	180 (9.9)	55 (3.0)	2140 (14.3)
Years < 16 (%)	180(13.7)	60 (4.6)	135 (10.3)	25 (1.9)	130 (9.9)	70 (5.3)	125 (9.5)	230 (17.6)	105 (8.0)	140 (10.7)	40 (3.1)	70 (5.3)	1310 (8.8)
Years > 65 (%)	180(4.9)	310 (8.5)	435 (11.9)	255 (7.0)	300 (8.2)	160 (4.4)	250 (6.9)	350 (9.6)	505 (13.9)	230 (6.3)	395 (10.8)	275 (7.5)	3645 (24.3)
Area in km^2 (%)	0.05 (11.9)	0.04 (9.5)	0.05 (11.9)	0.03 (7.1)	0.03 (7.1)	0.02 (4.7)	0.03 (7.1)	0.04 (9.5)	0.03 (7.1)	0.03 (7.1)	0.03 (7.1)	0.03 (7.1)	0.42
Pop. density inhab/km^2	19,600	27,750	25,300	21,167	40,167	27,000	44,500	53,625	66,000	56,000	49,334	20,834	35,667
Healthy environment	Low	Medium-Low	Medium-Low	High	Medium-Low	Medium-Low	Low	Low	Medium-Low	Low	Medium-Low	Medium-Low	

Abbreviations

GIS: Geographic Information System; MNI: Median Neighborhood Index; NEMS-S: Nutrition Environment Measures Survey in Stores; HFAI: Healthy Food Availability Index; M-SPACES: Madrid Systematic Pedestrian and Cycling Environment Scan; KDE: kernel density estimation.

Authors' contributions

AC conducted the geospatial analyses and drafted the manuscript. JD collected food environment data and contributed to manuscript preparation. PG collected physical activity environment data and contributed to manuscript preparation. UB provided revisions to the manuscript. MF conceived the Heart Healthy Hoods project and revised the manuscript. FE oversaw spatial analytical results and revisions the manuscript. All authors read and approved the final manuscript.

Author details

[1] Social and Cardiovascular Epidemiology Research Group, School of Medicine, University of Alcalá, Alcalá de Henares, Madrid, Spain. [2] Department of Geology, Geography and Environmental Sciences, University of Alcalá, Calle Colegios 2, 28801 Alcalá de Henares, Madrid, Spain. [3] Escuela Nacional de Sani- dad, Instituto de Salud Carlos III, Madrid, Spain. [4] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

Acknowledgements

We would like to thank the whole Social and Cardiovascular Epidemiology Research Group, for his collaboration and ideas on the study.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the European Research Council under the Euro- pean Union's Seventh Framework Programme (FP/2007-2013)/ERC Starting Grant HeartHealthyHoods Agreement n. 336893. Usama Bilal was supported by a fellowship from the Obra Social La Caixa and by a Johns Hopkins Center for a Livable Future-Lerner Fellowship.

Received: 31 May 2016 Accepted: 28 September 2016

Published online: 04 October 2016

References

1. WHO. The European health report 2015. Targets and beyond—reaching new frontiers in evidence. W.H.O.R.O.f. Europe, Editor 2015. p. 18.
2. Lake A, Townshend T. Obesogenic environments: exploring the built and food environments. J R Soc Promot Health. 2006;126(6):262–7.
3. Link B, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav. 1995;35:80–94.
4. Rose G. Sick individuals and sick populations. Int J Epidemiol. 1985;14(1):32–8.
5. Diez Roux A. Neighborhoods and health: where are we and were do we go from here? Rev Epidémiol Santé Publique. 2007;55(1):13–21.
6. Franco M, Bilal U, Diez-Roux A. Preventing non-communicable diseases through structural changes in urban environments. J Epidemiol Com- munity Health. 2015;69(6):509–11.
7. Hirsch J, et al. Change in walking and body mass index following residential relocation: the multi-ethnic study of atherosclerosis. Am J Public Health. 2014;104(3):e49–56.
8. Díez J, et al. Understanding differences in the local food environment across countries: a case study in Madrid (Spain) and Baltimore (USA). Prev Med. 2016;89:237–44.
9. Rosario AM, Pérez-Stable EJ. Making neighborhoods good for your health. Circulation. 2016;134(7):514–6.
10. Steiner R, et al. A76 Inequalities in the distribution of primary care providers: comparing healthcare disparity among urban and rural Floridian. J Transp Health. 2015;2(2, Supplement):S44–5.
11. Hillsdon M, et al. The relationship between access and quality of urban green space with population physical activity. Public Health. 2006;120(12):1127–32.
12. Bethlehem J, et al. The SPOTLIGHT virtual audit tool: a valid and reliable tool to assess obesogenic characteristics of the built environment. Int J Health Geogr. 2014;13(1):52.
13. Pikora T, et al. Developing a framework for assessment of the environmental determinants of walking and cycling. Soc Sci Med. 2003;56(8):1693–703.
14. Pikora T, et al. Neighborhood environmental factors correlated with walking near home: using SPACES. Med Sci Sports Exerc. 2006;38(4):708–14.
15. Franco M, Bilal U, Diez J. Food Environment. In: Caballero B, Finglas P, Toldrá F, editors. The encyclopedia of food and health, vol. 3. Oxford: Academic Press; 2015.
16. Franco M, et al. Availability of healthy foods and dietary patterns: the multi-ethnic study of atherosclerosis. Am J Clin Nutr. 2009;89(3):897–904.
17. Cobb L, et al. The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results. Obesity (Silver Spring). 2015;23(7):1331–44.
18. Feng J, et al. The built environment and obesity: a systematic review of the epidemiologic evidence. Health Place. 2010;16(2):175–90.
19. Leal C, Chaix B. The influence of geographic life environments on cardio-metabolic risk factors: a systematic review, a methodological assessment and a research agenda. Obes Rev. 2011;12(3):217–30.
20. Glass T, McAtee M. Behavioral science at the crossroads in public health: Extending horizons, envisioning the future. Soc Sci Med. 2006;62(7):1650–71.
21. Guthman J. Too much food and too little sidewalk? problematizing the obesogenic environment thesis. Environ Plan A. 2013;45(1):142–58.
22. Owen N, et al. Neighborhood walkability and the walking behavior of Australian adults. Am J Prev Med. 2007;33(5):387–95.
23. Glanz K, et al. Nutrition Environment Measures Survey in Stores (NEMS-S): development and evaluation. Am J Prev Med. 2007;32(4):282–9.
24. Economos C, et al. Food and physical activity environments: an energy balance approach for research and practice. Am J Prev Med. 2015;48(5):620–9.
25. Meyer K, et al. Combined measure of neighborhood food and physical activity environments and weight-related outcomes: the CARDIA study. Health Place. 2015;33:9–18.
26. Jones A, et al. Tackling obesities: future choices–obesogenic environments evidence review. London: Government Office for Science; 2007.
27. Wall M, et al. Patterns of obesogenic neighborhood features and adolescent weight: a comparison of statistical approaches. Am J Prev Med. 2012;42(5):e65–75.
28. Kelly-Schwartz A, et al. Is sprawl unhealthy? a multilevel analysis of the relationship of metropolitan sprawl to the health of individuals. J Plan Educ Res. 2004;24(2):184–96.
29. Matthews S, Moudon A, Daniel M. Work group II: using Geographic Information Systems for enhancing research relevant to policy on diet, physical activity, and weight. Am J Prev Med. 2009;36(4 Suppl):S171–6.
30. HHH. Heart Healthy Hoods project. 2014 (Cited 12 Dec 2015). https://hhh-project.eu/.
31. INE. Censos de Población y Viviendas 2011. 2016 (Cited 27 Nov 2015). http://www.ine.es/censos2011_datos/cen11_datos_inicio.htm.
32. Bilal U, et al. Population cardiovascular health and urban environments: the Heart Healthy Hoods exploratory study in Madrid, Spain. BMC Med Res Methodol. 2016;16(1):1–12.
33. Terán FD. Historia del urbanismo en España III: Siglos XIX y XX. Cátedra. Madrid; 1999.
34. Díaz-Olalla J, Benítez-Robredo T. Estudio de Salud de la Ciudad de Madrid 2014. Madrid: Madrid Salud; 2015.
35. Díez-Gañán L. Hábitos de salud en la población adulta de la Comunidad de Madrid, 2013. Resultados del Sistema de Vigilancia de Factores de Riesgo Asociados a Enfermedades No Transmisibles en población adulta (SIVFRENT-A), in Boletín Epidemiológico de la Comunidad de Madrid 2014, C.d.S. Madrid: Servicio de Epidemiología, Editor 2013. p. 45.
36. WHO, FAO. Diet, nutrition and the prevention of chronic diseases: report of a Joint World Health Organization/Food and Agriculture Organization Expert Consultation. WHO Technical Report Series. 2003. p. 149.
37. Franco M, et al. Neighborhood characteristics and availability of healthy foods in Baltimore. Am J Prev Med. 2008;35(6):561–7.
38. Franco M, et al. Neighborhood characteristics and availability of healthy foods in Baltimore. Am J Prev Med. 2008;35(6):561–7.

Study 2. Geographic and statistic stability of deprivation aggregated measures at different spatial units in health research

Applied Geography 95 (2018) 9–18

Contents lists available at ScienceDirect

Applied Geography

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

Geographic and statistic stability of deprivation aggregated measures at different spatial units in health research

Alba Cebrecos[a,b,c], M. Felícitas Domínguez-Berjón[c,d], Ignacio Duque[c,e], Manuel Franco[a,f], Francisco Escobar[a,b,*]

[a] Social and Cardiovascular Research Group, University of Alcala, Spain
[b] Dept. of Geology, Geography and Environmental Sciences, University of Alcala, Spain
[c] Group of Social Determinants of Health. Spanish Society of Epidemiology, Spain
[d] Dirección General de Salud Pública, Consejería de Sanidad, Madrid, Spain
[e] Subdirección General de Tecnologías de la Información y las Comunicaciones, Instituto Nacional de Estadística, Madrid, Spain
[f] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, USA

ARTICLE INFO

Keywords:
Modifiable areal unit problem
Socioeconomic factors
Cardiovascular disease
Geographic scale

ABSTRACT

Deprivation indices constitute a valuable tool for assessing health inequalities. A key issue when analyzing deprivation is the choice of the geographical scale and spatial unit of analysis. Our objective was to evaluate statistical and geographical stability of an Area Based Deprivation Index (ABDI) computed at different spatial scales and to study their relation with cardiovascular disease.

The present study has been conducted in the city of Madrid, Spain. Madrid divides its territory in three different administrative units nested within each other: census section, neighborhoods and districts. For each unit a deprivation index was calculated through Principal Component Analysis (PCA). The data source was the 2011 national census from where a range of socioeconomic and demographic indicators were selected. To study statistical and geographical stability of deprivation we used an Exploratory Spatial Data Analysis and bivariate Local Indicators of Spatial Association analysis. We also conducted Pearson correlation analyses to study the change in the relationship between deprivation and the prevalence of cardiovascular disease (CVD) across the three scales.

At census section and neighborhood level, first component showed four and five factors loading higher than 0.6, respectively. These factors loading related to occupancy/labor market and education. However at district level, first component showed seven factors loading higher than 0.6 and related to occupancy/labor market, education and immigration. With indicators of these factors loading, deprivation indices were calculated for each administrative unit by extracting a single PCA axis. Variance explained for each index was 65%, 86% and 79%, respectively. Bivariate local autocorrelation analyses showed aggregated areas of low and high stability with variable degree of significance in the three scales. The ABDIs calculated at census section level, neighborhood level and district level presented different significant correlations with CVD prevalence ($r = 0.328$; $r = 0.635$; and $r = 0.739$ respectively). These results show that the deprivation index did not remain stable across the three scales, neither were the correlations between deprivation and age-adjusted CVD prevalence.

Understanding the stability of a spatial phenomenon across different scales is essential to determine the best unit of aggregation of data when studying an important process such as socioeconomic deprivation and its possible health impacts.

1. Introduction

Mortality and morbidity increase as the social position of a population decreases. This progressive and lineal phenomenon was defined by Marmot as social gradient (Marmot, 2005). This is a universal phenomenon, although the magnitude and extent may vary between countries and change overtime (Marmot, 2005; Sir Michael, 2006). In Nova Scotia (Canada) 35,266 premature deaths over a 11 year-period were studied and concluded that about 40% were attributable to socioeconomic inequalities (Saint-Jacques, Dewar, Cui, Parker, &

* Corresponding author. Dpt of Geology, Geography and Environmental Sciences, University of Alcala, Calle Colegios 2, 28801, Alcala de Henares, Madrid, Spain. E-mail addresses: alba.cebrecos@uah.es (A. Cebrecos), felicitas.dominguez@salud.madrid.org (M.F. Domínguez-Berjón), ignacio.duque.rodriguez@ine.es (I. Duque), manuel.franco@uah.es (M. Franco), francisco.escobar@uah.es (F. Escobar).

https://doi.org/10.1016/j.apgeog.2018.04.001
Received 22 June 2017; Received in revised form 30 March 2018; Accepted 2 April 2018
0143-6228/

Dummer, 2014). In Europe, researchers studied 26,229,104 European inhabitants from 16 cities and found that up to 30% of excess deaths were attributable to socioeconomic disparities (Borrell et al., 2014). In Spain, mortality excess related to deprivation was 59,445 deaths among men and 23,292 among women (Borrell, Marí-Dell'Olmo, Serral, Martínez-Beneito, & Gotsens, 2010).

A given area may be socioeconomically deprived as a result of multiple interrelated factors. Deprived indices were first developed in the U.K (Carstairs & Morris, 1989; Townsend, 1987) as a multivariate tool which allows to study the level of deprivation in an area, com-paring deprivation effects across a variety of geographic regions and as a proxy of individual-level socioeconomic status (Smith, Hart, Watt, Hole, & Hawthorne, 1998). These indices provide a synthetic measure of different aspects of deprivation.

In Spain, the project "Socio-economic and environmental inequalities in mortality in small areas of Spanish cities", or MEDEA project (http://www.proyectomedea.org/), describes the geographical patterns of mortality and their relationship with the socioeconomic and environmental characteristics. To assess the socioeconomic status of areas in Spain, the MEDEA team developed an deprivation index based on 2001 national census (MF. Domínguez-Berjón et al., 2008). Results of this work have been used in a variety of research areas such as the study of the association between air pollution, Socio-Economic Status (SES) and health (Barceló, Saez, & Saurina, 2009; Cambra et al., 2013; García-Pérez et al., 2009; Ramis et al., 2009); health inequalities and mortality (M. F. Domínguez-Berjón et al., 2010; Gandarillas et al., 2011; Gotsens et al., 2011; Nolasco et al., 2009; Segura del Pozo et al., 2010) and deprivation and cancer incidence (Saurina et al., 2010) among others. In addition to their scientific impact, these results have been used as complementary information in the prioritization of interventions with an equity approach, published in the State Health Reports of Madrid Region (Primaria, 2007–2014). These reports are the base-document for public health planning in Madrid.

Given the dramatic social and economic changes occurred in Spain since late 2007, a redeveloped deprivation index with 2011 census was necessary. The general criticism of the deprivation indices focuses on the selection of the indicators that build the index, however less attention has been given to its geographical variability, a problem known as the Modifiable Areal Unit Problem (MAUP) (Schuurman, Bell, Dunn, & Oliver, 2007).

MAUP is an inherent problem of overlapping artificial spatial units (e.g. administrative areas) over a continuous geographical phenomenon (Openshaw, 1984). That implies a potential measurement error due to the aggregation of statistical data into these artificial units. Boundaries of these areas are defined by historical, political or operational reasons. There are not designed to define homogeneous zones with respect to social, economic or health characteristics of population. Social, economic or environmental phenomenon which promote or restrict health risks, are not limited by the boundaries of these artificial units.

The MAUP is composed by two interrelated effects (Openshaw, 1984). First, the scale effect which is the variation in results that can be obtained when data for one set of areal units are progressively aggregated into fewer and larger units of analysis (Openshaw, 1984). As an example, when census data are aggregated in neighborhoods, districts, and municipalities, results may change with increasing scale. Second, the zoning effect which refers to the different configurations of zones of the same size that may generate different results (Houston, 2014). Another example, when results using a 100-m continuous grid system differed from results using a 100-m grid system oriented in different ways (Houston, 2014).

In the international public health research field, there have been pioneering works that studied the implications of this problem in health research. The Geocoding project, studied which level of geography would be most apt for monitoring US socioeconomic inequalities in health, overall and within diverse racial/ethnic-gender groups (Krieger et al., 2003a; Krieger, Waterman, Chen, Soobader, & Subramanian,

Another Canadian study highlighted the effect of scale on indices by mapping ABDIs at multiple census scales in an urban area and compared self-rated health data with ABDIs at two census scales (Schuurman et al., 2007).

Selecting the spatial scale involves setting the spatial resolution of the study, this means, the ability to distinguish objects on the earth surface. On a larger scale, statistical aggregation of data on smaller surface units (e.g. census sections) will be required, but these areas contain fewer cases and therefore less stable rates. By contrast, at smaller scales, aggregation occurs in larger areas (e.g. districts), blur-ring significant variability and can sometimes lead to interpretations that are contradictory to those derived from finer-resolution data (Nelson & Brewer, 2015). Thus, it is necessary to consider that MAUP affects most statistics and has an impact on variance, standard deviation, correlation, regression analysis and any other statistical result (Flowerdew, Manley, & Sabel, 2008).

Any public health phenomenon under analysis is inseparably related to the scale because the scale provides its meaning. However, Geographic Information Systems (GIS) have been underused in health studies and wherever used, limited analyses selecting the relevant scale area have been conducted. The present study assesses the statistic and geographic (in)stability that arises when using deprivation data aggregated at different scales when understanding social determinants of health with a geographical perspective.

The aims of the study are twofold: first, to build an Area Based Deprivation Index (ABDI) at three different spatial scales (census section, neighborhoods and districts) for the city of Madrid. Second, to study the statistical and geographical stability of the ABDI throughout the three scales and its relationship with cardiovascular disease (CVD) prevalence.

2. Methods

2.1. Study setting and sample

This study is framed within a larger project, the Hearth Healthy Hoods (https://hhhproject.eu/). The HHH general objective is to understand the urban physical and social environment of the city of Madrid (Spain) and its relation with residents cardiovascular health (Bilal et al., 2016; Cebrecos et al., 2016; Julia; Díez et al., 2016; J Díez et al., 2017; Gullón et al., 2015). The HHH project includes the socio-demographic characterization of Madrid administrative spatial units. In this task, HHH researchers collaborate closely with the MEDEA project, with extensive experience in study of social deprivation and its association with health in Spain (Barceló et al., 2009; Cambra et al., 2013; MF.; Domínguez-Berjón et al., 2008; Gandarillas et al., 2011; García-Pérez et al., 2009; Nolasco et al., 2009; Ramis et al., 2009; Saurina et al., 2010; Segura del Pozo et al., 2010).

Madrid municipality, with an extension of 605 sqkm, consists of 21 districts, which are in turn divided into 128 neighborhoods and in 2409 basic spatial units called census sections (based on the 2011 Census data of National Statistical Institute -INE-). Madrid municipality had a total population of 3,186,595 in 2016 (INE, 2016). The administrative configuration of Madrid began at XII and XIII centuries and increasing its size as the capital of the Spanish Kingdom during the second half of XVIII century with different administrative divisions ("cuarteles", "barrios" and others). It was in the second half of XIX century, due to important migration flows from rural to urban areas that appeared for the first time the denomination of "district". Population increases and restructuring of administrative division continued with subsequent urban plans and developments until 1988 with the final and current number of district and neighborhoods (Prado Martínez, 2004). Ac-cording to the 2011 census, neighborhood average population is 24,895 (max. 69,300 and min. 770 inhabitants) and average surface is 4.7 sqkm (max. 187.5 and min. 0.25 sqkm). Districts have an average population of 151,743 inhabitants (max. 246,020 and min. 45,625 inhabitants) and an average surface of 28.8 sqkm (max. 237.9 and min. 4.7 sqkm).

Census sections are the smallest administrative unit. These are established and delimited by operational criteria for fieldwork in statistical operations and is essentially defined by criteria of population size with a mean population of approximately 1500 censed residents. Boundaries are easily recognizable by natural terrain features, permanent buildings or roads. The entire surface of Spain is divided in these spatial units, and in the case of Madrid Municipality, these sections are fully embedded within the neighborhood units which in turn are nested within district units. In Madrid municipality the average census section has 1323 inhabitants (max. 4510 and min. 100 inhabitants) and a surface of 0.2 sqkm (max. 94.5 and min. 0.008 sqkm).

2.2. Data sources

The information source to compute the ABDI was the 2011 census (National Statistical Institute -INE-). Indicators considered for the ABDI calculation were conditioned by the variables available in the 2011 census and also by a criterion of homogeneity with the indicators considered in the index calculated in 2001 by the MEDEA project (MF. Domínguez-Berjón et al., 2008). These criteria facilitate the comparison of deprivation measures in different periods, although this is not the purpose of this study.

Cardiovascular health data are collected, managed and anonymized by the Primary Health Care System of Madrid using the Electronic Health Records as the main source (Bilal et al., 2016). This database includes information on cardiovascular health and risk factors (tobacco use, obesity, hypertension, diabetes mellitus, dyslipidemia) and socio-demographic variables (age, sex) of 1,446,994 Madrid residents (ages of 45–70) for the year 2014. Appropriate anonymization process was conducted by removing all personally identifiable information (address, name, identifiers).

Cartographic information about the study area was collected from the Spanish National Mapping Agency (http://www.ign.es) and Spanish National Spatial Data Infrastructure (http://www.idee.es/), allowing to generate a georeferenced database to integrate and map the results. Administrative boundaries (census section, neighborhoods and districts) were obtained in vector polygon format. ArcGIS 10.1 software was used to integrate, standardize, manage and visualize these datasets. All spatial information was projected to a common system (ETRS89 UTM 30N). Contextual layers were introduced to the final maps as reference information.

2.3. Indicators contemplated for possible inclusion in the index

As a result of the bibliographic review and the data available from the 2011 census, twelve indicators were initially selected (Table 1). Indicators were separated into two large groups: socioeconomic and demographic, differentiating the former into 3 types, depending on whether they were related to occupation, labor market and education level. Percentages of the 12 indicators were calculated in the three spatial scales considered.

Occupation types expose workers to different occupational hazards leading to different psychosocial individual processes as stress, control and autonomy (Galobardes, Shaw, Lawlor, Lynch, & Davey Smith, 2006). In addition, occupation type is usually correlated to income and therefore is one of the measures commonly used to obtain social class. The fact of belonging to a disadvantaged social class places families in circumstances of limited access to material and social resources. The chosen indicator (manual workers) is based on the British classification of social class that classified the occupations according to the level of education or learning and that was adapted for Spain (Domingo Salvany Marcos Alonso, 1989; Sociedad Española de Epidemiologia (SEE) and SemFYC (2000). The categories can be reduced to 2 large groups: manual and non-manual occupations.

Unemployment reflects the lack of income and material resources and is a marker of both material and social deprivation (Campbell, Radford, & Burton, 1991). In addition, there are now increasingly frequent forms of underemployment or precarious employment that share common characteristics with unemployment.

Education level is an indicator of material deprivation and is a strong determinant of employment (or at least the first job) and income (MF Domínguez-Berjón, Borrell, Benach, & Pasarín, 2001; Galobardes et al., 2006). It is also an indicator of social deprivation, because through education skills are acquired to meet social demands or resolve potentially stressful situations.

The initial selection of demographic indicators aimed to identify the population groups with a disadvantaged socio-economic situation. This selection includes indirect indicators as none of them implies deprivation by itself in a single fashion (Thunhurst, 1985). Initially selected indicators refered to aging, people foreign born in low-income countries and to single-parent households.

2.4. Analysis

2.4.1. Area Based Deprivation Index (ABDI)

Based on the MEDEA project method to calculate the 2001 ABDI (MF. Domínguez-Berjón et al., 2008) a new index was calculated using the 2011 census data. IBM SPSS software was adopted for the statistical analysis. Previously to the analysis, z-scores for each of the indicators were calculated.

As a first step, a correlation analysis was performed to check for covariance between variables. We found covariance between the indicators of manual workers, temporary employees and foreigners born. Three indicators were eliminated to overcome the problem of covariance (Manual workers ≥ 16 years old among the employed, temporary employees ≥ 16 years old among the occupied; and born in low income countries). Finally, nine indicators were considered for the calculation of the new ABDI. ABDI was calculated by a Principal Component Analysis (PCA) with extraction of eigenvalues greater than or equal to one. That assumes the extraction of different components that are capable of explaining an amount of total variance greater than or equal to 1. The interpretation of the factors was performed on the varimax rotation. From this first analysis were selected the indicators with factors loading higher than 0.6 in the first component. Finally, the ABDI was constructed by aggregation of these indicators with the extraction of a single axis of PCA.

2.4.2. Statistic and geographic stability study

In order to study statistic and geographic stability of ABDI across three scales, we adopted an Exploratory Spatial Data Analysis (ESDA) approach. ESDA extends the definition of Exploratory Data Analysis (EDA) to spatial data increasing the set of visual tools, as choropleth maps; and the set of numerical as spatial cluster detection statistics (Fischer & Nijkamp, 2014). Visual analytics is an emerging inter-disciplinary and sensemaking approach to exploring, analyzing, communicating, and generating hypotheses from large, complex, and heterogeneous data (Nelson & Brewer, 2015). This is one of ESDA advantages, allows to generate new hypotheses from a large volume of data instead of validating them a priori.

A classic non-spatial analytical approach assumes that variables are independent of the region where they occur. This assumption directly conflicts with the first law of the Geography “everything is related to everything else, but near things are more related than distant things” (Tobler, 1970). The law expresses spatial dependency process or spatial autocorrelation. Spatial autocorrelation not only affects statistical analysis of social phenomena, it also takes into account the influence that observation units have among themselves.

Thus, election of ESDA techniques is the best option to evaluate the MAUP allowing for visualization and estimation of spatial auto-correlation. Spatial statistics can take a global or local form. Spatial statistics abstracts complex

Table 1
Operational definitions for area-based socioeconomic indicators of the 2011 census included in the construction of the area based deprivation index.

	Indicator	Operational definition
Socioeconomic group	Manual workers ≥16 years old among the employed	Percentage of people aged 16 years or over employed in sectors services, agriculture, fishing, craftwork, skilled workers in manufacturing industries, construction, mining, installations operators, and non-skilled workers; with respect to the total employed population aged 16 years or over
	Manual workers ≥16 years old among the employed or unemployed who have worked before	Percentage of people aged 16 years or over, employed or unemployed who have worked before as manual workers (employed in sectors: services, agriculture, fishing, craftwork, skilled workers in manufacturing industries, construction, mining, installations operators, and non-skilled workers) with respect to the total employed population aged 16 years or over
	Unemployment ≥16 years old	Percentage of people aged 16 years or over without a job (unemployed and those seeking work for the first time), with respect to the total economically active population
	Temporary employees ≥ 16 years old among the occupied	Percentage of people aged 16 or older occupied as eventual employees.
	Temporary employees ≥16 years old among the occupied or unemployed who have worked before	Percentage of people aged 16 or older occupied or unemployed who have worked before as eventual employees.
	Insufficient instruction ≥16 years old	Percentage of people aged 16 years and over who, according to a list of the National Statistics Institute, cannot read or write; can read and write but have less than 5 years schooling; went to school for 5 years or more but did not complete basic compulsory education, with respect to the total population aged 16 years and over.
	Insufficient instruction 16–29 years old	Percentage of people aged between 16 and 29 years with low educational level, with respect to the total population aged from 16 to 29 years.
Demographic group	Aging	Percentage of population aged 65 or older respect to the total population.
	Born in low income countries	Percentage of population born in countries other than the following: Spain, Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Iceland, Italy, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, Andorra, Germany, San Marino, Vatican City, Sweden, Switzerland, Canada, United States of America, Japan, Australia and New Zealand; respect to the total population
	Born in low income countries arriving in Spain after 2006	Percentage of population born in countries other than the following: Spain, Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Iceland, Italy, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, Andorra, Germany, San Marino, Vatican City, Sweden, Switzerland, Canada, United States of America, Japan, Australia and New Zealand, and who arrived in Spain between 2007 and 2011; respect to the total population.
	Born in low income countries or born in Spain whose father or mother was born in low income countries	Percentage of population born or with their father or mother born in countries other than listed above, respect to the total population.
	Population in single-parent households	Population in households with a single-father who lives at least, with a child ≤25 years.

and multivariate spatial, temporal, and scalar relationships into a single numeric statistic, such as Moran's Index of spatial autocorrelation (Moran, 1950). Local spatial statistics allow the decomposition of global indicators into the contribution of each individual observation (Anselin, 1995). They produce a surface of geographically weighted values, which capture local variability, but often require a cartographic approach to make sense of the result. In this work, we conducted a bivariate Local Indicators of Spatial Association (or bivariate LISA). Spatial analyses were run using Geoda Software (GeoDa Center for Geospatial Analysis, Chicago, Illinois).

Before conducting bivariate analysis, it was necessary to process data through a spatial union. This technique allows to join the attributes of one entity with the attributes of another based on the spatial relationship existing between them. In these analyses, each census section received index values of the neighborhood and district in which census section is spatially nested. The administrative units selected foster a more direct MAUP assessment due to clean nesting relationship between the three aggregate levels. Once the data processing was finished, bivariate LISA analyses were conducted. The analyses allow to locate areas or clusters where the indexes were stable or unstable throughout the three scales. Unlike the classical LISA analysis individually identifying statistically significant correlation clusters for one variable; bivariate analyses study the correlation between two different variables (Lee, 2017).

The type of cluster obtained by the bivariate LISA analyses might be:

Positive autocorrelation: Cluster High-High (HH) or Low-Low (LL). A HH-cluster represents census sections with an above-average value of ABDI, surrounded significantly by sections with values of ABDI calculated at neighborhood level or at district level; above their average. The LL-cluster represents the opposite case. These clusters indicate stability throughout the three scales.

Negative autocorrelation: Cluster High-Low (HL) or Low-High (LH). A HL-cluster represents census sections with an above-average value of ABDI, surrounded significantly by sections with values of ABDI calculated at neighborhood level or at district level; below their average. The LL-cluster represents the opposite case. These clusters indicate in-stability throughout the three scales.

This analysis allowed to know how ABDI values are correlated on a given scale with the ABDI values of a different scale (census versus neighborhood section, census versus district and neighborhood vs. district section). To study the stability of ABDI across the 3 scales the definition of contiguous neighborhood of first order queen was used as spatial weight matrix.

Finally, to corroborate the pattern of socioeconomic deprivation may be different at different scales, we studied its correlation with the prevalence of CVD in the whole municipality of Madrid. To do this, we calculated age-adjusted rates of CVD prevalence for the three aggregation scales (census section, neighborhood and district).

3. Results

As a result of our first aim, we built an Area Based Deprivation Index (ABDI) at three different spatial scales (census section, neighborhoods and districts) for the city of Madrid. The PCA analysis for the three scales, we found different factors loading greater than 0.6 in the first component for each scale. At census section and neighborhood level, first component showed four and five factors loading higher than 0.6, respectively. These factors loading were related to occupancy/labor market and education. However, at district level, first component showed seven factors loading higher than 0.6 and were related with occupancy/labor market, education and immigration. Table 2 shows the factors loading that finally was combined to calculate the ABDI

Table 2
Factors loading obtained by principal component extraction from a single axis for selected indicators at each scale and variability explained.

	Census section	Neighborhood	District
Manual workers	0,882	0957	0,969
Unemployment	0,785	0948	0,958
Eventual employees	0,741	0908	0,952
Insufficient instruction	0,814	0963	0,924
Young insufficient instruction	–	0,853	0906
Recent foreigners born in low income country	–	–	0,657
Foreigners whose father or mother born in low income country	–	–	0,832
Variability explained (%)	65.2	85.9	79.4

extracting a single PCA axis. Variance explained for each index was 65%, 86% and 79% respectively.

On the left, Fig. 1 shows the distribution of the ABDI calculated on the three scales. On the right, the distribution of age-adjusted rates of CVD prevalence is shown. In both cases, the fifth quintile represents a more unfavorable socioeconomic situation and a highest adjusted rates of prevalence of cardiovascular disease. For both variables there is a greater variability at the census section level, while increasing the scale, there is a smoothing of the results. Therefore, a possible north-south pattern is more clearly seen in both cases (more clearly in the pre-valence of disease) at the neighborhood and district level.

The average values of indicators that compose index and age-adjusted rates of cardiovascular prevalence in each of the quintiles of the index are presented in Table 3. Fifth quintile (Q5) represents the most disadvantaged areas. For all scales, the percentages of CVD rates adjusted for age are greater as deprivation increases.

To unravel what relationships there are between the deprivation indices across the scales, how stable is the relationship and how the geographical distribution is, Fig. 2 was developed. The figure is the cartographic representation of bivariate analysis result. GeoDA soft-ware provides results of spatial autocorrelation and significance in in-dependent maps for each one of them, which limits the interpretation of spatial autocorrelation and significance in the context of space simultaneously. Thus, this new cartography was developed. The choropleth bivariate mapping showed in Fig. 2 integrates spatial auto-correlation and level of significance across-scales of analysis to facilitate visual analysis of ABDI stability.

Fig. 2a shows the results of comparing the ABDI calculated at the census section level against the ABDI at the neighborhood level. Fig. 2b presents the results of comparing the ABDI calculated at the census section level against the ABDI at the district level. Fig. 2c shows the results of comparing the ABDI calculated at the neighborhood level against the ABDI at the district level.

Color symbology shows grade of stability or instability obtained by the Local Moran Index analysis. Color pink represents instability (negative spatial autocorrelation). Let's take as an example Fig. 2b; the pink areas in the southern of the study area showed significant in-stability between the ABDI calculated at the census tract level and the ABDI calculated at the district level. These were clusters of type HL or LH, which represent areas in which the ABDI at the level of the census section presented high values surrounded by low values of ABDI calculated at the neighborhood level, and vice versa. On the other hand, the blue color represents the stability of the indexes through the scales (or the positive spatial autocorrelation). Following the example in Fig. 2b, the predominant blue colors in the central area of the municipality represent stability through the census and district section escalations. That is, they were HH or LL clusters, where high values of the ABDI at the census section level were surrounded by high ABDI values at the neighborhood level. Gray colors represent values of Local Moran Index close to zero, which implies low clustering or randomization.

Color tones inform about the signification level, dark colors imply high signification while light colors imply low signification.

Fig. 2c presents the largest number of unstable areas (a total of 270 census sections), but is the analysis of Fig. 2b which presents the largest number of unstable areas with the greatest significance (90 pink census sections with a level of significance less than 0.05 in Fig. 2c against 144 census sections in Fig. 2b). The greatest stability (blue colors) encompasses 1850 census section in the relationship between neighbor-hood and its next scale of aggregation (Fig. 2c) of which 1084 have a significance less than or equal to 0.01. Fig. 2a shows the higher random results found; 1126 census section with little or none autocorrelation, or with a random spatial distribution.

Secondly, we aimed to study the statistical and geographical stability of the ABDI throughout the three scales and its relationship with cardiovascular disease (CVD) prevalence.

The correlation analysis between the ABDI calculated at different scales (and the indicators that compose it) and the age-adjusted rates of the prevalence of CVD are presented in Table 4. The ABDI presented a significant correlation for all scales of CVD being greater as we increase the administrative unit. The indicators that make up the index in each scale were also significantly correlated, with the exception of immigration indicators at the district level. At the census section level, the indicator with the highest correlation was manual workers ($r = 0.341$; < 0.01) as well as at the neighborhood level ($r = 0.677$, $p < 0.01$). At the district level, the indicator that presented the highest correlation with the prevalence of CVD was the indicator young insufficient instruction ($r = 0.865$, $p < 0.01$).

4. Discussion

The general aim of this study was to highlight the implications of unit of analysis and scale choice when using spatially deprivation aggregated data in urban health research. We found that for different administrative spatial units, deprivation index yielded different results and different correlations with the prevalence of CVD. These results add to the already published evidence warnings about the need to examine the effects of varying the geographic extent when calculating measures and associations between urban contextual environment and health outcomes (Hanigan, Cochrane, & Davey, 2017; Houston, 2014; Rey, Jougla, Fouillet, & Hémon, 2009). These findings have clear implications for the effective targeting of resources and public health interventions.

Our results showed that by varying the geographical unit in the ABDI calculation, the indicators that make up the ABDI also varied. At the census section and at neighborhood level indicators of the ABDI were related to occupancy/labor market and education domains. But at district level, two indicators related with immigration obtained factors loading greater than 0.6 in the first component. This could indicate that immigration is a social phenomenon not limited locally in a city as densely populated as Madrid; but it extends across a wider territory. Therefore, it becomes a statistically significant component for socio-economic deprivation when studying in a larger observation unit such as district.

These results add to the evidence that social phenomena have multiscale characteristics. Root (2012) found that poverty showed a stronger effect at smaller scales and unemployment at larger scales in relation with cleft palate (Root, 2012). It makes some sense that un-employment rate might influence at a higher level of geography be-cause high unemployment tends to depress the economy in much larger areas (e.g., a whole city), whereas poverty is often localized, existing in “pockets” around a city (Root, 2012). Johnston et al. (2016) concluded that segregation in Australia was greater in macro- and micro-scales than in intermediate scales and it varies in its intensity (Johnston, Forrest, Jones, & Manley, 2016). At different spatial scales and/or aggregation level yields different patterns and degrees of spatial auto-correlation that will impact on the results of any statistical analysis.

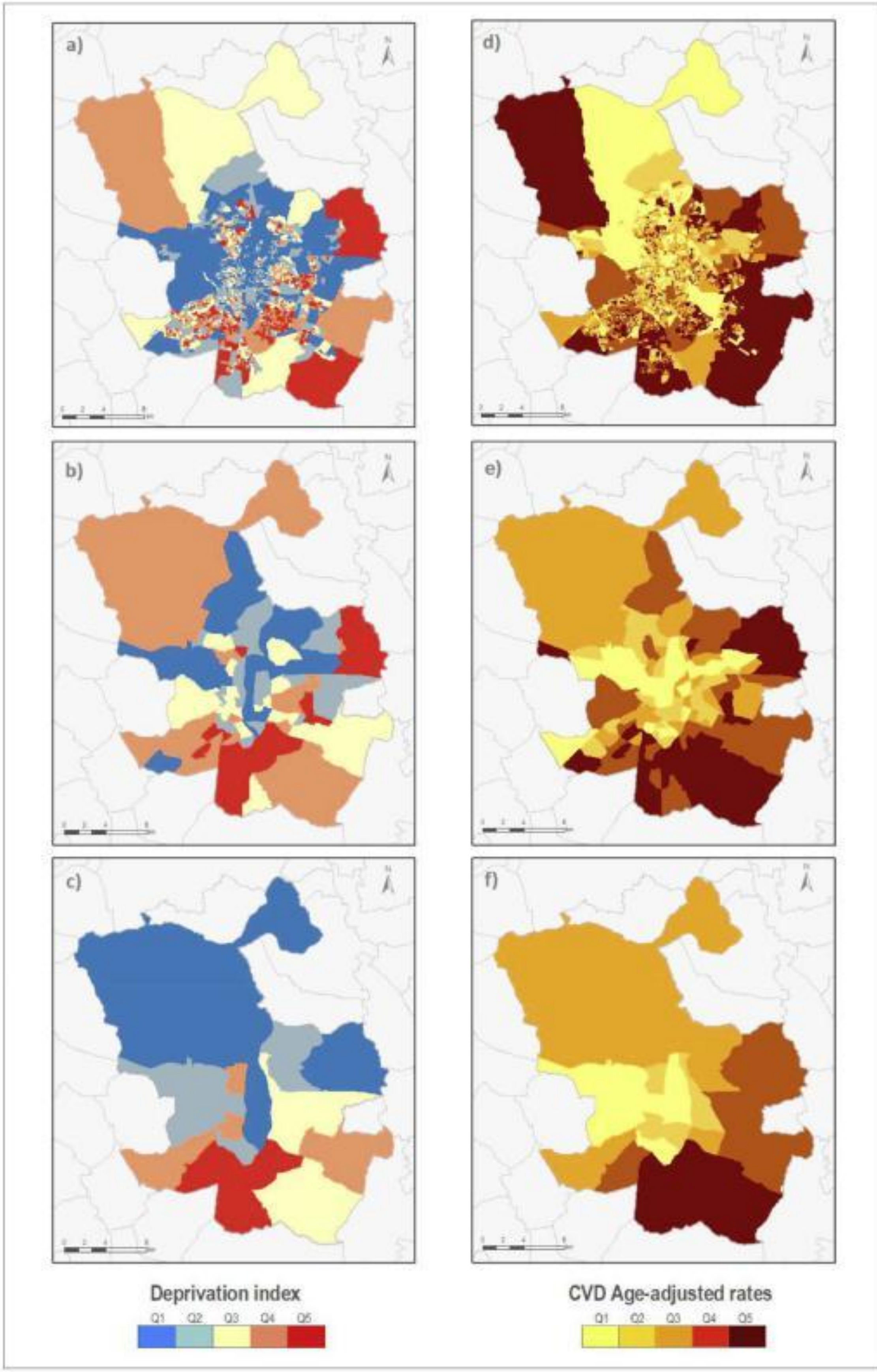


Fig. 1. Area-Based Deprivation Index (ABDI), on the left panels, by quintiles in Madrid Municipality at a) the census section scale, b) the neighborhood scale and c) the district scale. Age-adjusted rates of Cardiovascular disease (CVD) prevalence, on right side, at f) the census section scale, g) the neighborhood scale and g) the district scale.

In the statistical study and geographical stability of this work, it is observed how the ABDI did not remain stable throughout the three administrative units. These results were corroborated with the correlation analysis between socioeconomic deprivation and the prevalence of CVD. All the correlations were positive and statistically significant ($p < 0.01$) across the three scales. This association being strongest when the unit of geographical analysis increased: $r = 0.328$ at census section level, $r = 0.635$ at neighborhood level and $r = 0.739$ at district level. In general, any more disadvantaged geographical unit had a greater correlation with the prevalence of CVD.

It can be seen from Fig. 1 the variation captured at census section level in higher that at neighborhood or district level. Some of this variance increases undoubtedly due to "noise" introduced by random events that are more influential on rate calculations where the under-lying population denominators are small (Hanigan et al., 2017). An-other important factor, from the perspective of statistical modeling, is the loss of statistical power associated with a smaller number of spatial units for inclusion in the analysis at district level ($N = 21$). Thus, de-spite the presence of added variance at census section scale, we con-sider it is a "truer" representation of the exposure measures with a larger number of sampling units providing a better overall representation of the available data.

While the Modifiable Areal Unit Problem (MAUP) is an old friend of geosciences it has been also identified as a methodological challenge in social epidemiology. Diez Roux, already posed in 2001 that the size and definition of the relevant geographic area may vary according to the processes through which the area effect is hypothesized to operate and the outcome being studied (Diez Roux, 2001). In the current study we showed how it is possible to use a bivariate autocorrelation analysis to explore how is correlated a socioeconomic deprivation index with the deprivation values closer at a different scale. Along with visualization, this approach has allowed us to characterize the relationships between MAUP and the spatial scale.

Obviously, our aim was not to solve the MAUP, but rather to emphasize the need to analyze associated and underlying phenomena of data aggregated in area units. For that, we developed a straightforward and easily replicable ABDI in any given country and study its stability across scales. The phenomenon of socioeconomic deprivation was chosen because it is a universal phenomenon, a challenge for public health policies and usually is evaluated through based at spatially aggregated data. In addition, we assessed the changes. In addition, we evaluated the changes suffered by socioeconomic deprivation when it is calculating at different spatial units; as well as the changes that occur in their relationship with CVD. The rates adjusted for age of CVD pre-valence were calculated from data from the Primary Health Care System of 1,446,994 residents of Madrid. This study is unique in Spain and provides new relevant knowledge when taking into account variations in health outcomes using aggregate data at the area level.

However, our study is not without limitations. The data source for calculating ABDI is the national census which is collected every 10 years. This implies a large time lag that makes it difficult to update the data for longitudinal or cross-sectional studies throughout the decade. Besides, the information may not be comprehensive enough to reflect all aspects of interest, such as selected social deprivation indicators. Another limitation is the correlation analysis carried out between socioeconomic deprivation and the prevalence of CVD does not allow to infer the causality between the variables. However, this correlational investigation allows to corroborate the changes of relationship between socioeconomic deprivation and the prevalence of CVD due to the choice of the scale of analysis.

This research results may inform policy decisions regarding health surveillance and public health interventions by taking into account the impact of the aggregation scale on the association between socio-economic disadvantage and health outcomes. Based on these results, our recommendation is that the descriptive analysis and notification should be carried out at the neighborhood or district level, since these spatial units are the ones best capturing the wide distribution of disease. However, for explanatory or inferential studies, it may be necessary to drill down to census section level to delineate those areas for specific focus, both for understanding the likely exposures or for targeting re-sources for intervention. In the case of Madrid, within the HHH study, we closely collaborate with the regional and local primary health care and public health administrations assuring these knowledge may be translated into local policy decisions.

5. Conclusion

The choice of the spatial scale of analysis is key for epidemiological studies to understand the associations between the urban context and health. From the results of this study it can be concluded that the scale of analysis influences the understanding of the geographical patterns of socio-economic disadvantage, specifically in relation cardiovascular

Table 3

Average Area-Based Deprivation Index indicators values and age-adjusted rates of cardiovascular prevalence per quintiles. Q5 represents the most disadvantaged areas.

		% Manual Workers	% Unemployment	% Eventual employees	% Insufficient instruction	% Young insufficient instruction	% Recent foreigners	% Foreigners parents	% CVD rates
Census section	Q1	19.4	13.4	12.4	8.0	–	–	–	3.4
	Q2	29.5	17.5	18.0	13.7	–	–	–	3.6
	Q3	40.8	20.5	21.6	19.3	–	–	–	3.1
	Q4	53.6	24.6	24.8	26.1	–	–	–	4.0
	Q5	67.8	32.4	32.6	33.2	–	–	–	4.6
Neighborhood	Q1	20.1	14.8	12.0	7.3	4.3	–	–	3.5
	Q2	27.1	17.0	16.8	11.6	5.4	–	–	3.6
	Q3	37.3	20.4	21.3	16.8	6.7	–	–	3.8
	Q4	48.6	23.5	23.6	23.5	11.0	–	–	4.0
	Q5	60.8	29.0	28.1	30.6	16.2	–	–	4.7
District	Q1	26.6	16.6	16.0	11.9	5.6	3.1	13.1	3.7
	Q2	29.8	18.3	18.6	13.9	5.7	3.2	14.9	3.5
	Q3	42.6	22.3	20.8	20.5	9.7	2.9	16.4	4.1
	Q4	43.3	22.4	22.7	19.4	9.6	4.6	22.6	3.9
	Q5	58.4	27.2	26.9	27.5	14.6	4.8	24.0	4.6

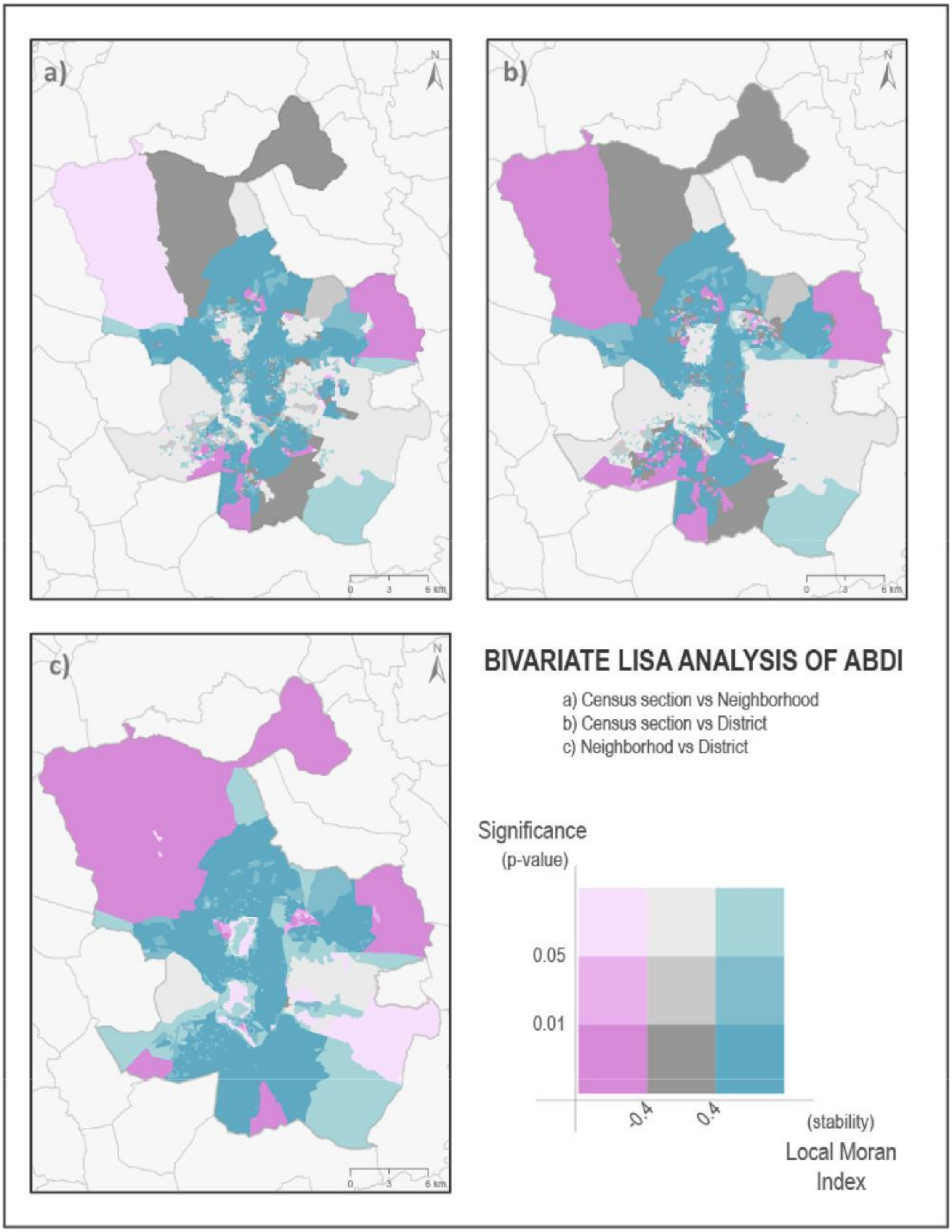


Fig. 2. Bivariate analysis of Local Indicator of Spatial Association (bivariate LISA) of the relationship of the Area Based Deprivation Indices (ABDIs) calculated throughout the three scales. The legend depicts the local indicator value against significance (p-value) for relationship between a) census section and neighborhood, b) census section and district, and c) neighborhood and district.

health. The correlations between deprivation and CVD at the three study scales were positive and significant, being greater as the size of the spatial unit increased. Deprivation and adjusted rates of CVD pre-valence showed greater variability at the census section level than when they were added at the district level. It is because the aggregation to a larger geographical unit softens the variation located within the district. It is warranted to take into account the implications of these results when developing interventions or allocating funds for the fight against social inequalities in health.

Deprivation indices are a valuable tool for assessing socio-demographic and health inequalities in our cities. A strong theoretical framework is warranted to help researchers choosing the geographical scale best representing the pathways acting between specific social determinants and health outcomes. If there is uncertainty about the best scale, several geographic scales and explicit hypotheses shall be explored.

Table 4
Results of the Pearson correlation between the deprivation index calculated for the three administrative units (and the indicators that build them) with the age-adjusted rates of cardiovascular disease prevalence.

	Census section (N = 2409)	Neighborhood (N = 128)	District (N = 21)
% Manual Workers	0.341**	0.677**	0.837**
% Unemployment	0.215**	0.586**	0.788**
% Eventual employees	0.186**	0.493**	0.599**
% Insufficient instruction	0.305**	0.604**	0.783**
% Young insufficient instruction	0.190**	0.582**	0.865**
% Recent foreigners	0.042*	0.101	0.209
% Foreigners parents	0.085**	0.205*	0.398
ABDI	0.328**	0.635**	0.739**

*p value < 0.05.
**p value < 0.01.

Acknowledgements

We would like to thank the whole Social and Cardiovascular Research Group for their collaboration and ideas on the study. Thanks also to MEDEA project for yielding its long experience to this work and to the Spanish Society of Epidemiology for funding to provide the 2011 census data.

List of abbreviations

ABDI	Area Based Deprivation Index
CVD	Cardiovascular Disease
EDA	Exploratory Data Analysis
ESDA	Exploratory Spatial Data Analysis
GIS	Geographic Information System
PCA	Principal Component Analysis
SES	Socio-Economic Status

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Authors' contributions

AC conducted the geospatial analyses and drafted the manuscript. FD help with the conceptual framework to calculate the index, and revised the manuscript. ID developed the strategy of validation and robustness of the data and revised the manuscript. OK reviewed all spatial analyzes and the manuscript. MF conceived the Heart Healthy Hoods project and revised the manuscript. FE oversaw spatial analytical results and revisions the manuscript. All authors read and approved the final manuscript.

Conflicts of interest

The authors of this paper do not report any conflict of interests.

Availability of data and materials

Not applicable.

Funding

This study was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/ 2007–2013)/ERC Starting Grant HeartHealthyHoods Agreement n. 336893.

References

Anselin, L. (1995). Local indicators of spatial association—LISA. Geographical Analysis, 27(2), 93–115.
Barceló, M. A., Saez, M., & Saurina, C. (2009). Spatial variability in mortality inequalities, socioeconomic deprivation, and air pollution in small areas of the Barcelona Metropolitan Region, Spain. The Science of the Total Environment, 407(21), 5501–5523.
Bilal, U., Díez, J., Alfayate, S., Gullón, P., del Cura, I., Escobar, F., et al. (2016). Population cardiovascular health and urban environments: The Heart Healthy Hoods exploratory study in Madrid, Spain. BMC Medical Research Methodology, 16(1), 104.
Borrell, C., Marí-Dell'olmo, M., Palència, L., Gotsens, M., Burström, B., Domínguez-Berjón, F., et al. (2014). Socioeconomic inequalities in mortality in 16 European ci-ties. Scandinavian Journal of Public Health, 42(3), 245–254.
Borrell, C., Marí-Dell'Olmo, M., Serral, G., Martínez-Beneito, M., & Gotsens, M. (2010). Inequalities in mortality in small areas of eleven Spanish cities (the multicenter MEDEA project). Health & Place, 16(4), 703–711.
Cambra, K., Martínez-Rueda, T., Alonso-Fustel, E., Cirarda, F. B., Audicana, C., Esnaola, S., & Ibáñez, B. (2013). Association of proximity to polluting industries, deprivation and mortality in small areas of the Basque Country (Spain). The European Journal of Public Health, 23(1), 171–176.
Campbell, D. A., Radford, J. M., & Burton, P. (1991). Unemployment rates: An alternative to the jarman index? BMJ, 303(6805), 750–755.
Carstairs, V., & Morris, R. (1989). Deprivation, mortality and resource allocation. Journal of Public Health, 11(4), 364–372.
Cebrecos, A., Díez, J., Gullón, P., Bilal, U., Franco, M., & Escobar, F. (2016). Characterizing physical activity and food urban environments: A GIS-based multi-component proposal. International Journal of Health Geographics, 15(1), 35.
Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. American Journal of Public Health, 91(11), 1783–1789.
Díez, J., Bilal, U., Cebrecos, A., Buczynski, A., Lawrence, R. S., Glass, T., et al. (2016). Understanding differences in the local food environment across countries: A case study in Madrid (Spain) and Baltimore (USA). Preventive Medicine, 89, 237–244.
Díez, J., Conde, P., Sandin, M., Urtasun, M., López, R., Carrero, J., et al. (2017). Understanding the local food environment: A participatory photovoice project in a low-income area in Madrid, Spain. Health & Place, 43, 95–103.
Domingo Salvany, A., & Marcos Alonso, J. (1989). Propuesta de un indicador de la «clase social» basado en la ocupación. Gaceta Sanitaria, 3(10), 320–326.
Domínguez-Berjón, M., Borrell, C., Benach, J., & Pasarín, M. (2001). Medidas de privación material en los estudios de áreas geográficas pequeñas. Gaceta Sanitaria, 23–33.
Domínguez-Berjón, M., Borrell, C., Cano-Serral, G., Esnaola, S., Nolasco, A., Pasarín, M., et al. (2008). Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). Gaceta Sanitaria, 22(3), 179–187.
Domínguez-Berjón, M. F., Gandarillas, A., Segura del Pozo, J., Zorrilla, B., Soto, M. J., López, L., et al. (2010). Census tract socioeconomic and physical environment and cardiovascular mortality in the Region of Madrid (Spain). Journal of Epidemiology & Community Health, 64(12), 1086–1093.
Fischer, M. M., & Nijkamp, P. (2014). Handbook of regional science. Berlin Heidelberg: Springer-Verlag.
Flowerdew, R., Manley, D. J., & Sabel, C. E. (2008). Neighbourhood effects on health: Does it matter where you draw the boundaries? Social Science & Medicine, 66(6), 1241–1255.
Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Davey Smith, G. (2006). Indicators of socioeconomic position (part 2). Journal of Epidemiology & Community Health, 60(2), 95–101.
Gandarillas, A., Domínguez-Berjón, M. F., Zorrilla, B., Galán, I., Duque, I., & Segura del Pozo, J. (2011). Deaths in collective dwellings and inequalities in small-area mortality: An ecological study in the Madrid region (Spain). Journal of Epidemiology & Community Health, 65(4), 310–314.
García-Pérez, J., Pollán, M., Boldo, E., Pérez-Gómez, B., Aragonés, N., Lope, V., et al. (2009). Mortality due to lung, laryngeal and bladder cancer in towns lying in the vicinity of combustion installations. The Science of the Total Environment, 407(8), 2593–2602.
Gotsens, M., Marí-Dell'Olmo, M., Martínez-Beneito, M.Á., Pérez, K., Pasarín, M. I., Daponte, A., et al. (2011). Socio-economic inequalities in mortality due to injuries in small areas of ten cities in Spain (MEDEA Project). Accident Analysis & Prevention, 43(5), 1802–1810.
Gullón, P., Badland, H. M., Alfayate, S., Bilal, U., Escobar, F., Cebrecos, A., Diez, J., & Franco, M. (2015). Assessing walking and cycling environments in the streets of Madrid: Comparing on-field and virtual audits. Journal of Urban Health, 92(5), 923–939.
Hanigan, I. C., Cochrane, T., & Davey, R. (2017). Impact of scale of aggregation on associations of cardiovascular hospitalization and socio-economic disadvantage. PLoS One, 12(11) e0188161.
Houston, D. (2014). Implications of the modifiable areal unit problem for assessing built environment correlates of moderate and vigorous physical activity. Applied

Geography, 50, 40–47.
INE (2016). In I. N. d. Estadística (Vol. Ed.), Censos de Población y Viviendas 2011: Vol. 2015Instituto Nacional de Estadística.
Johnston, R., Forrest, J., Jones, K., & Manley, D. (2016). The scale of segregation: Ancestral groups in sydney, 2011. Urban Geography, 37(7), 985–1008.
Krieger, N., Chen, J. T., Waterman, P. D., Soobader, M. J., Subramanian, S. V., & Carson, R. (2002). Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter?: the public health disparities geocoding project. American Journal of Epidemiology, 156(5), 471–482.
Krieger, N., Chen, J., Waterman, P., Soobader, M., Subramanian, S., & Carson, R. (2003a). Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). Journal of Epidemiology & Community Health, 57(3), 186–199.
Krieger, N., Waterman, P. D., Chen, J. T., Soobader, M. J., & Subramanian, S. V. (2003b). Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence: Geocoding and choice of area-based socioeconomic measures–the public health disparities geocoding project (US). Public Health Reports, 118(3), 240–260.
Lee, S. I. (2017). Correlation and spatial autocorrelation. In S. Shekhar, H. Xiong, & X. Zhou (Eds.). Encyclopedia of GIS (pp. 360–368). Cham: Springer International Publishing.
Marmot, M. (2005). Social determinants of health inequalities. Lancet, 365(9464), 1099–1104.
Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. Biometrika, 37(1/2), 17–23.
Nelson, J. K., & Brewer, C. A. (2015). Evaluating data stability in aggregation structures across spatial scales: Revisiting the modifiable areal unit problem. Cartography and Geographic Information Science, 1–16.
Nolasco, A., Melchor, I., Pina, J. A., Pereyra-Zamora, P., Moncho, J., Tamayo, N., et al. (2009). Preventable avoidable mortality: Evolution of socioeconomic inequalities in urban areas in Spain, 1996–2003. Health & Place, 15(3), 732–741.
Openshaw, S. (1984). The modifiable areal unit problem. Geo Abstracts University of East Anglia.
Prado Martínez, P. (2004). La Organización de los Distritos. Ayuntamiento de Madrid. Libro de las Jornadas sobre la Aplicación de la Ley de Grandes Ciudades.
Primaria, D. G. d. A. (2007-2014). Informe del Estado de Salud de la Población de la Comunidad de Madrid. Madrid: Consejería de Sanidad.
Ramis, R., Vidal, E., García-Pérez, J., Lope, V., Aragonés, N., Pérez-Gómez, B., et al. (2009). Study of non-Hodgkin's lymphoma mortality associated with industrial pollution in Spain, using Poisson models. BMC Public Health, 9(1), 26.
Rey, G., Jougla, E., Fouillet, A., & Hémon, D. (2009). Ecological association between a deprivation index and mortality in France over the period 1997 – 2001: Variations with spatial scale, degree of urbanicity, age, gender and cause of death. BMC Public Health, 9(1), 33.
Root, E. D. (2012). Moving neighborhoods and health research Forward: Using geo-graphic methods to examine the role of spatial scale in neighborhood effects on health. Annals of the Association of American Geographers, 102(5), 986–995.
Saint-Jacques, N., Dewar, R., Cui, Y., Parker, L., & Dummer, T. J. (2014). Premature mortality due to social and material deprivation in Nova Scotia, Canada. International Journal for Equity in Health, 13(1), 94.
Saurina, C., Saez, M., Marcos-Gragera, R., Barceló, M. A., Renart, G., & Martos, C. (2010). Effects of deprivation on the geographical variability of larynx cancer incidence in men, Girona (Spain) 1994–2004. Cancer Epidemiology, 34(2), 109–115.
Schuurman, N., Bell, N., Dunn, J., & Oliver, L. (2007). Deprivation indices, population health and Geography: An evaluation of the spatial effectiveness of indices at multiple scales. Journal of Urban Health, 84(4), 591–603.
Segura del Pozo, J., Gandarillas, A., Domínguez Berjón, F., Soto, M. J., López, L., Marta, I., et al. (2010). Chronic liver disease and cirrhosis mortality and social deprivation: A spatial analysis in small areas of Madrid region. Nutricion Hospitalaria, 25(4), 597–605.
Sir Michael, M. (2006). Health in an unequal world. Lancet, 368(9552), 2081–2094.
Smith, G. D., Hart, C., Watt, G., Hole, D., & Hawthorne, V. (1998). Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: The renfrew and paisley study. Journal of Epidemiology & Community Health, 52(6), 399–405.
Sociedad Española de Epidemiologia (SEE), & SemFYC, G. (2000). Una propuesta de medida de la clase social. Atención Primaria, 25(5), 350–363.
Thunhurst, C. (1985). The analysis of small area statistics and planning for health. Journal of the Royal Statistical Society. Series D (The Statistician), 34(1), 93–106.
Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. Economic Geography, 46(Suppl. 1), 234–240.
Townsend, P. (1987). Deprivation. Journal of Social Policy, 16(2), 125–146

Study 3. A multicomponent method assessing healthy cardiovascular urban environments: The Heart Healthy Hoods Index

Health and Place 55 (2019) 111–119

Contents lists available at ScienceDirect

Health & Place

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

A multicomponent method assessing healthy cardiovascular urban environments: The Heart Healthy Hoods Index

Alba Cebrecos[a,b], Francisco Escobar[a,b], Luisa N. Borrell[c], Julia Díez[a], Pedro Gullón[a], Xisca Sureda[a], Olivier Klein[d], Manuel Franco[a,e,]

[a] Social and Cardiovascular Epidemiology Research Group, School of Medicine, University of Alcalá, Alcalá de Henares, Madrid, Spain
[b] Department of Geology, Geography and Environmental Sciences, University of Alcalá, Alcalá de Henares, Madrid, Spain
[c] Department of Epidemiology and Biostatistics, City University of New York, New York, United States
[d] Department Urban Development and Mobility, LISER- Luxembourg Institute of Socio-Economic Research, Esch-sur-Alzette, Luxembourg
[e] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

ARTICLE INFO

Keywords:
Built environment
Cardiovascular diseases
Health
Residence characteristics
Geographic Information System
Spatial analysis
Spain

ABSTRACT

Previous studies have examined the built environment mostly focusing on a single exposure construct (e.g. walkability) to examine its association with health outcomes. This study developed a multicomponent Heart Healthy Hoods Index to characterize heart-healthy urban environments and examined its relationship with the prevalence of cardiovascular disease (CVD) in Madrid, Spain. Using spatial methods, we generated two index models (model 0 unweighted and model 1 weighted) using the percentage of deaths for the main behavioral risk factors for CVD (diet, physical activity, alcohol, and tobacco environments). We performed global (Ordinal Least Square) and local (Geographically Weighed Regression) regression analyses to assess the relationship between both index models and CVD prevalence, and to identify the best index model. In the global analysis, both models showed a significant negative relationship with CVD prevalence. In the local analysis, Model 1 removed the spatial autocorrelation of residuals and showed the lowest values for the Akaike information criterion. This study provides evidence of a non-stationary relationship between the heart-healthy urban environment and CVD prevalence. The HHH index may be an effective tool to identify and prioritize geographical areas for CVD prevention.

1. Introduction

Previous studies have generally characterized the built environment narrowly (i.e., a single construct of exposure) examining its association with behavioral risk factors and health outcomes (Lee et al., 2015; Lytle and Sokol, 2017; McCormack and Shiell, 2011; Popova et al., 2009). For instance, a recent systematic review described how the food environment impacts health-related outcomes, particularly obesity risk (Lytle and Sokol, 2017). Other studies have explored the associations between availability and accessibility of alcohol or tobacco outlets with drinking and tobacco behaviors, respectively (Lee et al., 2015; Popova et al., 2009). Moreover, another systematic review summarized the evidence from studies linking aspects of the built environment (i.e., street connectivity, land use mix, residential density, sidewalks or walkability index, among others) with behaviors related to physical activity (McCormack and Shiell, 2011). Therefore, the heterogeneity of these studies highlights the challenges inherent to measuring the built environment (Glass and McAtee, 2006).

Although the evidence on integrated measures of the built environment is increasing, this evidence mostly focuses on physical activity or food environments measures, and their relationship with overweight or obesity (DeWeese et al., 2018; Hobbs et al., 2018; Nau et al., 2015). However, recent studies have found spatial associations between built environment domains, socioeconomic status or deprivation, and health behaviors (D'Angelo et al., 2015; Schneider and Gruber, 2013; Shortt et al., 2015). These studies call attention to a new research area combining multidimensional measures of the health-related built environment. For example, a study in Germany found that

Correspondence to: Dpt of Surgery and Medical Sciences, University of Alcala, School of Medicine, Campus Universitario - C/ 19, Av. de Madrid, Km 33,600, Alcalá de Henares, 28871 Madrid, Spain.
E-mail addresses: alba.cebrecos@uah.es (A. Cebrecos), francisco.escobar@uah.es (F. Escobar), Luisa.Borrell@sph.cuny.edu (L.N. Borrell), julia.diez@uah.es (J. Díez), pedro.gullon@edu.uah.es (P. Gullón), francisca.sureda@uah.es (X. Sureda), olivier.klein@liser.lu (O. Klein), manuel.franco@uah.es (M. Franco).

https://doi.org/10.1016/j.healthplace.2018.11.010
Received 9 July 2018; Received in revised form 16 November 2018; Accepted 28 November 2018
Available online 11 December 2018

tobacco, alcohol and fast food outlets were more likely to be clustered in low-income neighborhoods (Schneider and Gruber, 2013). This study concluded that obesogenic and addictive environments may have a contextual influence on an individual's lifestyle and contribute to health risks. Similarly, a study in Glasgow, Scotland, showed that besides clusters of alcohol, tobacco, and fast food, gambling outlets were also found in the most deprived neighborhoods (Macdonald et al., 2018).

To address these challenges, the current work focuses on the intertwined nature of the multidimensional features of the health-related built environment. These environmental features do not act in isolation, but rather are the result of social forces affecting neighborhoods (Guthman, 2013). Although most research focuses on a single exposure construct, some authors have used a multidimensional approach. Examples of studies considering multiple factors found a significant association between these integrated measures and health-related, but not between a component of the measure and health outcomes (Kelly-Schwartz et al., 2004; Meyer et al., 2015; Wall et al., 2012). For instance, Hobbs et al. (2018) have assessed the obesogenic environment and documented interactions between the food and the physical environments (Hobbs et al., 2018). Saelens et al. (2012) found a significant association of physical activity and food environments combined only with obesity, suggesting that neighborhood characteristics may work together to build obesogenic environments (Saelens et al., 2012). Thus, it is imperative to integrate the different domains that constitute the urban (D'Angelo et al., 2015; Green et al., 2018; Guthman, 2013; Meyer et al., 2015).

In this context, the Heart Healthy Hoods (HHH) project (https://hhhproject.eu/) has examined how the urban environment relates to residents' cardiovascular health in Madrid (Spain) (Bilal et al., 2016; Franco et al., 2015). Specifically, the HHH project has examined the food (Bilal et al., 2018), the physical activity (Gullón et al., 2017) the alcohol (Sureda et al., 2017) (ref), and the tobacco environment (Valiente et al., 2018) separately. Thus, the aim of this study is twofold: first, to develop the HHH index, a multi-component index integrating the characteristics of heart-healthy urban environments (food, physical activity, tobacco, and alcohol) for small areas using Geographic Information Systems (GIS); and second, to examine the association of the index with the prevalence of cardiovascular diseases (CVD) at the area-level.

2. Methods

2.1. Study area

Madrid municipality is the capital city of Spain, located in the central area of the country with a population of 3165,235 inhabitants (INE, 2017). The city of Madrid is administratively divided into 21 districts further divided into 128 neighborhoods and 2415 census sections in 2014. The census section is the smallest administrative area for which population count data are available in Spain. An average census section comprises 1311 residents (ranging from 583 to 3865 residents) and a surface of 0.2 km^2 (ranging from 0.007 to 94.7 km^2). For analytical purposes, the unit of analysis was at the census section area-level.

2.2. Heart-healthy urban environmental data

The methodology used in this work is based on a previous pilot study (Cebrecos et al., 2016). Briefly, for the pilot study, we developed a synthetic index to characterize the obesogenic environment integrating measures of food and physical activity environments using twelve census sections of Madrid. For the current study, our aim was to characterize the cardiovascular environment for the city of Madrid, and thus, to consider of four environments shown to affect CVD risk (food, physical activity, alcohol, and tobacco). The geocoding of the spatial data and the whole development process of the index was developed using ArcGIS 10.1 software (ESRI, Redlands, California).

This section describes the indicators and secondary data sources used to create our index. All the data, except the walkability measure (collected at the census section level), were obtained in point format. Below we describe our indicators and data related to our framework for four domains of heart-healthy urban environments: food, physical activity, alcohol, and tobacco (Bilal et al., 2016).

2.2.1. Food environment data

To characterize the retail food environment, we obtained a city-wide retail spaces census from the Department of Statistics of Madrid City Council (Censo de Locales y Actividades). This database is publicly available and curated by the local government for licensing purposes including data on economic activities of all occupied commercial spaces. It includes outlet name, a classification of the outlet type, and locational information (latitude and longitude, along with the business address). This database is updated annually and includes for every commercial space, its registered economic activity based on the National Classification of Economic Activities 2009 coding system. This coding system, similar to the North American Industry Classification System, is the standard used to classify business establishments by Spanish statistical agencies.

Using these data, and consistent with previous research (Díez et al., 2016; Lake et al., 2010), we classified food stores into: 1) supermarkets (including discounters); 2) small grocery stores; 3) convenience stores; and 4) specialized stores (including fruit and vegetable stores, fish-mongers, butchers, and bakeries). Each of these store types were weighted according to their healthy food availability. For instance, fruits and vegetables stores (and other specialized stores such as fish-mongers) were considered the healthiest stores because they carry healthy products and lack processed and ultra-processed food (Bilal et al., 2016; Díez et al., 2016). Small grocery and convenience stores were considered the least healthy store types. We geocoded a total of 7771 food outlets and weighted each type according to their avail-ability of healthy food. We used this weighting for the Kernel Density Estimation (KDE) of the food environment when developing the HHH index.

2.2.2. Physical activity environment data

We used a composite walkability index developed by the research team for the city of Madrid. Further details of how this index was developed have been published elsewhere (Gullón et al., 2017). Briefly, this index includes residential density (occupied dwellings/ km2), population density (residents/ km2), retail destinations (retail and services destinations/ km2) and street connectivity (Kernel density in 3 m × 3 m pixels of the density of street intersections). All four indicators were equally weighted in the walkability index ranging from -13.2 to 7.61 and higher values indicating more walkable census sections. We used this weighting for the KDE to score the physical activity environment in the HHH index. Data came from different secondary databases, such as the Spanish Census (that includes data on occupied dwellings), the Padron (population density data), the retail spaces census (retail and commercial services where you can go walking) and CARTOCIUDAD (a National Mapping Agency initiative that collects and makes available official geo-referenced urban data, including street structure and administrative boundaries in shapefile format).

2.2.3. Alcohol environment data

We considered alcohol availability to characterize the alcohol environment including all off-premise alcohol outlets. Off-premises where defined as outlets where people purchase alcohol and included 1) supermarkets; 2) small grocery stores; 3) convenience stores; and 4) wine or liquor stores.

Table 1
Description of urban environment data, indicators and sources.

Environment	Measures	Indicators	Format	Year	Source
Food	Healthy food availability	Food store density Availability of food store diversity	Point	2014	Retail spaces census
Physical activity	Walkability index	Residential Density Population Density Retail Destinations Street Connectivity	Census section	2014	Housing census Padron Retail spaces census CARTOCIUDAD
Alcohol	Alcohol availability	Alcohol outlet density	Point	2014	Retail spaces census
Tobacco	Tobacco availability	Tobacco store density	Point	2014	Commissioner for the Tobacco Market

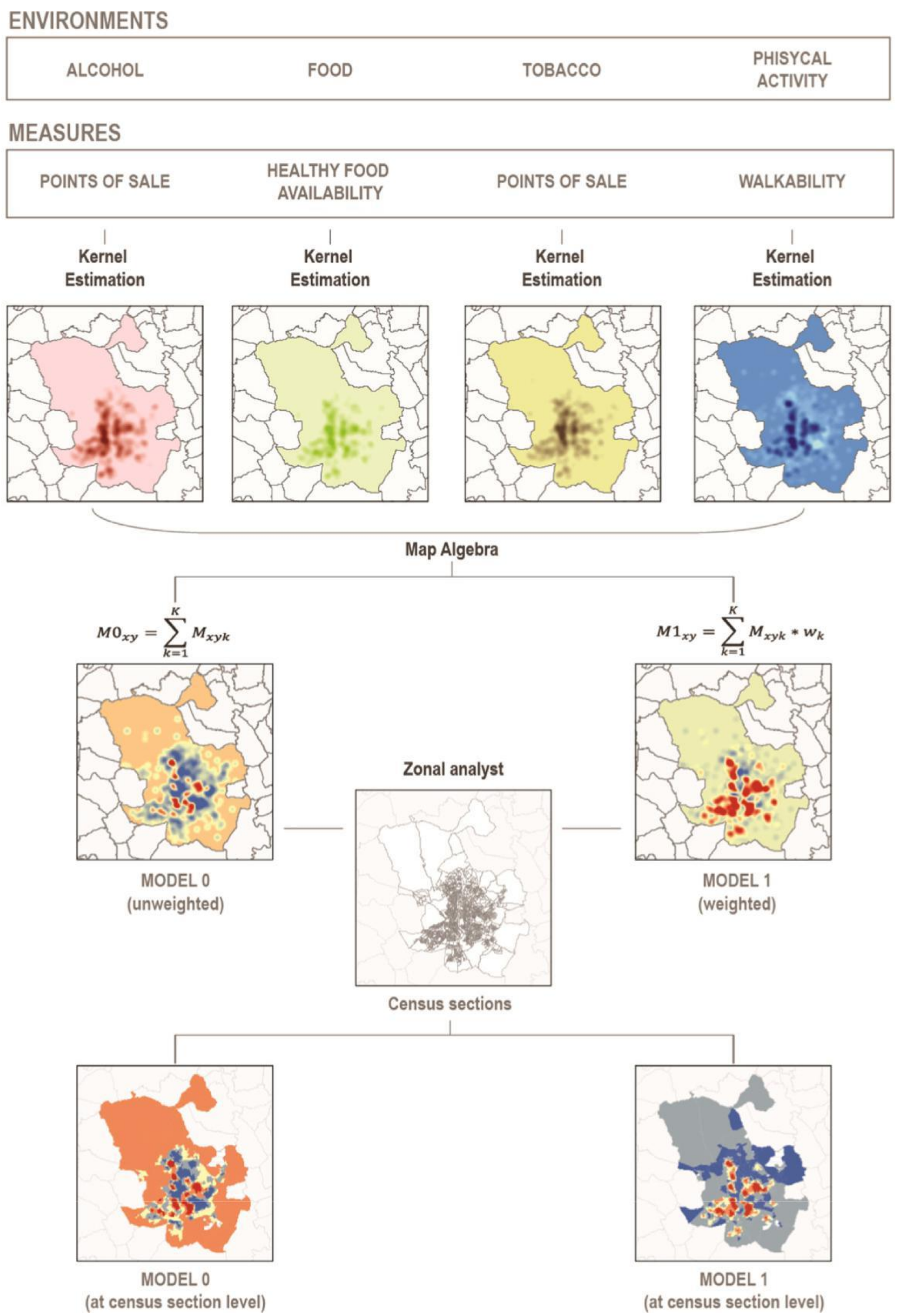


Fig. 1. GIS model for the construction of the HHH index.

Data came from the retail spaces census from the Department of Statistics of Madrid City Council (Censo de Locales y Actividades), de-scribed above (see food environment data). We geocoded a total of 4352 off-premise alcohol outlets and we use them to calculate the availability of off-premise alcohol outlets using an unweighted KDE.

2.2.4. Tobacco environment data

We used the availability of tobacco stores as a proxy of tobacco exposure in the urban environment. Information about the addresses of all tobacco stores in the city was provided by the Tobacco Market Commission (TMC) for the year 2014. TMC is a governmental organization tasked with monitoring the activity of tobacco stores and its spatial distribution to ensure the "free competition" of the tobacco market in Spain. We geocoded the 639 tobacco stores included in the database and we used these data to calculate the availability of tobacco stores using an unweighted KDE.

Table 1 describes each neighborhood domains used to characterize and assess the neighborhood environment.

2.3. Prevalence of cardiovascular disease

We used Electronic Health Records from the Primary Health Care System of Madrid (Bilal et al., 2016). This database included information on: (1) cardiovascular disease prevalence specified as a person with a diagnosis of myocardial infarction, angina, chronic ischemic heart disease, stroke or peripheral artery disease, (2) CVD risk factors (tobacco use, obesity, hypertension, diabetes mellitus, dyslipidemia) and (3) sociodemographic variables (age, sex) for 1,446,994 adults aged 45–70 residing in Madrid in 2014. For analytical purposes, we only considered data on CVD prevalence for this analysis. We removed all personally identifiable information (address, name, identifiers) and aggregated the results to the census section level. Finally, we calculated the age-standardized prevalence rate (by 1000 inhabitants). This indicator was used as our dependent variable in the regression models.

2.4. Neighborhood deprivation

We used an existing deprivation composite index to measure the area-level deprivation and considered as a confounder in the statistical analysis (Cebrecos et al., 2018). Briefly, this area-based deprivation index is constructed at the census section level using four indicators from 2011 Spanish Census data: 1) manual workers, 2) unemployment, 3) eventual employees and 4) insufficient instruction. These indicators are strongly related to social class (e.g. manual workers), and both material and social deprivation (e.g. unemployment or educational level). The deprivation index scores ranged from 1.93 to 3.36 with in-creasing values indicating greater deprivation.

For more details on the deprivation index construction please see (Cebrecos et al., 2018).

2.5. Design of index models

The method to build the model of the index is summarized as follows: 1) we used KDE to convert data from each urban environment dimension into a continuous surface for the entire city of Madrid. The KDE were integrated to ArcGIS employing the Silverman's quadratic Kernel function (Silverman, 1986); 2) we calculated z-scores to make each surface comparable with the rest; 3) we performed a map algebra analysis with all surfaces generated and standardized, to combine complete measurements; and 4) we developed two different index models: model 0 (unweighted) and model 1 (weighted) weighted for each indicator using mortality data related to the main CVD risk factors. Fig. 1 graphically depicts the approach used to create the HHH index, the multidimensional features included, the indicators selected, the data we acquired, and the map algebra computed.

2.5.1. MODEL 0 (unweighted model)

For this model, all environments (food, physical activity, alcohol, and tobacco) were equally weighted in the final characterization of the heart-healthy environment. We considered that if the urban environment was health-promoting (such as healthy food access or walkability), the sign was positive; and if the environment promoted unhealthy behaviors (tobacco stores and alcohol retailers), the sign was negative.
For constructing KDE, we applied a cell size of 10 m, and a fixed bandwidth of 1042 m. These distances were defined by observing the minimum and maximum distances between the points of sale, the minimum area of the census sections and the size of the study area. The operation adopted for map algebra was a local unweighted average computing the mean of pixels at the same location for all environment surfaces, treating all domains as equally weighted, where k is a sub-index for each of the domains (food, alcohol, physical activity, and tobacco); and x and y are sub-indices for the geographical coordinates of each cell.

$$M0_{xy} = \sum_{k=1}^{K} M_{xyk} \tag{1}$$

2.5.2. MODEL 1 (weighted model)

We wanted to assess the specific influence of each urban environment on the heart-healthy environment. We used the most recent Global Burden of Disease report as theoretical justification (http:// ghdx.healthdata.org/gbd-results-tool)for weighting each domain (food, physical activity, alcohol, and tobacco environments) by the proportion of CVD deaths attributable to the four related behavioral risk factors (unhealthy diet, physical inactivity, alcohol use and smoking). We acquired the 2016 data for Spain for adults aged 40–75 years (IHME, 2016). As shown in Table 2 of the supplemental material, the behavioral risk factor with the greatest impact on CVD was the dietary risk (46.3% of total deaths due to CVD), followed by tobacco (23.5%), alcohol (17.1%) and low levels of physical activity (6.9%). We used these proportions for the specific weight per environmental indicator. The weighted model was calculated as:

$$M1_{xy} = \sum_{k=1}^{K} M_{xyk} * w_k \tag{2}$$

In the map algebra analysis, each environmental domain was weighted by the proportion of CVD deaths attributable to each behavioral risk factor (w). For example, the food environment domain was weighted by the proportion of CVD death attributable to dietary risk factors.

Finally, to fully integrate the HHH index into the geographic context of the area, we assigned each census section a single value by means of zonal analysis. This analysis calculated a single output value for each

Table 2
Global regression modeling the relationship between both HHH index models - unweighted (Model 0) and weighted (model 1) – and the age-adjusted CVD prevalence rates in the city of Madrid, Spain.

	MODEL 0	*MODEL 1*
OLS_{crude}		
R2	0.000	0.001
β_{model}	-0.001	0.001
AICc	7441.64	7439.62
Moran's I $_{residual}$	0.278[a]	0.277[a]
$OLS_{adjusted}$		
R2	0.158	0.159
β_{model}	−0.037[b]	−0.002[b]
$\beta_{deprivation}$	0.459[b]	0.474[b]
AICc	7032.46	7028.65
Moran's I $_{residual}$	0.127[a]	0.124[a]

OLSc is the crude Ordinary Least Square analysis between CVD rates and the index models. OLSa is the Ordinary Least Square analysis between CVD rates and the index models adjusted by deprivation index.
[a] $p < 0.01$.
b $p < 0.05$.

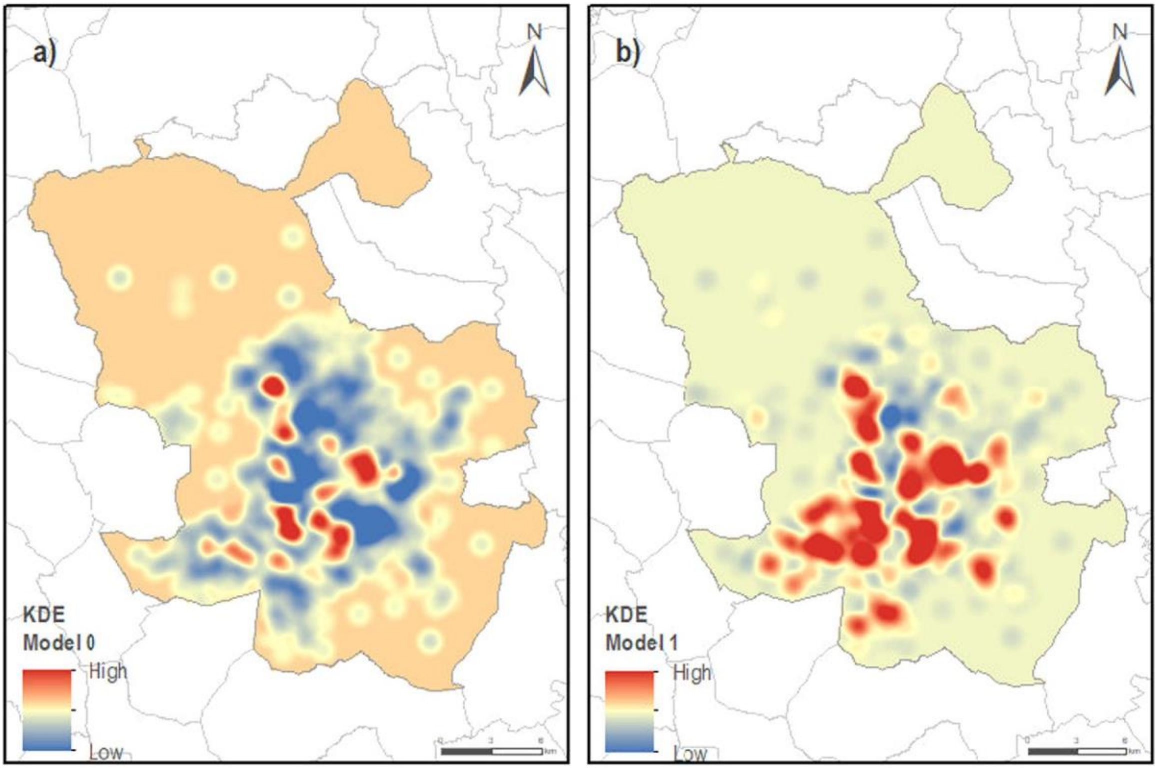


Fig. 2. Cartographic representation of the index models. Kernel Density Estimation surface for a) Model 0 and b) Model 1. High values of model (red color) represent heart-healthier environments in terms of food, physical activity, tobacco and alcohol. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

census section averaging all pixels that fall within each area. Thereby, we obtained both index models at the census section level (see Fig. 1). For interpretation of the HHH index, higher values show census sections that have better heart-healthy environments

We collected contextual information on the study area from the Spanish National Mapping Agency (www.ign.es) and Spanish National Spatial Data Infrastructure (www.idee.es), allowing us to generate a georeferenced database to integrate and map all the urban environment and cardiovascular data.

2.6. Spatial and statistical analysis

We examined the spatial autocorrelation (spatial dependency) of the dependent variable (CVD adjusted rate) and of the independent variables (both heart-healthy index models and the area-based deprivation index) by computing the Moran's Index. This analysis allows identifying whether the distribution pattern of each variable is either aggregated (a positive spatial correlation and statistically significant), dispersed (negative spatial correlation and statistically significant) or randomly distributed (without correlation) (Moran, 1948). The conceptualization of spatial relationship was first order polygon contiguity of edges and corners with row standardization.

We performed an Ordinary Least Squares (OLS) global regression to study the relationship between the HHH index models and the CVD prevalence rate, adjusting for area-level deprivation index (Charreire et al., 2010; Feng et al., 2010). First, we only considered the HHH index models as explanatory variables (OLS_{crude} or OLS_c); then we computed adjusted models for the socioeconomic deprivation ($OLS_{adjusted}$ or OLS_a).

The OLS model does not consider the underlying spatial dependence (spatial autocorrelation) and spatial heterogeneity (spatial non-stationarity) inherent to the data (Anselin 2002). Thus, to consider spatial non-stationarity in the relationship between the heart-healthy urban environment (calculated by our HHH index models) and the rates of CVD prevalence, we performed a Geographically Weighted Regression (GWR). GWR performs a set of local regressions with coefficients that vary in magnitude and direction according to location (Fotheringham et al., 2002). In addition, the coefficients can be mapped and visualized to explore the varying spatial relationship between health outcomes and explanatory variables (Gebreab, 2018).

We performed two GWR models. GWR_{crude} (GWR_C) assess the relationship between both HHH index models and age-adjusted CVD rates; whereas $GWR_{adjusted}$ (GWR_a) was adjusted for the deprivation index. In GWR analyses, we used a kernel with an adaptive Gaussian function to fit the size of bandwidth according to the distribution of the census sections. The kernel bandwidth was determined by minimizing the corrected Akaike's Information Criterion (AICc). The AICc is a relative goodness-of-fit statistics for comparing competing models, where the model with the smallest AICc provides the closest or best approximation to reality (Fotheringham et al., 2002; Gebreab, 2018). The GWR with the smallest AIC was considered to have a better fit, and a difference in AICc of more than 3 values was considered as a notable difference between two models (Fotheringham et al., 2002).

Finally, we studied the spatial autocorrelation of residuals using the Moran's Index. It is a useful indicator of problems in regression modeling (Fotheringham et al., 2002). In general, a random distribution of the residuals indicates a more effective regression model. The conceptualization of spatial relationship was the same that the previous Moran's Index analysis. We used ArcGIS 10.1 software (ESRI, Redlands, California) to perform the spatial association analysis, the GWR and the cartographic results.

3. Results

3.1. Results of both HHH index models

The cartographic representation of the models calculated is shown in Fig. 2: The left-hand side shows results from model 0 (weighting the heart-healthy environments equally) whereas the right-hand side shows results from model 1, specifically weighting each environmental indicator. Red tones represent areas with healthier environments, whereas blue tones represent unhealthier urban environments for cardiovascular health. We found a strong correlation between models 0 and 1 ($r = 0.82$; $p < 0.01$).

The most heart-healthy environments (high values in red tones) were observed, for both models, in the city core. In model 0, these heart-healthy hot spots (in red tones) were found to be smaller and more isolated. Model 0 characterized most of the built environment of Madrid with low values (blue colors). In contrast, model 1 showed larger and more connected heart-healthy areas in the urban core (shown in red) and worse values in the less populated surrounding areas.

3.2. Spatial analysis results

The global autocorrelation analysis, calculated by the Moran's Index, showed a positive association ($p < 0.01$) for both explanatory and dependent variables. These results, shown in Table 3 of the supplemental material, show the existing spatial dependency within the variables.

Table 2 shows the results of the OLS global regression analysis. The HHH index weighted (model 1) presented a higher R2 in the regression adjusted by deprivation (OLS_a). OLS_a, showed lower AIC_c values and a significant negative change for the age-adjusted CVD rate. In OLS_a, a one unit increase in the heart-healthy environment decreases the ad-justed rate of CVD by 0.002 cases per 1000 inhabitants. However, we found spatial dependency of residual for these analyses.

Table 3 shows the results of the GWR. Compared to the OLS global regression models, the GWR models showed smaller AICc values and higher R^2 for all the analyses. The spatial autocorrelation of residuals was eliminated only in index model 1 after adjusting for socioeconomic deprivation (GWR_a). In addition, GWR_a in model 1 presented the lowest AICc values. In addition, GWR_a showed a non-stationarity relationship between heart healthy environment and the age-adjusted CVD rates, with this relationship varying across space in direction and magnitude (Fig. 3).

Table 3
Local regression modeling the relationship between both HHH index models - unweighted (Model 0) and weighted (model 1) – and the age-adjusted CVD prevalence rates in the city of Madrid, Spain.

	MODEL 0	MODEL 1
GWR_{crude}		
R^2	0.177	0.179
β_{model}	−0.667, 6.094	−0.683, 0.053
AICc	7003.94	6999.76
Moran's $I_{residual}$	0.118[a]	0.117[a]
$GWR_{adjusted}$		
R^2	0.336	0.324
β_{model}	−0.555, 0.415	−0.023, 0.022
$\beta_{deprivation}$	−0.349, 1.01	−0.309, 0.939
AICc	6741.19	6732.24
Moran's I $_{residual}$	−0.020[b]	−0.015

WGR_c is the crude Geographically Weighted Regression analysis between CVD rates and the index models. WGR_a is the Geographically Weighted Regression between CVD rates and the index models adjusted by deprivation index.

[a] $p < 0.01$.

[b] $p < 0.1$.

Fig. 3 shows the results of the GWR_a analysis for model 1. The figure allows simultaneously studying the adjustment (R^2) and the coefficients (β) of model 1. Fig. 3a shows higher R^2 in the southeast of the study area. In the southern area, the greatest R^2 occurs in areas with negative values of the index model (Fig. 3b). In other words, in the south of Madrid, a healthier cardiovascular environment is associated with a reduction of the adjusted rates of CVD. In some areas of southern Madrid, a one unit increase in the cardiovascular environment is associated with a decline in CVD prevalence of 0.023. However, in other parts of Madrid, the relationship is positive, suggesting higher CVD prevalence in areas with better heart-healthy environments.

4. Discussion

We developed a multicomponent index that integrates four dimensions related to heart-healthy urban environments: the food, physical activity, tobacco, and alcohol environments; and used two models for this Heart Healthy Hoods (HHH) index: one equally weighting all environmental domains; and one specifically weighting each domain. We used two methods to examine the relationship between the HHH index and the prevalence of CVD, the OLS regression analysis and the GWR. The weighted model for the HHH index shows a better fit in both methods. As expected, the GWR shows a better fit as indicated by the lowest values of AIC, and it reveals a non-stationary relationship be-tween the heart-healthy environment and CVD prevalence. In addition, the weighted model better captured the spatial variation in neighbor-hood environments, underscoring a remarkable difference between the connected heart-healthy areas in the urban core and the less populated surrounding areas. Specifically, we found areas with inverse relation-ships showing that improving the cardiovascular environment would contribute to reduced prevalence rates. However, we also identified areas with a positive relationship between HHH index models and CVD prevalence rates.

Several studies have examined the effects of a single built environment measure on health behaviors and health outcomes. Examples are the studies about food environment (Lytle and Sokol, 2017), alcohol (Popova et al., 2009), tobacco (Lee et al., 2015) or physical activity (McCormack and Shiell, 2011). The novelty of our study lies in the use of an integrated approach to combine all those characteristics of the built environment. These characteristics may act synergistically to influence health-related behaviors (Nelson et al., 2006), which would also influence health outcomes. In addition, our work goes beyond previous studies, as it includes clinical data on cardiovascular health rather than overall mortality or self-reported health (Chalkias et al., 2013; Chen et al., 2010; Clary et al., 2016; Feuillet et al., 2016).

While we did not find past research on multicomponent indexes integrating measures of the built environment and cardiovascular health, we found works that have integrated measures of the built environment and obesity (Tseng et al., 2014; Wall et al., 2012). For in-stance, Tseng et al. (2014) created an obesogenic index (covering 3 constructs: food, activity resources and walkability) to quantify neighborhood obesogenicity and examined its association with body mass index (BMI) in socio-economically disadvantaged neighborhoods in Australia. Consistent with our study, they found a spatial heterogeneity in the relationship between their index and the health outcomes (non-stationary relationship), with generally positive associations for BMI in urban areas and inverse associations in rural areas. Moreover, Wall et al. (2012) integrated neighborhood measures of food, physical ac-tivity, street/transportation, and socioeconomic characteristics to ex-amine their associations with adolescent weight status (BMI). Using a spatial latent class analysis, they created composite scores representing closely correlated neighborhood characteristics that potentially relate to obesity (i.e., away-from-home food and recreation accessibility, community disadvantage, green space, retail/transit density, and supermarket accessibility). Despite using a different methodology, Wall et al. (2012) found results similar to ours identifying clusters with

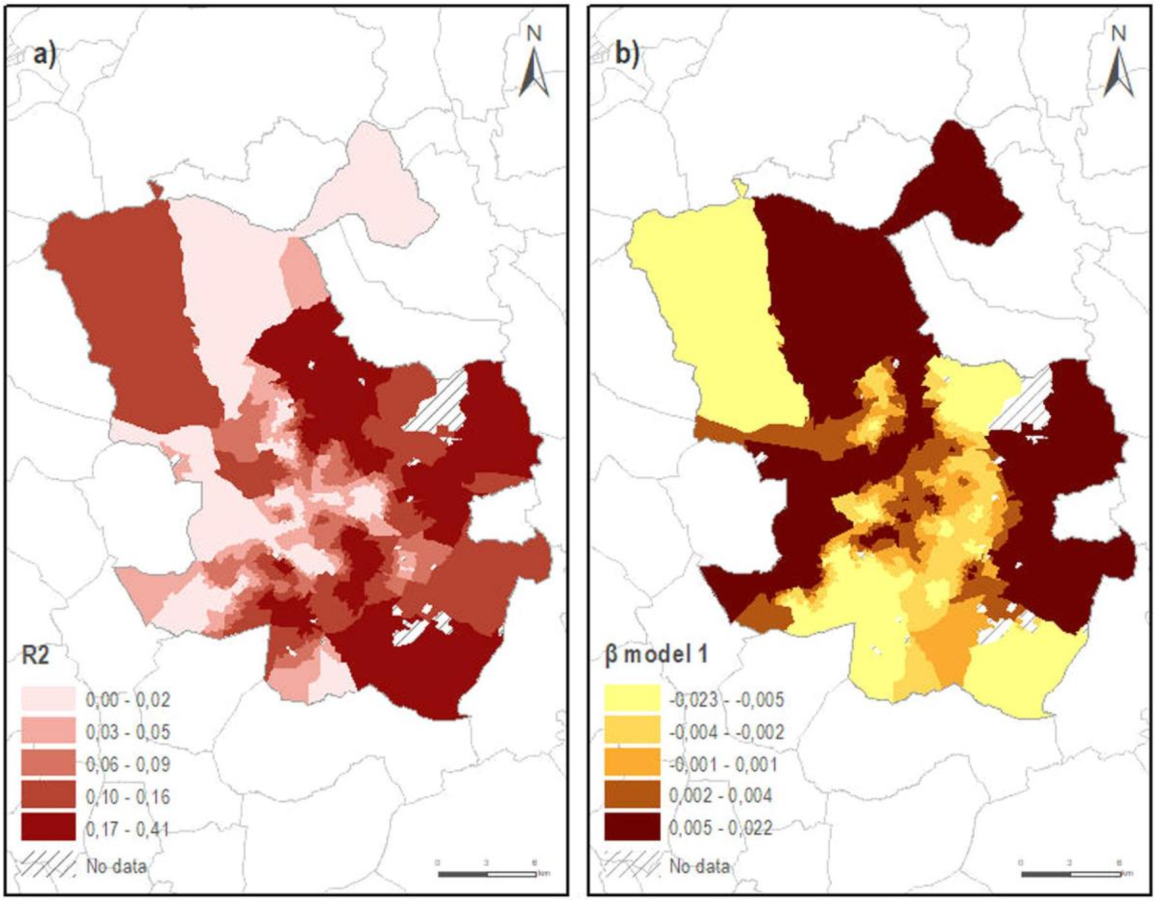


Fig. 3. Results of geographically weighed regression between the heart-healthy environment and age-adjusted CVD prevalence, adjusting for socioeconomic de-privation. 2a) Local regression adjustment (R2). 2b) Local regression coefficients (β) for heart-healthy environment.

complex combinations of both positive and negative environmental influences on health outcomes. Thus, our findings concur with the existing evidence that urban environments may influence CVD risk factors.

Previous studies (Borrell et al., 2014; Havard et al., 2008; Shortt et al., 2015) led us to consider that socioeconomic deprivation as an important factor in these analyses. It is possible that environments with healthy cardiovascular resources could mitigate the age-standardized CVD rate but not eliminate it completely (Clary et al., 2016; Reed et al., 2013). In environments with a diversity of options (healthy and/or unhealthy), socioeconomic or personal influences may buffer the influence of exposure to the environment (Lytle, 2009). Our findings showed that controlling for the deprivation index may mitigate the effect of the HHH index on CVD risk. Alongside model misspecifications, non-stationarity may therefore also stem from intrinsic differences in the way individuals respond to specific characteristics of their local environment (Fotheringham et al., 2002).

Several limitations of this study should be noted. For the measurement of urban domains, a very specific set of indicators were selected. Therefore, we may be missing other environmental characteristics that could influence residents' health behaviors (Hillsdon et al., 2006; Jeffery et al., 2006). For example, we only considered the food and alcohol retailers without considering food services (e.g. fast-foods) or in-premises (bars). This may be underestimating the impact of the urban environment in terms of alcohol and unhealthy food exposure. Furthermore, when GIS measures of the urban environment are derived from secondary database there could be inaccuracies (Liese et al., 2013, 2010). However, we have no reasons to believe that inaccuracies affect one more than another area, and thus, the likelihood of bias may be non-differential throughout the study area.

Another inherent limitation to the use of aggregate data in area units is the Modifiable Areal Unit Problem or MAUP (Openshaw, 1983). In this study, we used the census section to aggregate environmental measures. However, the measures and associations of the urban environment may differ according to the geographical delimitation used (Barnes et al., 2016). An example of these differences is shown by a work developed in Madrid (Cebrecos et al., 2018), where an index of deprivation varied when calculated for different administrative units as well as its association with CVD. The border effect is a limitation that arises when conducting studies infinite regions (Griffith, 1983). The urban environment domains considered, CVD prevalence and deprivation, all extend beyond geographical delimitations of Madrid municipality. Yet, we did not consider the vicinity outside the municipality to perform the spatial association analysis because we had no data outside the municipality of Madrid. The entire city of Madrid is part of the regional electronic health record system that is also universal and, therefore, the possibility of undiagnosed CVD is minimal, but not impossible.

Despite these limitations, this study has several strengths. First, the use of composite indexes reduces collinearity and over-adjustment, confers ease of interpretation, and may reduce measurement errors (Feng et al., 2010). Second, integrating different indicators within an index can detect associations not previously found (Kelly-Schwartz et al., 2004). To reduce the impact that the built environment has on health behaviors, it is important to understand the mechanisms driving them independently and/or jointly. Thus, these index models and the cartography developed can provide an effective analytical tool to sup-port public health policy decisions in that identifies the geographical areas where interventions should be prioritized, those with the most significant relationship. Lastly and most importantly, we used and compared two methodological approaches, OLS global and GWR local regression analyses. Methodologically, the GWR analysis, as a heuristic approach (Wheeler and Paez, 2010), provides more useful information about the relationships between CVD and the heart-healthy environment than the OLS analysis. Thus, we consider very useful and necessary to complete the results of a global analysis with a local analysis that take into account the spatial dependence and non-stationary nature of the data. The GWR analysis improved the ability to explain local CVD rates. Unlike the OLS global model, GWR findings allowed us to identify locally what type of relationship exists between the heart-healthy environment and CVD and where this relationship is stronger. This modeling approach has the potential for developing future public health programs and interventions accounting for contextual characteristics to reduce and prevent CVD.

To our knowledge, this is the first study examining a wide variety of neighborhood contextual characteristics influencing the cardiovascular health of Madrid's citizens. The study area of the current work is limited to the city of Madrid, but we would like to underscore the international application that this type of index may have. Comprehensive studies of urban health in Europe (Macdonald et al., 2018; Schneider and Gruber, 2013), the USA (Auchincloss et al., 2008; Brenner et al., 2015; Moore et al., 2008) and Australia (Feng et al., 2018; Koohsari et al., 2016) could implement this methodology proposed to characterize urban environments in relation to cardiovascular health when secondary data are available.

Considering the balance or combination of resources can be theoretically and empirically important to understand the influence of the built environment on health behaviors. Consistent with theories of behavioral economics (Hursh and Roma, 2013; Reed et al., 2013), human behavior is the result of the distribution and availability of a diverse and synergistic set of resources. However, most investigations examine discrete characteristics of exposure to the environment (Caspi et al., 2012; Malambo et al., 2016). Specific exposures can be more easily addressed from public health policies but it may be a more ac-curate reflection considering human behavior within a system of multiple environmental characteristics.

5. Conclusion

The Heart Healthy Hoods Index is a multicomponent method to characterize, in an exhaustive way, heart-healthy urban environments, using the example of the city of Madrid. In addition, we tested the index models by studying their spatial relationship with CVD prevalence highlighting variations of this association through spatial local analysis. We found that the relationship between CVD and the urban environment was non-stationary but varied along the territory of Madrid both in direction and strength. The index and the cartography associated can provide an effective analytical tool to support public health policy decisions in that identifies the geographical areas where preventive and health promotion interventions for heart diseases might be prioritized. The Heart Healthy Hoods Index, could be a relevant tool to identify urban environments susceptible to cardiovascular preventive interventions.

Acknowledgments

We would like to thank the whole Social and Cardiovascular Epidemiology Research Group, for his collaboration and ideas on the study.

Conflict of interest

The authors of this paper do not report any conflict of interests.

Funding

This work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/ 2007–2013)/ERC Starting Grant Heart Healthy Hoods Agreement no. 336893.

References

Auchincloss, A.H., Diez Roux, A.V., Brown, D.G., Erdmann, C.A., Bertoni, A.G., 2008. Neighborhood resources for physical activity and healthy foods and their association with insulin resistance. Epidemiology. https://doi.org/10.1097/EDE. 0b013e31815c480.
Barnes, T.L., Colabianchi, N., Hibbert, J.D., Porter, D.E., Lawson, A.B., Liese, A.D., 2016. Scale effects in food environment research: implications from assessing socio-economic dimensions of supermarket accessibility in an eight-county region of South Carolina. Appl. Geogr. https://doi.org/10.1016/j.apgeog.2016.01.004.
Bilal, U., Díez, J., Alfayate, S., Gullón, P., Del Cura, I., Escobar, F., Sandín, M., Franco, M., 2016. Population cardiovascular health and urban environments: the Heart Healthy Hoods exploratory study in Madrid, Spain. BMC Med. Res. Methodol. 16, 1–12. https://doi.org/10.1186/s12874-016-0213-4.
Bilal, U., Jones-Smith, J., Diez, J., Lawrence, R.S., Celentano, D.D., Franco, M., 2018. Neighborhood social and economic change and retail food environment change in Madrid (Spain): the heart healthy hoods study. Heal. Place 51. https://doi.org/10.1016/j.healthplace.2018.03.004.
Borrell, C., Marí-Dell'olmo, M., Palència, L., Gotsens, M., Burström, B.O., Domínguez-Berjón, F., Rodríguez-Sanz, M., Dzúrová, D., Gandarillas, A., Hoffmann, R., Kovacs, K., Marinacci, C., Martikainen, P., Pikhart, H., Corman, D., Rosicova, K., Saez, M., Santana, P., Tarkiainen, L., Puigpinós, R., Morrison, J., Pasarín, M.I., Díez, È., 2014. Socioeconomic inequalities in mortality in 16 European cities. Scand. J. Public Health. https://doi.org/10.1177/1403494814522556.
Brenner, A.B., Diez Roux, A.V., Barrientos-Gutierrez, T., Borrell, L.N., 2015. Associations of alcohol availability and neighborhood socioeconomic characteristics with drinking: cross-sectional results from the Multi-Ethnic Study of Atherosclerosis (MESA). Subst. Use Misuse. https://doi.org/10.3109/10826084.2015.1027927.
Caspi, C.E., Sorensen, G., Subramanian, S.V., Kawachi, I., 2012. The local food environment and diet: a systematic review. Health Place. https://doi.org/10.1016/j.healthplace.2012.05.006.
Cebrecos, A., Díez, J., Gullón, P., Bilal, U., Franco, M., Escobar, F., 2016. Characterizing physical activity and food urban environments: a GIS-based multicomponent proposal. Int. J. Health Geogr. 15. https://doi.org/10.1186/s12942-016-0065-5.
Cebrecos, A., Domínguez-Berjón, M.F., Duque, I., Franco, M., Escobar, F., 2018. Geographic and statistic stability of deprivation aggregated measures at different spatial units in health research. Appl. Geogr. 95. https://doi.org/10.1016/j.apgeog. 2018.04.001.
Charreire, H., Casey, R., Salze, P., Simon, C., Chaix, B., Banos, A., Badariotti, D., Weber, C., Oppert, J.M., 2010. Measuring the food environment using geographical in-formation systems: A methodological review. Public Health Nutr. https://doi.org/10.1017/S1368980010000753.
Chalkias, C., Papadopoulos, A.G., Kalogeropoulos, K., Tambalis, K., Psarra, G., Sidossis, L., 2013. Geographical heterogeneity of the relationship between childhood obesity and socio-environmental status: empirical evidence from Athens, Greece. Appl.Geogr. https://doi.org/10.1016/j.apgeog.2012.10.007.
Chen, V.Y.J., Wu, P.C., Yang, T.C., Su, H.J., 2010. Examining non-stationary effects of social determinants on cardiovascular mortality after cold surges in Taiwan. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2009.11.044.
Clary, C., Lewis, D.J., Flint, E., Smith, N.R., Kestens, Y., Cummins, S., 2016. The local food environment and fruit and vegetable intake: a geographically weighted regression approach in the ORiEL study. Am. J. Epidemiol. https://doi.org/10.1093/aje/ kww073.
D'Angelo, H., Evenson, K.R., Rose, S.W., Fleischhacker, S., Myers, A.E., Ribisl, K.M., 2015. Examination of community and consumer nutrition, tobacco and physical activity environments at food and tobacco retail stores in three diverse North Carolina communities. Prev. Med. Rep. https://doi.org/10.1016/j.pmedr.2015.08.018.
DeWeese, R.S., Ohri-Vachaspati, P., Adams, M.A., Kurka, J., Han, S.Y., Todd, M., Yedidia, M.J., 2018. Patterns of food and physical activity environments related to children's food and activity behaviors: a latent class analysis. Heal. Place. https://doi.org/10.1016/j.healthplace.2017.11.002.
Díez, J., Bilal, U., Cebrecos, A., Buczynski, A., Lawrence, R.S., Glass, T., Escobar, F., Gittelsohn, J., Franco, M., 2016. Understanding differences in the local food environment across countries: a case study in Madrid (Spain) and Baltimore (USA). Prev. Med. 89. https://doi.org/10.1016/j.ypmed.2016.06.013.
Feng, J., Glass, T.A., Curriero, F.C., Stewart, W.F., Schwartz, B.S., 2010. The built environment and obesity: a systematic review of the epidemiologic evidence. Heal. Place. https://doi.org/10.1016/j.healthplace.2009.09.008.
Feng, X., Astell-Burt, T., Badland, H., Mavoa, S., Giles-Corti, B., 2018. Modest ratios of fast food outlets to supermarkets and green grocers are associated with higher body mass index: Longitudinal analysis of a sample of 15,229 Australians aged 45 years and older in the Australian National Liveability study. Heal. Place. https://doi.org/10.1016/j.healthplace.2017.10.004.
Feuillet, T., Salze, P., Charreire, H., Menai, M., Enaux, C., Perchoux, C., Hess, F., Kesse-Guyot, E., Hercberg, S., Simon, C., Weber, C., Oppert, J.M., 2016. Built environment in local relation with walking: why here and not there? J. Transp. Heal. https://doi.

org/10.1016/j.jth.2015.12.004.
Fotheringham, A.S.S., Brunsdon, C., Charlton, M., 2002. Geographically weighted re-gression: the analysis of spatially varying relationships. Notes. https://doi.org/10. 1111/j.1538-4632.2003.tb01114.x.
Franco, M., Bilal, U., Diez-Roux, A.V., 2015. Preventing non-communicable diseases through structural changes in urban environments. J. Epidemiol. Community Health. https://doi.org/10.1136/jech-2014-203865.
Gebreab, S.Y., 2018. Statistical methods in spatial epidemiology. In: Duncan, D.T., Kawachi, I. (Eds.), Neighborhoods and Health. Oxford University Press, New York, pp. 371.
Glass, T.A., McAtee, M.J., 2006. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. Soc. Sci. Med. https://doi.org/10.1016/j.socscimed.2005.08.044.
Green, M.A., Daras, K., Davies, A., Barr, B., Singleton, A., 2018. Developing an openly accessible multi-dimensional small area index of 'Access to Healthy Assets and Hazards' for Great Britain, 2016. Health Place 54, 11–19. https://doi.org/10.1016/J.HEALTHPLACE.2018.08.019.
Griffith, D.A., 1983. The boundary value problem in spatial statistical analysis. J. Reg. Sci.
Gullón, P., Bilal, U., Cebrecos, A., Badland, H.M., Galán, I., Franco, M., 2017. Intersection of neighborhood dynamics and socioeconomic status in small-area walkability: the
Heart Healthy Hoods project. Int. J. Health Geogr. 16. https://doi.org/10.1186/ s12942-017-0095-7.
Guthman, J., 2013. Too much food and too little sidewalk? Problematizing the obeso-genic environment thesis. Environ. Plan. A. https://doi.org/10.1068/a45130.
Havard, S., Deguen, S., Bodin, J., Louis, K., Laurent, O., Bard, D., 2008. A small-area index of socioeconomic deprivation to capture health inequalities in France. Soc. Sci. Med. https://doi.org/10.1016/j.socscimed.2008.09.031.
Hillsdon, M., Panter, J., Foster, C., Jones, A., 2006. The relationship between access and quality of urban green space with population physical activity. Public Health. https://doi.org/10.1016/j.puhe.2006.10.007.
Hobbs, M., Griffiths, C., Green, M.A., Jordan, H., Saunders, J., McKenna, J., 2018. Neighbourhood typologies and associations with body mass index and obesity: a cross-sectional study. Prev. Med. https://doi.org/10.1016/j.ypmed.2017.11.024.
Hursh, S.R., Roma, P.G., 2013. Behavioral economics and empirical public policy. J. Exp.Anal. Behav. https://doi.org/10.1002/jeab.7.
IHME, 2016. Global health data exchange [www document]. Inst. Heal. Metr. Eval. ⟨http://ghdx.healthdata.org/gbd-results-tool⟩.
INE, 2017. Cifras oficiales de población resultantes de la revisión del Padrón municipal a 1 de enero 2014 [www document]. Inst. Nac. Estad. ⟨http://www.ine.es/jaxiT3/Tabla.htm?t=2881&L=0⟩.
Jeffery, R.W., Baxter, J., McGuire, M., Linde, J., 2006. Are fast food restaurants an en-vironmental risk factor for obesity? Int. J. Behav. Nutr. Phys. Act. https://doi.org/10. 1186/1479-5868-3-2.
Kelly-Schwartz, A.C., Stockard, J., Doyle, S., Schlossberg, M., 2004. Is sprawl unhealthy? A multilevel analysis of the relationship of metropolitan sprawl to the health of in-dividuals. J. Plan. Educ. Res. https://doi.org/10.1177/0739456X04267713.
Koohsari, M.J., Sugiyama, T., Mavoa, S., Villanueva, K., Badland, H., Giles-Corti, B., Owen, N., 2016. Street network measures and adults' walking for transport: appli-cation of space syntax. Heal. Place. https://doi.org/10.1016/j.healthplace.2015.12. 009.
Lake, A.A., Burgoine, T., Greenhalgh, F., Stamp, E., Tyrrell, R., 2010. The foodscape: classification and field validation of secondary data sources. Heal. Place. https://doi.org/10.1016/j.healthplace.2010.02.004.
Lee, J.G.L., Henriksen, L., Rose, S.W., Moreland-Russell, S., Ribisl, K.M., 2015. A sys-tematic review of neighborhood disparities in point-of-sale tobacco marketing. Am. J. Public Health. https://doi.org/10.2105/AJPH.2015.302777.
Liese, A.D., Barnes, T.L., Lamichhane, A.P., Hibbert, J.D., Colabianchi, N., Lawson, A.B., 2013. Characterizing the food retail environment: impact of count, type, and geos-patial error in 2 secondary data sources. J. Nutr. Educ. Behav. https://doi.org/10.1016/j.jneb.2013.01.021.
Liese, A.D., Colabianchi, N., Lamichhane, A.P., Barnes, T.L., Hibbert, J.D., Porter, D.E., Nichols, M.D., Lawson, A.B., 2010. Validation of 3 food outlet databases: complete-ness and geospatial accuracy in rural and urban food environments. Am. J.Epidemiol. https://doi.org/10.1093/aje/kwq292.
Lytle, L.A., 2009. Measuring the food environment. State of the science. Am. J. Prev. Med. https://doi.org/10.1016/j.amepre.2009.01.018.
Lytle, L.A., Sokol, R.L., 2017. Measures of the food environment: a systematic review of the field, 2007–2015. Heal. Place. https://doi.org/10.1016/j.healthplace.2016.12. 007.
Macdonald, L., Olsen, J.R., Shortt, N.K., Ellaway, A., 2018. Do 'environmental bads' such as alcohol, fast food, tobacco, and gambling outlets cluster and co-locate in more deprived areas in Glasgow city, Scotland? Heal. Place. https://doi.org/10.1016/j.healthplace.2018.04.008.
Malambo, P., Kengne, A.P., De Villiers, A., Lambert, E.V., Puoane, T., 2016. Built environment, selected risk factors and major cardiovascular disease outcomes: a systematic review. PLoS One. https://doi.org/10.1371/journal.pone.0166846.
McCormack, G.R., Shiell, A., 2011. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. Int. J. Behav. Nutr. Phys. Act. ⟨https://doi.org/1479-5868-8-125[pii]10.1186/1479-5868-8-125⟩.
Meyer, K.A., Boone-Heinonen, J., Duffey, K.J., Rodriguez, D.A., Kiefe, C.I., Lewis, C.E., Gordon-Larsen, P., 2015. Combined measure of neighborhood food and physical activity environments and weight-related outcomes: the CARDIA study. Heal. Place. https://doi.org/10.1016/j.healthplace.2015.01.004.
Moore, L.V., Diez Roux, A.V., Nettleton, J.A., Jacobs, D.R., 2008. Associations of the local food environment with diet quality–a comparison of assessments based on surveys and geographic information systems: the multi-ethnic study of atherosclerosis. Am. J. Epidemiol. https://doi.org/10.1093/aje/kwm394.
Moran, P.A., 1948. The interpretation of statistical maps. J. R. Stat. Soc. Ser. B. https://doi.org/10.2307/2983777.
Nau, C., Ellis, H., Huang, H., Schwartz, B.S., Hirsch, A., Bailey-Davis, L., Kress, A.M., Pollak, J., Glass, T.A., 2015. Exploring the forest instead of the trees: an innovative method for defining obesogenic and obesoprotective environments. Heal. Place. https://doi.org/10.1016/j.healthplace.2015.08.002.
Nelson, M.C., Gordon-Larsen, P., Song, Y., Popkin, B.M., 2006. Built and social environments. Associations with adolescent overweight and activity. Am. J. Prev. Med. https://doi.org/10.1016/j.amepre.2006.03.026.
Openshaw, S., 1983. The modifiable area unit problem. Concepts Tech. Mod. Geogr. https://doi.org/10.1177/1077558707312501.
Popova, S., Giesbrecht, N., Bekmuradov, D., Patra, J., 2009. Hours and days of sale and density of alcohol outlets: impacts on alcohol consumption and damage: a systematic review. Alcohol Alcohol. https://doi.org/10.1093/alcalc/agp054.
Reed, D.D., Niileksela, C.R., Kaplan, B.A., 2013. Behavioral economics: a tutorial for behavior analysts in practice. Behav. Anal. Pract. https://doi.org/10.1007/ BF03391790.
Saelens, B.E., Sallis, J.F., Frank, L.D., Couch, S.C., Zhou, C., Colburn, T., Cain, K.L., Chapman, J., Glanz, K., 2012. Obesogenic neighborhood environments, child and parent obesity: the neighborhood impact on kids study. Am. J. Prev. Med. https://doi.org/10.1016/j.amepre.2012.02.008.
Schneider, S., Gruber, J., 2013. Neighbourhood deprivation and outlet density for tobacco, alcohol and fast food: first hints of obesogenic and addictive environments in Germany. Public Health Nutr. https://doi.org/10.1017/S1368980012003321.
Shortt, N.K., Tisch, C., Pearce, J., Mitchell, R., Richardson, E.A., Hill, S., Collin, J., 2015. A cross-sectional analysis of the relationship between tobacco and alcohol outlet density and neighbourhood deprivation. BMC Public Health. https://doi.org/10. 1186/s12889-015-2321-1.
Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman Hall, London. ⟨https://doi.org/10.2307/2347507⟩.
Sureda, X., Espelt, A., Villalbí, J.R., Cebrecos, A., Baranda, L., Pearce, J., Franco, M., 2017. Development and evaluation of the OHCITIES instrument: assessing alcohol urban environments in the Heart Healthy Hoods project. BMJ Open 7. https://doi.org/10.1136/bmjopen-2017-017362.
Tseng, M., Thornton, L.E., Lamb, K.E., Ball, K., Crawford, D., 2014. Is neighbourhood obesogenicity associated with body mass index in women? Application of an obesogenicity index in socioeconomically disadvantaged neighbourhoods. Heal. Place. https://doi.org/10.1016/j.healthplace.2014.07.012.
Valiente, R., Sureda, X., Bilal, U., Navas-Acien, A., Pearce, J., Franco, M., Escobar, F., 2018. Regulating the local availability of tobacco retailing in Madrid, Spain: a GIS study to evaluate compliance. Tob. Control. https://doi.org/10.1136/tobaccocontrol-2018-054269.
Wall, M.M., Larson, N.I., Forsyth, A., Van Riper, D.C., Graham, D.J., Story, M.T., Neumark-Sztainer, D., 2012. Patterns of obesogenic neighborhood features and adolescent weight: a comparison of statistical approaches. Am. J. Prev. Med. https://doi.org/10.1016/j.amepre.2012.02.009.
Wheeler, D.C., Paez, A., 2010. Geographically weighted regression. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. . ⟨https://doi.org/10. 1007/978-3-642-03647-7⟩.

2. CONTEXT AND RESULTS OVERVIEW

The first study of this PhD dissertation was to develop an innovative method based on GIS to evaluate the obesogenic environment of a "medium" neighborhood of Madrid. This neighborhood was used as the pilot area of the HHH project and was formed by 12 census sections (Bilal, Díez, *et al.*, 2016). Through the development of this synthetic index, measures of the food and physical activity environments that were generated by a KDE were integrated into a composite index. The index ranged from 0 to 100, where higher scores represented a healthier environment. There was a heterogeneous spatial distribution of the obesogenic contextual determinants. In the study area, an average score of 25.5 was obtained. About 75% of the area had scores below 36.8, and 50%, below 25.5. The study area was also characterized in four categories according to the average score of each census section: low (17.7-21.6), medium-low (21.7-30.8), medium-high (30.9-35.1) and high (35.2- 43.8). Four of the 12 census sections were classified as low, four as medium-low, three as medium-high and one as high.

The second study addressed the implications that the MAUP has on the choices of the unit of analysis and the scale when data of socioeconomic deprivation is aggregated at different areal units. We found that, for different administrative units (census section, neighborhood, and district), the deprivation index obtained different results and correlations with CVD prevalence. Also, we found that the components of the index varied when building the index at different geographical units. At the census section and neighborhood level, the index indicators were related to occupation/labor market and education domains. However, at the district level, two indicators related to immigration were part of the first component of PCA. This could indicate immigration is a social phenomenon not limited locally in a city as densely populated as Madrid, but it extends through a wider territory. Thus, it became a statistically significant component for socioeconomic deprivation only when it was incorporated in a larger observation unit, such as the district.

In the third study, we developed a multicomponent index that integrated four domains related to heart-healthy urban environments: food, physical activity, tobacco, and alcohol environments. Two models were performed to calculate the Heart Healthy Hoods index (HHH index): one with un-weighted environment domains, and another with specific weights for each domain. We performed two regression analysis (OLS and GWR), to examine the relationship between the HHH index and CVD prevalence. The weighted model showed a better fit in both regression analyses. The GWR had a better fit (based on the AIC) and showed a non-stationary relationship between the heart-healthy environment and CVD prevalence. In addition, the weighted model captured better the spatial variation in neighborhood environments; there was a considerable difference in the relationship HHH index-CVD between the urban core and the less populated outskirts areas of the city. Specifically, we found areas with inverse relationships showing that improve cardiovascular environment would contribute to reducing prevalence rates of CVD. However, we also identified areas with a positive relationship between the HHH index models and the prevalence rates of CVD.

SECTION IV.
DISCUSSION

Throughout this PhD dissertation, we have described the current concept of Health Geography within the published works under an ecological approach, focusing mainly on how the effects of neighborhood and context on cardiovascular health have been articulated. Much of the progress in this research field has been driven from Epidemiology using concepts and geographic techniques such as GIS. However, there are aspects of this epidemiological research that are still unresolved from a geography perspective, being the most general the insufficient rigor necessary to conceptualize the "place". Therefore, there is a demand for evidence about the inherent challenges in ecological studies that use spatially aggregated data. In this sense, the current PhD dissertation has considered the impact of geographic properties of data on health outcomes. Specifically, spatial dependence, MAUP, scale effect, and spatial heterogeneity have been evaluated by studying the relationship between urban contexts and CVD prevalence. Thus, it is possible to appreciate a double contribution of our research on urban health scientific body: 1) it considers the necessary geographic conceptualization in ecological studies and 2) it applies it to a global health problem such as CVD.

Furthermore, geographers should have a central role in bringing the idea that individual health is a product of interactions with multiple geographical context and scales which can harm and promote health. In this work, a novel research line has been presented by focusing on the intertwined nature of the built environment multidimensional characteristics and how they are associated with CVD. This multidimensional approach has been considered in previous research relating the built environment and obesity, as shown by studies based on obesogenic environments (Frank *et al.*, 2012; Guthman, 2013; Economos *et al.*, 2015). However, to date, we have not found studies that synergistically consider the characteristics of the built environment characteristics that might impact CVD by shaping the 4 main risk factors of CVD: diet, physical activity, tobacco and alcohol. We consider this approach is crucial to better understand the place's role in health promotion and the reduction of health inequities.

Based on our research results, living in a healthier built environment would imply a small change for each residents' CVD risk. Therefore, it would be understandable if a policymaker has doubts whether to make an investment to improve the built environment if such investment does not produce obvious individual changes. However, a moderate improvement in the whole population can have a profound effect on the general burden of the disease (Rose, 1985). This is the basis of the population strategy for disease prevention. The relationship between the HHH index and the CVD prevalence might appear to be weak, so an improvement in the cardiovascular environment would imply a small improvement in individual CVD risk. However, if this little change happens in all individuals, it would imply a great change at the population level.

Most of the work has been developed throughout Madrid City, one of the main European capitals. Currently, several reviews have summarized the available evidence on the relationship between the built environment and different health outcomes or health behaviors (Feng *et al.*, 2010; Giskes *et al.*, 2011; Lee and Maheswaran, 2011; Caspi *et al.*, 2012; Carroll-Scott *et al.*, 2013). However, most studies were conducted in North America and Australia, while the number of studies conducted in Europe is scarce. The attributes of the physical environment in Europe are likely to differ from the United States or Australian context. For example, European cities (Kasanko *et al.*, 2006) are characterized by a more compact (dense) structure compared to cities in North America or Australia. Consequently, the evidence generated in this work could be the basis for future studies evaluating the relationship between European urban contexts and health outcomes. In addition, all the methods proposed in this research have the potential to be replicable in other parts of the world, with other administrative hierarchies, as long as the data is collected at the appropriate level. Both the indices to characterize the urban environment, as well as the socioeconomic deprivation, can be replicated in other area units, such as municipalities or countries, always taking into account the effort and resourced needed for data collection.

In the next sections, a detailed discussion around the results of the three studies is presented, relating them to the evidence already published. Finally, the limitations of the research and its strengths are presented.

Characterization of the obesogenic environment

In the first study of this research, the obesogenic environment was characterized taking into account the interrelationships between food and physical activity environments. These interrelationships arise due to the correlation between the metrics of these two urban environments (Guthman, 2013; Townshend and Lake, 2016). To characterize the obesogenic environment, a synthetic index based on GIS was developed by integrating on-field measures of both environments. This perspective, whose results are increasingly abundant, was also used in the third study of this dissertation. It was considered that, if the food and physical activity environments are related and form the so-called obesogenic environments, the interrelation between the contextual characteristics of the urban environment affecting cardiovascular health and its risk factors could also be considered. Under this premise, in the third study a heart-healthy index was developed integrating built environment measures of food, physical activity, alcohol, and tobacco.

The use of composite indexes reduces collinearity and excessive adjustment conferring ease of interpretation and reduction of measurement errors (Feng *et al.*, 2010). In addition, the integration of different indicators in an index can bring up associations not previously found (Kelly-Schwartz *et al.*, 2004). In the case of the first study, a systematic observation was carried out on-field using validated audit tools that provided highly detailed spatial data on physical activity and food environments. This also ensured both variability and more statistical power. The use of a broad sampling strategy to maximize variation among environmental factors reduced the sample size needed to evaluate associations between built environments and obesity outcome (Frank *et al.*, 2012).

Previous studies have considered both food and physical activity environment to characterize the obesogenic environment, but have not obtained a composite value (Frank *et al.*, 2012; Wall *et al.*, 2012; Meyer *et al.*, 2015). Most of these studies used GIS to integrate the information from several sources, mainly from secondary databases. The KDE remains underused compared to proximity analysis or statistically defined areas (Charreire *et al.*, 2010; Buck *et al.*, 2011; Thornton *et al.*, 2012). However, the number of studies that use KDE to study the obesogenic environment has increased in the last years (Rundle *et al.*, 2009; Charreire *et al.*, 2010; Thornton *et al.*, 2012). KDE overcomes the limitations in the binary definitions of the analysis based on fixed geographical limits (for example, number of stores per census section). Smooth transitions through administrative boundaries better represent the reality of urban environments (Chaix *et al.*, 2009). In addition, the resulting KDE surface of both the obesogenic index and the heart-healthy index can be used as an independent variable in statistical models (Carlos *et al.*, 2010).

Finally, it is necessary mentioning the potential of replicability of this methodology. Having used a selected study area through "median neighborhood" methodology, it could be useful for future international comparisons of median neighborhoods in different environments (Franco, Bilal and Diez-Roux, 2015; Díez *et al.*, 2016).

Scale effect in the study of social phenomena

The results obtained in this second study add to recommendations found in the literature about the need to examine the scale effects when using aggregated measures and studying associations between urban context and health indicators (Rey *et al.*, 2009; Houston, 2014; Hanigan, Cochrane and Davey, 2017). Researchers must take into account the MAUP when changing the data aggregation unit or when changing the scale (Openshaw, 1983; Fotheringham and Wong, 1991). For example, Root (2012) found poverty showed a stronger effect on smaller scales and unemployment on larger scales in relation to the cleft palate (Root, 2012). These results can be explained as unemployment rate may influence a higher level of geography since high unemployment tends to depress the economy in much larger areas (for example, an entire city), while poverty is often localized in small clustering areas (Root, 2012). Another evidence of MAUP implications can be found in the work of Johnston et all. (2016). They concluded that segregation in Australia was greater in the macro and micro scales than in the intermediate scales and varies in intensity (Johnston *et al.*, 2016). At different spatial scales and/or levels of aggregation, different patterns and degrees of spatial autocorrelation are obtained that will impact on the results of any statistical analysis. The results of the second study follow the line established by these previous works, evidenced that social phenomena have multi-scale characteristics.

As a result of the statistical and geographical stability analysis for socioeconomic deprivation, it was observed that deprivation index did not remain stable through the three administrative units (census section, neighborhood, and district). These results were corroborated with the correlation analysis between socioeconomic deprivation and the CVD prevalence. At the three scales, the correlations were positive and statistically significant ($p < 0.01$), but the association was stronger when the geographical unit of analysis was higher (e.g. districts). In general, any more disadvantaged geographical unit had a greater correlation with the prevalence of CVD. An explanation could be that residents of these disadvantaged areas are suffering a double burden of risk due to deprivation amplification (Macintyre, 2007). People may be suffering from a double disadvantage, the individual socioeconomic and the lack of infrastructure in the neighborhood that allows and encourages a healthy life. This could translate into the development of unhealthy habits (smoking, an unhealthy diet or sedentary lifestyle) affecting the cardiovascular health of those who live there.

The size and definition of the spatial unit of analysis should vary according to the phenomenon under study and the processes by which the area effect on the health outcome is hypothesized (Diez-Roux *et al.*, 2001). This study showed is possible to use a bivariate autocorrelation analysis to explore how an index of socioeconomic deprivation correlates with the closest deprivation values on different scales. Along with visualization, this approach has allowed to characterize the relathionship between MAUP and the spatial scale. Obviously, the aim was not to solve the MAUP, but to emphasize the need to analyze the associated and underlying phenomena of the aggregated data in areal units. To do so, a replicable socioeconomic deprivation index was developed and its stability was studied ghrough the scales.

The phenomenon of socioeconomic deprivation was selected because it constitutes a universal problem, a challenge for public health policies and, in general, it is evaluated through spatially aggregated data. Socioeconomic deprivation is the driver of social inequalities in health (Burrows *et al.*, 2011; Laraia *et al.*, 2012; Weng *et al.*, 2017) since people living in disadvantaged areas present worse health indicators and tend to group in neighborhoods with worse social and economic conditions. (Macintyre, 2007; Borrell *et al.*, 2014). Therefore, knowing the implications of the spatial unit selection is of great interest for all studies focused on inequalities in health with a spatial focus. In addition, the results of this research can inform health policy decision makers. Taking into account the impact of the aggregation scale on the association between socio-economic disadvantage and health outcomes can be useful for health surveillance and evaluation of public health interventions.

Heart-Healthy environment and its spatial relationship with CVD prevalence

Following the research line of the first study, an integrative approach was used to characterize the CVD-related built environment of the entire city of Madrid. As already mentioned in this document, several studies have addressed the effects of a built environment single measure on health-related behaviors. Some examples are studies on food environment (Lytle and Sokol, 2017), alcohol (Popova *et al.*, 2009), tobacco (Lee *et al.*, 2015) or physical activity (McCormack and Shiell, 2011). The novelty of our contribution lies in the use of an integrated approach to combine several measures of these environments to characterize the cardiovascular environment of Madrid. The measures of these environments can act synergistically to influence health-related behaviors (Nelson *et al.*, 2006), which would also influence health outcomes. Furthermore, it is convenient to emphasize this work goes beyond previous studies since it includes clinical data on cardiovascular health instead of general mortality or self-reported health (Chen *et al.*, 2010; Chalkias *et al.*, 2013; Clary *et al.*, 2016; Feuillet *et al.*, 2016).

Although we did not find previous research on multicomponent indexes integrating a whole perspective of the built environment related to cardiovascular health, we did find studies that integrate measures of different domains of the built environment related to obesity (Tseng et al., 2014; Wall et al., 2012) (Tseng et al., 2014; Wall et al., 2012). For example, Tseng et al. (2014) proposed an obesity index (covering three domains: food, activity resources, and ease of use) in Australia to quantify neighborhood obesogenicity. In addition, they examined its association with body mass index (BMI) in neighborhoods with socioeconomic disadvantages. In line with the third study of this PhD dissertation, Tseng's work found spatial heterogeneity in the relationship between the index and health outcomes (non-stationary relationship), with generally positive associations for BMI in urban areas and inverse associations in rural areas. Another example is the work of Wall et all., (2012) that integrated measures of food environment, physical activity environment, and public transport along with socioeconomic characteristics and examined their associations with the weight of adolescents (BMI). Using a latent class spatial analysis, they created composite scores representing neighborhood characteristics closely correlated and potentially related to obesity (i.e., accessibility of food and recreation outside the home, community disadvantages, green space, traffic density, and accessibility to supermarkets). Despite using a different methodology, they identified groups with complex combinations of both positive and negative environmental influences on health outcomes (Wall et al., 2012). Therefore, the findings of this dissertation coincide with the evidence in other countries about the influence of urban environments on CVD risk factors and the heterogeneity of these relationships.

Previous studies (Havard *et al.*, 2008; Borrell *et al.*, 2014; Shortt *et al.*, 2015) led to consider socioeconomic deprivation as a potentially relevant factor for the analyses of this study. It is possible that environments with healthy cardiovascular resources can mitigate CVD prevalence, but not eliminate it completely (Reed, Niileksela and Kaplan, 2013; Clary *et al.*, 2016). In environments with a variety of options (healthy and unhealthy), socio-economic or personal influences can buffer the influence of exposure to the built environment (Lytle, 2009). The results of this study showed that the control of socioeconomic deprivation can mitigate the effect of the HHH index on the risk of CVD. Together with the erroneous specifications of the model, the non-stationarity of the relationship between the HHH index and the prevalence of CVD may also be due to intrinsic differences in the way in which individuals respond to the specific characteristics of their local environment (Brunsdon, Fotheringham and Charlton, 2002).

According to our knowledge, this is the first study examines a wide variety of neighborhood contextual characteristics that influence the cardiovascular health of Madrid citizens. But it is worth noting the international application that HHH index can have. Comprehensive urban health studies in Europe (Schneider and Gruber, 2013; Macdonald *et al.*, 2018), U.S.A (Auchincloss *et al.*, 2008; Moore *et al.*, 2008; Brenner *et al.*, 2015) and Australia (Koohsari *et al.*, 2016; Feng *et al.*, 2018) could implement this proposal to characterize urban environments in relation to cardiovascular health. Considering the balance or combination of resources can be theoretically and empirically important to understand the influence of the built environment on health-related behaviors. According to the theories of behavioral economics (Hursh and Roma, 2013; Reed, Niileksela and Kaplan, 2013), human behavior is the result of the distribution and availability of a diverse and synergistic set of resources. However, most research examines the discrete characteristics of exposure to the urban environment (Caspi *et al.*, 2012; Malambo *et al.*, 2016). Specific exposures can be more easily addressed from public health policies, but it can be a more precise reflection considering human behavior within a system with multiple contextual characteristics.

1. LIMITATIONS

This PhD research implies a series of limitations that it has not been possible to tackle throughout this dissertation, but that must be highlighted and recognized. However, certain limitations have been rectified as the different research objectives were reached. For example, the first study of this dissertation required the collection of primary data through systematic observation, a process that requires resources and time. However, in the third study involving a similar approach, these costs were drastically reduced by using secondary databases with spatial information and new remote devices to collect geocoded primary data (Matthews, Moudon and Daniel, 2009). On the other hand, when GIS measurements of the urban environment are derived from a secondary database, there may be measurement errors as this data was collected for other purposes (Liese *et al.*, 2010, 2013). However, there is no reason to believe that inaccuracies affect one area more than another, and therefore, the probability of measurement error may not be differential in the study area. In the first study, the relative importance of physical activity and food environments was not considered, treating both environments with the same weight by means of an unweighted local average. This limitation was addressed in the third study of this research by developing two models of the heart-healthy index. Following the methodology of the first study, one of the models considered that all the evaluated environments (food, physical activity, alcohol, and tobacco) had the same weight. On the other hand, the second model treated the relative importance of each environment by the percentage of CVD deaths attributable to the four main risk factors (unhealthy diet, lack of physical activity, alcohol consumption and smoking). These percentages were used as weights when developing this second model.

A limitation that could not be remedied originates in the measurement of urban environments by selecting a very specific set of indicators. Therefore, other characteristics of the urban context that could influence residents' health behaviors may be lacking (Hillsdon *et al.*, 2006; Jeffery *et al.*, 2006). For example, the off-premise sale of food and alcoholic beverages was considered, but on-premise (for example restaurants, cafeterias, bars) or nightlife (such as discos or nightclubs) were not considered. In the case of tobacco environment, all tobacco stores were taken into account, but tobacco vending machines were not. This may be underestimating accessibility and availability to the contextual determinants of CVD, that is, exposure to food, alcohol and tobacco environments.

Another notable limitation related to data source selected is that to calculate socioeconomic deprivation was used National Census, which is collected every 10 years. This implies a large time lag that makes it difficult to update the data for longitudinal or cross-sectional studies throughout the decade. In addition, it is possible that the indicators selected to calculate socioeconomic deprivation did not provide enough information to reflect all aspects of this social phenomenon. Furthermore, the correlation analysis performed between socioeconomic deprivation and the prevalence of CVD does not allow to infer the causality between the variables. However, our results allowed to corroborate the importance of spatial scaling, and how choices of scale impact on the relationship between socioeconomic deprivation and CVD.

In the first and third studies of this dissertation, the census section was used to aggregate measures of the urban built environment. This might seem an inconsistency since throughout this dissertation the MAUP (Openshaw, 1983) has been evidenced as a critical issue when using aggregated data (Barnes *et al.*, 2016). However, this study provides two types of results for the characterization of the built environment, one continuous and one delimited by census sections. The continuous result is based on a mathematical surface that overcomes the limitations of administrative limits, so it better presents the reality of urban environments (Chaix *et al.*, 2009). The second result limited by the census section of Madrid is necessary since this area is the smallest unit in which can be disposed of sociodemographic, economic, and health statistics.

Finally, the study area of this work is limited to the municipality of Madrid, so the metropolitan area of the city has not been included. This has generated the so-called "edge effect" and is a limitation that arises when studies are performed in finite regions. Urban environments, prevalence of CVD and socioeconomic deprivation extend beyond the geographical boundaries of Madrid municipality. However, we do not consider the surroundings of the municipality since the data is provided by the city council of Madrid and is not available for other municipalities that are part of Madrid's metropolitan region.

2. STRENGTHS

Despite the limitations, this doctoral dissertation has remarkable strengths that have been discussed throughout this document. One of them is that the need to develop a tool to help policy makers in their choice of health interventions and the establishment of priorities has been overcome. Epidemiological research has used a variety of highly sophisticated analysis to measure the correlation and causality between risk factors and health outcomes. Their findings are key to public health, but offer limited guidance to policymakers in the choice, scope, and area of intervention. Therefore, there is a demand for works that, using a spatial approach, can guide political decision-makers to establish priorities and strategies when developing the most appropriate local intervention.

The urban context indices developed in this dissertation showed the spatial distribution of the main contextual risk factors of CVD. This distribution is depicted by a continuous surface allowing policymakers to properly choose between different approaches that are not affected by the aggregation of spatial data in administrative areal units. Both indices and its associated cartography are an effective analytical tool to support public health policy decisions. In this sense, another strength is the use and comparison of two methodological approaches, one global and one local. Methodologically, local analysis (GWR), as a heuristic approach (Wheeler and Paez, 2010), provides more useful information about the relationships between CVD and the heart-healthy environment than global analysis (OLS). It is considered useful and necessary to complete the results of a global analysis with a local analysis taking into account the spatial dependence and non-stationary nature of the data. The GWR analysis improved the ability to explain local CVD rates. Unlike the global OLS model, the GWR findings allowed us to locally identify what kind of relationship exists between the heart-healthy environment and CVD and where this relationship is strongest. This local diagnosis can help policymakers to plan community interventions where there are greater needs and the impact will be higher. Furthermore, thanks to the evidence generated in the second study, the results of this research can help policy decisions about the scale in which the intervention should focus.

Another strong point of this research is to have health data obtained from the universal Primary Health Care electronic records, which covers almost all citizens, and the risk of unreported events is low. Thus, we used data from 1,446,994 residents between 45 and 75 years old living in the city of Madrid, and therefore, age-adjusted prevalence of CVD at the census section level.

Once exposed the main strengths and limitations of the research, we will conclude this section highlighting the development of novel geospatial methodologies to study the urban contextual determinants of CVD and its spatial distribution pattern for the entire city of Madrid. Under a rigorous geographical perspective, this research will help advance the field of epidemiology and public health policy.

SECTION V.
GENERAL CONCLUSION

In this section, a discussion on the degree of achievement of the proposed objectives is presented. The main objective of this PhD dissertation was to develop a GIS-based methodological proposal to characterize urban contextual determinants of CVD. To reach this objective and being able to verify the hypothesis, it was necessary to reach the proposed secondary objectives.

SO1. To design a GIS-based multicomponent index to integrate information from food and physical activity environments to characterize the obesogenic environment of an urban area.

This secondary objective was reached in study 1 of this dissertation that culminated with the publication in International Journal of Health Geography. In order to understand the obesogenic environment, it is necessary to consider the inter-relations between food and physical activity environments. For this, we proposed a synthetic index providing a feasible way to integrate different measures of healthy urban environments in terms of food and physical activity. This multicomponent method opens up new ways to capture the interrelationships between physical activity and healthy food availability urban environment domains that did not emerge when they were studied in an isolated way.

SO2. To build a socioeconomic deprivation index and to study its statistical and geographic stability through three different spatial observation units (census section, neighborhoods and districts).

SO3. To study the scale effect on the relationship between socioeconomic deprivation and prevalence of CVD.

Both objectives were achieved in studio 2 which was published in the journal Applied Geography . A deprivation index was developed allowing the detection of areas of large cities with an unfavorable socioeconomic situation. In addition, were evidenced the changes that occur in the relationship between deprivation index and CVD prevalence when they are studied at different spatial aggregation units. We concluded deprivation index can contribute to the study of social inequalities in health in Madrid, and be an instrument of great utility for health planning. However, the scale effect influences the spatial pattern and the relationship between deprivation and CVD. The correlations between deprivation and CVD at the three study scales were positive and significant, being greater as the size of the spatial unit increased.

Our recommendation is that descriptive analysis and notification of results, be made at the neighborhood or district level, since these spatial units best capture the wide distribution of the disease and deprivation. For explanatory or inferential studies, it may be necessary to deepen to census section level to delineate those areas for a specific approach, both to understand the possible exposures and to guide the resources for the intervention. And finally, if there is uncertainty about the best scale to represent the pathways acting between specific social determinants and health outcomes, several geographic scales, and explicit hypotheses shall be explored.

SO4. To develop and implement a multicomponent index, the Heart Healthy Hoods index (HHH index), integrating characteristics of healthy urban environments for the heart (food, physical activity, tobacco, and alcohol) using GIS tools.

SO5. To examine the association of the HHH index with the prevalence of CVD at the area level.

These secondary objectives were developed in study 3 which published in the journal Health & Place. We developed the Heart Healthy Hoods Index (HHH index), a multi-component method to characterize, in an exhaustive way, heart-healthy urban environments. We concluded, in relation to SO5, the relationship between the Heart Healthy Hoods Index and CVD prevalence was non-stationary but varied along the territory of Madrid both in direction and strength. Examining the association between HHH index and CVD prevalence, the local regression analysis (GWR) provides more useful information about the relationship between CVD and the heart-healthy environment than the global regression analysis, since allowing to identify locally what type of relationship exists and where is stronger. The Heart Healthy Hoods Index could be a relevant tool to identify urban environments susceptible to cardiovascular preventive interventions.

In summary, this PhD dissertation has deepened from a robust geospatial approach in the development of assessment tools for the contextual determinants of health-related environments. In addition, we analyze their relationship with the prevalence of CVD from 1.5 million citizens. From an integrative perspective, we have been developing and enriching different methodologies to characterize urban environments (physical and social) and demonstrating their applicability testing the indices by studying their spatial relationship with CVD prevalence highlighting variations of this association through space and scale. The indices and the cartography associated developed in this PhD dissertation provide effective analytical tools to support public health policy decisions in that identifies the geographical areas where preventive and health promotion interventions for heart diseases and its risk factors associated might be prioritized.

Finally, we remind the hypothesis raised:

GIS-based multicomponent indexes may capture the intertwined nature of the contextual determinants of CVD in a more complete fashion than single indexes.

As a consequence of the achievement of specific and general objectives proposed in this PhD research, we conclude that the hypothesis has been accepted.

REFERENCES

Anselin, L. (1988) *Spatial econometrics: methods and models*. Springer Science & Business Media. doi: 10.1007/978-94-015-7799-1.

Anthamatten, P. and Hazen, H. (2011) *An introduction to the geography of health*. 1st edn. London: Routledge. doi: 10.4324/9780203877463.

Aránguez, E. *et al.* (2005) 'Sistemas de información geográfica y salud pública en la Comunidad de Madrid', *Departamento de Geografía. Serie Geográfica*. Alcalá de Henares: Universidad de Alcalá, 12, pp. 137–146.

Arnott, B. *et al.* (2014) 'Efficacy of behavioural interventions for transport behaviour change: Systematic review, meta-analysis and intervention coding', *International Journal of Behavioral Nutrition and Physical Activity*. doi: 10.1186/s12966-014-0133-9.

Auchincloss, A. H. *et al.* (2008) 'Neighborhood resources for physical activity and healthy foods and their association with insulin resistance.', *Epidemiology (Cambridge, Mass.)*. doi: 10.1097/EDE.0b013e31815c480.

Auchincloss, A. H. *et al.* (2012) *A Review of Spatial Methods in Epidemiology, 2000–2010*, *SSRN*. doi: 10.1146/annurev-publhealth-031811-124655.

Auchincloss, A. H. *et al.* (2013) 'Neighborhood health-promoting resources and obesity risk (the multi-ethnic study of atherosclerosis).', *Obesity (Silver Spring, Md.)*. United States, 21(3), pp. 621–628. doi: 10.1002/oby.20255.

Barcellos, C. and Buzai, G. D. (2006) 'La dimensión espacial de las desigualdades sociales en salud: aspectos de su evolución conceptual y metodológica', *Departamento de Ciencias Sociales. Universidad Nacional de Luján: Anuario de la División Geografía*, pp. 275–292. doi: http://dx.doi.org/10.19137/huellas-2018-2205.

Barnes, T. L. *et al.* (2016) 'Scale effects in food environment research: Implications from assessing socioeconomic dimensions of supermarket accessibility in an eight-county region of South Carolina', *Applied Geography*. doi: 10.1016/j.apgeog.2016.01.004.

Barton, P. *et al.* (2011) 'Effectiveness and cost effectiveness of cardiovascular disease prevention in whole populations: Modelling study', *BMJ (Online)*. doi: 10.1136/bmj.d4044.

Berkman, L., Kawachi, I. and Glymour, M. (2014) *Social Epydemiology*. 2nd edn. Oxford, England: Oxford University Press.

Bethlehem, J. R. *et al.* (2014) 'The SPOTLIGHT virtual audit tool: A valid and reliable tool to assess obesogenic characteristics of the built environment', *International Journal of Health Geographics*. doi: 10.1186/1476-072X-13-52.

Bilal, U. *et al.* (2014) 'Validation of a method for reconstructing historical rates of smoking prevalence', *American Journal of Epidemiology*. doi: 10.1093/aje/kwt224.

Bilal, U., Beltrán, P., *et al.* (2016) 'Gender equality and smoking: A theory-driven approach to smoking gender differences in Spain', *Tobacco Control*. doi: 10.1136/tobaccocontrol-2014-051892.

Bilal, U., Díez, J., *et al.* (2016) 'Population cardiovascular health and urban environments: The Heart Healthy Hoods exploratory study in Madrid, Spain', *BMC Medical Research Methodology*. BMC Medical Research Methodology, 16(1), pp. 1–12. doi: 10.1186/s12874-016-0213-4.

Bilal, U. *et al.* (2018) 'Neighborhood social and economic change and retail food environment change in Madrid (Spain): The heart healthy hoods study', *Health and Place*, 51. doi: 10.1016/j.healthplace.2018.03.004.

Bivoltsis, A. *et al.* (2018) 'Food environments and dietary intakes among adults: does the type of spatial exposure measurement matter? A systematic review.', *International journal of health geographics*. England, 17(1), p. 19. doi: 10.1186/s12942-018-0139-7.

Borrell, C. *et al.* (2014) 'Socioeconomic inequalities in mortality in 16 European cities.', *Scandinavian journal of public health*. doi: 10.1177/1403494814522556.

Bosque, J. (1992) *Sistemas de Información Geográfica*. RIALP.

Brenner, A. B. *et al.* (2015) 'Associations of alcohol availability and neighborhood socioeconomic characteristics with drinking: Cross-sectional results from the Multi-Ethnic Study of Atherosclerosis (MESA)', *Substance Use and Misuse*. doi: 10.3109/10826084.2015.1027927.

Brown *et al.* (2018) *Health geographies: A critical introduction*. John Wiley & Sons. Available at: https://population.un.org/wup/Download/.

Brown, A. F. *et al.* (2011) 'Neighborhood disadvantage and ischemic stroke: The Cardiovascular Health Study (CHS)', *Stroke*. doi: 10.1161/STROKEAHA.111.622134.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. (2002) 'Geographically weighted summary statistics - a framework for localised exploratory data analysis', *Computers, Environment and Urban Systems*. doi: 10.1016/S0198-9715(01)00009-6.

Buck, C. *et al.* (2011) 'Development and application of a moveability index to quantify possibilities for physical activity in the built environment of children', *Health and Place*. doi: 10.1016/j.healthplace.2011.08.011.

Burrows, S. *et al.* (2011) 'Influence of social and material individual and area deprivation on suicide mortality among 2.7 million Canadians: a prospective study.', *BMC public health*. England, 11, p. 577. doi: 10.1186/1471-2458-11-577.

Caballero, B., Finglas, P. M. and Toldrá, F. (2016) *Encyclopedia of Food and Health*, *Volume 1: A-Che*.

doi: 10.1016/B978-0-12-384947-2.00329-9.

Carlos, H. A. *et al.* (2010) 'Density estimation and adaptive bandwidths: A primer for public health practitioners', *International Journal of Health Geographics*. doi: 10.1186/1476-072X-9-39.

Carroll-Scott, A. *et al.* (2013) 'Disentangling neighborhood contextual associations with child body mass index, diet, and physical activity: The role of built, socioeconomic, and social environments', *Social Science & Medicine*. Pergamon, 95, pp. 106–114. doi: 10.1016/J.SOCSCIMED.2013.04.003.

Caspi, C. E. *et al.* (2012) 'The local food environment and diet: a systematic review.', *Health & place*. doi: 10.1016/j.healthplace.2012.05.006.

Chaix, B. *et al.* (2006) 'Disentangling contextual effects on cause-specific mortality in a longitudinal 23-year follow-up study: Impact of population density or socioeconomic environment?', *International Journal of Epidemiology*. doi: 10.1093/ije/dyl009.

Chaix, B. *et al.* (2009) 'Neighbourhoods in eco-epidemiologic research: Delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur', *Social Science and Medicine*. doi: 10.1016/j.socscimed.2009.07.018.

Chaix, B., Merlo, J. and Chauvin, P. (2005) 'Comparison of a spatial approach with the multilevel approach for investigating place effects on health: The example of healthcare utilisation in France', *Journal of Epidemiology and Community Health*. doi: 10.1136/jech.2004.025478.

Chaix, B., Rosvall, M. and Merlo, J. (2007) 'Neighborhood socioeconomic deprivation and residential instability: Effects on incidence of ischemic heart disease and survival after myocardial infarction', *Epidemiology*. doi: 10.1097/01.ede.0000249573.22856.9a.

Chalkias, C. *et al.* (2013) 'Geographical heterogeneity of the relationship between childhood obesity and socio-environmental status: Empirical evidence from Athens, Greece', *Applied Geography*. doi: 10.1016/j.apgeog.2012.10.007.

Charreire, H. *et al.* (2010) 'Measuring the food environment using geographical information systems: A methodological review', *Public Health Nutrition*. doi: 10.1017/S1368980010000753.

Chen, V. Y. J. *et al.* (2010) 'Examining non-stationary effects of social determinants on cardiovascular mortality after cold surges in Taiwan', *Science of the Total Environment*. doi: 10.1016/j.scitotenv.2009.11.044.

Chen, X. and Kwan, M. P. (2015) 'Contextual uncertainties, human mobility, and perceived food environment: The uncertain geographic context problem in food access research', *American Journal of Public Health*. doi: 10.2105/AJPH.2015.302792.

Chetty, R. *et al.* (2016) 'The association between income and life expectancy in the United States, 2001-2014', *JAMA - Journal of the American Medical Association*. doi: 10.1001/jama.2016.4226.

Christian, W. J. (2012) 'Using geospatial technologies to explore activity-based retail food environments', *Spatial and Spatio-temporal Epidemiology*. doi: 10.1016/j.sste.2012.09.001.

Clary, C. *et al.* (2016) 'The local food environment and fruit and vegetable intake: A geographically weighted regression approach in the ORiEL study', *American Journal of Epidemiology*. doi: 10.1093/aje/kww073.

Cobb, L. K. *et al.* (2015) 'The relationship of the local food environment with obesity: A systematic review of methods, study quality, and results.', *Obesity (Silver Spring, Md.)*. United States, 23(7), pp. 1331–1344. doi: 10.1002/oby.21118.

Coutts, C. and Hahn, M. (2015) 'Green infrastructure, ecosystem services, and human health', *International Journal of Environmental Research and Public Health*. doi: 10.3390/ijerph120809768.

Cummins, S. and Macintyre, S. (2006) 'Food environments and obesity - Neighbourhood or nation?', *International Journal of Epidemiology*, pp. 100–104. doi: 10.1093/ije/dyi276.

Curtis, S. (2004) *Health and inequality: Geographical perspectives*. London: Sage.

D'Angelo, H. *et al.* (2015) 'Examination of community and consumer nutrition, tobacco and physical activity environments at food and tobacco retail stores in three diverse North Carolina communities', *Preventive Medicine Reports*. doi: 10.1016/j.pmedr.2015.08.018.

DeWeese, R. S. *et al.* (2018) 'Patterns of food and physical activity environments related to children's food and activity behaviors: A latent class analysis', *Health and Place*. doi: 10.1016/j.healthplace.2017.11.002.

Diez-Roux (2001) 'Investigating neighborhood and area effects on health', *American Journal of Public Health*. doi: 10.2105/AJPH.91.11.1783.

Diez-Roux *et al.* (2001) 'Neighborhood of Residence and Incidence of Coronary Heart Disease', *New England Journal of Medicine*. doi: 10.1056/NEJM200107123450205.

Diez-Roux, A. V. (2003) 'Residential Environments and Cardiovascular Risk', in *Journal of Urban Health*. doi: 10.1093/jurban/jtg065.

Diez-Roux, A. V. and Mair, C. (2010) 'Neighborhoods and health', *Annals of the New York Academy of Sciences*. doi: 10.1111/j.1749-6632.2009.05333.x.

Diez-Roux, A. V (1998) 'Bringing context back into epidemiology: variables and fallacies in multilevel analysis.', *American Journal of Public Health*. American Public Health Association, 88(2), pp. 216–222. doi: 10.2105/AJPH.88.2.216.

Díez, J. *et al.* (2016) 'Understanding differences in the local food environment across countries: A case study in Madrid (Spain) and Baltimore (USA)', *Preventive Medicine*, 89. doi: 10.1016/j.ypmed.2016.06.013.

Domínguez-Berjón, M. F. *et al.* (2010) 'Census tract socioeconomic and physical environment and cardiovascular mortality in the Region of Madrid (Spain)', *Journal of Epidemiology and Community Health*. doi: 10.1136/jech.2008.085621.

Dummer, T. J. B. (2008) 'Health geography: Supporting public health policy and planning', *CMAJ*. doi: 10.1503/cmaj.071783.

Duncan, D. T. and Kawachi, I. (2018) *Neighborhoods and health*. 2nd edn. New York: Oxford University Press.

Eagleson, S. *et al.* (2005) 'Spatial algorithm for detecting disease outbreaks in Australia', *Departamento de Geografía. Serie Geográfica*. Alcalá de Herares: Universidad de Alcalá, 12, pp. 39–55.

Eagleson, S., Escobar, F. and Williamson, I. (2002) 'Hierarchical spatial reasoning theory and GIS technology applied to the automated delineation of administrative boundaries', *Computers, Environment and Urban Systems*, 26(2–3), pp. 185–200. doi: 10.1016/S0198-9715(01)00040-0.

Eagleson, S., Escobar, F. and Williamson, I. (2003) 'Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems', *International Journal of Geographical Information Science*. doi: 10.1080/713811748.

Eckel, R. H. *et al.* (2014) '2013 AHA/ACC Lifestyle management to reduce cardiovascular risk', *Journal of the American College of Cardiology*, 63(25 PART B), pp. 2960–2984. doi: 10.1016/j.jacc.2013.11.003.

Economos, C. D. *et al.* (2015) 'Food and physical activity environments: An energy balance approach for research and practice', *American Journal of Preventive Medicine*. doi: 10.1016/j.amepre.2014.12.007.

Engel, S. G. *et al.* (2016) 'Ecological Momentary Assessment in Eating Disorder and Obesity Research: a Review of the Recent Literature', *Current Psychiatry Reports*. doi: 10.1007/s11920-016-0672-7.

Escobar, F. J. *et al.* (2004) 'Caracterización de los riesgos ambientales para la salud: puesta a punto de un SIG en dos distritos del sureste de Madrid', *Departamento de Geografía. Serie Geográfica*. Alcalá de Henares: Universidad de Alcalá, 12, pp. 11–37.

Estadísticas Madrid (2017) *Esperanza de vida por Distrito y Sexo según Edad*, *Ayuntamiento de Madrid*. Available at: http://www-2.munimadrid.es/CSE6/control/seleccionDatos?numSerie=3090000010 (Accessed: 31 January 2019).

Ezzati, M. and Riboli, E. (2012) 'Can noncommunicable diseases be prevented? Lessons from studies of populations and individuals', *Science*. doi: 10.1126/science.1227001.

Feng, J. *et al.* (2010) 'The built environment and obesity: A systematic review of the epidemiologic evidence', *Health and Place*. doi: 10.1016/j.healthplace.2009.09.008.

Feng, X. *et al.* (2018) 'Modest ratios of fast food outlets to supermarkets and green grocers are associated with higher body mass index: Longitudinal analysis of a sample of 15,229 Australians aged 45 years and older in the Australian National Liveability Study', *Health and Place*. doi: 10.1016/j.healthplace.2017.10.004.

Feuillet, T. *et al.* (2016) 'Built environment in local relation with walking: Why here and not there?', *Journal of Transport and Health*. doi: 10.1016/j.jth.2015.12.004.

Flowerdew, R., Manley, D. J. and Sabel, C. E. (2008) 'Neighbourhood effects on health: Does it matter where you draw the boundaries?', *Social Science and Medicine*. doi: 10.1016/j.socscimed.2007.11.042.

Foreman, K. J. *et al.* (2018) 'Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016-40 for 195 countries and territories.', *Lancet (London, England)*. doi: 10.1016/S0140-6736(18)31694-5.

Foster, S. *et al.* (2017) 'Liquor landscapes: Does access to alcohol outlets influence alcohol consumption in young adults?', *Health and Place*. doi: 10.1016/j.healthplace.2017.02.008.

Fotheringham, A. S. and Wong, D. W. S. (1991) 'The Modifiable Areal Unit Problem in Multivariate Statistical Analysis', *Environment and Planning A*. doi: 10.1068/a231025.

Franco, M. *et al.* (2009) 'Availability of healthy foods and dietary patterns: The Multi-Ethnic study of atherosclerosis', *American Journal of Clinical Nutrition*. doi: 10.3945/ajcn.2008.26434.

Franco, M. *et al.* (2011) 'Challenges and opportunities for cardiovascular disease prevention', *American Journal of Medicine*. doi: 10.1016/j.amjmed.2010.08.015.

Franco, M., Bilal, U. and Diez-Roux, A. V. (2015) 'Preventing non-communicable diseases through structural changes in urban environments', *Journal of Epidemiology and Community Health*. doi: 10.1136/jech-2014-203865.

Franco, R. Z. *et al.* (2016) 'Popular Nutrition-Related Mobile Apps: A Feature Assessment', *JMIR mHealth and uHealth*. doi: 10.2196/mhealth.5846.

Frank, L. D. *et al.* (2012) 'Objective assessment of obesogenic environments in youth: Geographic information system methods and spatial findings from the neighborhood impact on kids study', *American Journal of Preventive Medicine*. doi: 10.1016/j.amepre.2012.02.006.

Gebreab, S. Y. (2018) 'Statistical Methods in Spatial Epidemiology', in Duncan, D. T. and Kawachi, I. (eds) *Neighborhoods and Health*. 2nd edn. Oxford University Press, p. 371.

Ghosh, D. and Guha, R. (2013) 'What are we "tweeting" about obesity? Mapping tweets with topic modeling and Geographic Information System', *Cartography and Geographic Information Science*. doi:

10.1080/15230406.2013.776210.

Giskes, K. *et al.* (2011) 'A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments?', *Obesity Reviews*. John Wiley & Sons, Ltd (10.1111), 12(5), pp. e95–e106. doi: 10.1111/j.1467-789X.2010.00769.x.

Glass, T. A. and McAtee, M. J. (2006) 'Behavioral science at the crossroads in public health: Extending horizons, envisioning the future', *Social Science and Medicine*. doi: 10.1016/j.socscimed.2005.08.044.

Golden, S. D. *et al.* (2015) 'Upending the Social Ecological Model to Guide Health Promotion Efforts Toward Policy and Environmental Change', *Health Education & Behavior*. SAGE Publications Inc, 42(1_suppl), p. 8S–14S. doi: 10.1177/1090198115575098.

Gómez-Gracia, E. *et al.* (2013) 'Primary prevention of cardiovascular disease with a mediterranean diet', *Zeitschrift fur Gefassmedizin*, 10(2), p. 28. doi: 10.1056/NEJMoa1200303.

Goodchild, M. F. (1987) 'Geographic information systems and cartography', *Hospital Topics*. doi: 10.1080/00185868.1987.10543805.

Goodchild, M. F. (2004) 'The validity and usefulness of laws in geographic information science and geography', *Annals of the Association of American Geographers*. doi: 10.1111/j.1467-8306.2004.09402008.x.

Goodchild, M. F. (2009) 'What Problem? Spatial autocorrelation and geographic information science', *Geographical Analysis*. doi: 10.1111/j.1538-4632.2009.00769.x.

Goverment Office for Science (2003) 'Tackling Obesities : Future Choices – Obesogenic Environments – Evidence Review', *Foresight*. doi: 10.1108/14636680310471299.

Green, M. A. *et al.* (2018) 'Developing an openly accessible multi-dimensional small area index of "Access to Healthy Assets and Hazards" for Great Britain, 2016', *Health & Place*. Pergamon, 54, pp. 11–19. doi: 10.1016/J.HEALTHPLACE.2018.08.019.

Griffith, D. and Chun, Y. (2014) 'Spatial Autocorrelation and Spatial Filtering', in Fischer, M. M. and Nijkamp, P. (eds) *Handbook of Regional Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1477–1507. doi: 10.1007/978-3-642-23430-9_72.

Gustafson, A. *et al.* (2013) 'Food venue choice, consumer food environment, but not food venue availability within daily travel patterns are associated with dietary intake among adults, Lexington Kentucky 2011', *Nutrition Journal*. doi: 10.1186/1475-2891-12-17.

Guthman, J. (2013) 'Too much food and too little sidewalk? problematizing the obesogenic environment thesis', *Environment and Planning A*. doi: 10.1068/a45130.

Hägerstrand, T. (1985) 'Time-geography: focus on the corporeality of man, society, and environment', *The science and praxis of complexity*, pp. 193–216.

Hägerstrand, T. (1989) 'Reflections on "what about people in regional science?"', *Papers of the Regional Science Association*. doi: 10.1007/BF01954291.

Hanigan, I. C., Cochrane, T. and Davey, R. (2017) 'Impact of scale of aggregation on associations of cardiovascular hospitalization and socio-economic disadvantage', *PLoS ONE*. doi: 10.1371/journal.pone.0188161.

Havard, S. *et al.* (2008) 'A small-area index of socioeconomic deprivation to capture health inequalities in France', *Social Science and Medicine*. doi: 10.1016/j.socscimed.2008.09.031.

Hawelka, B. *et al.* (2014) 'Geo-located Twitter as proxy for global mobility patterns', *Cartography and Geographic Information Science*. doi: 10.1080/15230406.2014.890072.

Hillsdon, M. *et al.* (2006) 'The relationship between access and quality of urban green space with population physical activity', *Public Health*. doi: 10.1016/j.puhe.2006.10.007.

Hirsch, J. A. *et al.* (2014) 'Using MapMyFitness to Place Physical Activity into Neighborhood Context', *Frontiers in Public Health*. doi: 10.3389/fpubh.2014.00019.

Hiscock, R. *et al.* (2012) 'Socioeconomic status and smoking: a review', *Annals of the New York Academy of Sciences*. doi: 10.1111/j.1749-6632.2011.06202.x.

Hobbs, M. *et al.* (2018) 'Neighbourhood typologies and associations with body mass index and obesity: A cross-sectional study', *Preventive Medicine*. doi: 10.1016/j.ypmed.2017.11.024.

Horner, M. W. and Wood, B. S. (2014) 'Capturing individuals' food environments using flexible space-time accessibility measures', *Applied Geography*. doi: 10.1016/j.apgeog.2014.03.007.

Houston, D. (2014) 'Implications of the modifiable areal unit problem for assessing built environment correlates of moderate and vigorous physical activity', *Applied Geography*. doi: 10.1016/j.apgeog.2014.02.008.

Hursh, S. R. and Roma, P. G. (2013) 'Behavioral economics and empirical public policy', *Journal of the Experimental Analysis of Behavior*. doi: 10.1002/jeab.7.

Jeffery, R. W. *et al.* (2006) 'Are fast food restaurants an environmental risk factor for obesity?', *The international journal of behavioral nutrition and physical activity*. doi: 10.1186/1479-5868-3-2.

Jerrett, M., Gale, S. and Kontgis, C. (2010) 'Spatial modeling in environmental and public health research', *International Journal of Environmental Research and Public Health*. doi: 10.3390/ijerph7041302.

Jiang, B. (2015) 'Geospatial analysis requires a different way of thinking: the problem of spatial heterogeneity', *GeoJournal*. doi: 10.1007/s10708-014-9537-y.

Johnston, R. *et al.* (2016) 'The scale of segregation: ancestral groups in Sydney, 2011', *Urban*

Geography. doi: 10.1080/02723638.2016.1139867.

Jones, L. *et al.* (2015) 'Relationship between alcohol-attributable disease and socioeconomic status, and the role of alcohol consumption in this relationship: A systematic review and meta-analysis', *BMC Public Health*. doi: 10.1186/s12889-015-1720-7.

Jurdak, R. *et al.* (2015) 'Understanding human mobility from Twitter', *PLoS ONE*. doi: 10.1371/journal.pone.0131469.

Kasanko, M. *et al.* (2006) 'Are European cities becoming dispersed?: A comparative analysis of 15 European urban areas', *Landscape and Urban Planning*. Elsevier, 77(1–2), pp. 111–130. doi: 10.1016/J.LANDURBPLAN.2005.02.003.

Kelly-Schwartz, A. C. *et al.* (2004) 'Is sprawl unhealthy? A multilevel analysis of the relationship of metropolitan sprawl to the health of individuals', *Journal of Planning Education and Research*. doi: 10.1177/0739456X04267713.

Kim, A. E. *et al.* (2015) 'Using twitter data to gain insights into E-cigarette marketing and locations of use: An infoveillance study', *Journal of Medical Internet Research*. doi: 10.2196/jmir.4466.

Kiraly, C. *et al.* (2017) 'Applying Ecological Frameworks in Obesity Intervention Studies in Hispanic/Latino Youth: A Systematic Review', *Hispanic Health Care International*. SAGE Publications Inc, 15(3), pp. 130–142. doi: 10.1177/1540415317731069.

Kirby, R. S., Delmelle, E. and Eberth, J. M. (2017) 'Advances in spatial epidemiology and geographic information systems', *Annals of Epidemiology*. doi: 10.1016/j.annepidem.2016.12.001.

Kirchner, T. R. *et al.* (2013) 'Geospatial Exposure to Point-of-Sale Tobacco: Real-Time Craving and Smoking-Cessation Outcomes', *American Journal of Preventive Medicine*. Elsevier, 45(4), pp. 379–385. doi: 10.1016/J.AMEPRE.2013.05.016.

Koohsari, M. J. *et al.* (2016) 'Street network measures and adults' walking for transport: Application of space syntax', *Health and Place*. doi: 10.1016/j.healthplace.2015.12.009.

Kreatsoulas, C. and Anand, S. S. (2010) 'The impact of social determinants on cardiovascular disease', in *Canadian Journal of Cardiology*. doi: 10.1016/S0828-282X(10)71075-8.

Krieger, N. (2011) *Epidemiology and the People's Health: Theory and Context*, *Epidemiology and the People's Health: Theory and Context*. doi: 10.1093/acprof:oso/9780195383874.001.0001.

Kwan, M. P. (2012) 'The Uncertain Geographic Context Problem', *Annals of the Association of American Geographers*. doi: 10.1080/00045608.2012.687349.

Laffly, D. and Handschumacher, P. (2005) 'Spatial Expressions of Sleeping Sickness in Ivory Coast and Guinea Conakry remote sensing and spatial analysis contributions to map human african trypanosomiasis risk area', *Departamento de Geografía. Serie Geográfica*. Alcalá de Henares: Universidad de Alcalá.

Lake, A. and Townshend, T. (2006) 'Obesogenic environments: exploring the built and food environments', *The journal of the Royal Society for the Promotion of Health*. doi: 10.1177/1466424006070487.

Laraia, B. A. *et al.* (2012) 'Place matters: Neighborhood deprivation and cardiometabolic risk factors in the Diabetes Study of Northern California (DISTANCE)', *Social Science & Medicine*. Pergamon, 74(7), pp. 1082–1090. doi: 10.1016/J.SOCSCIMED.2011.11.036.

Leal, C. and Chaix, B. (2011) 'The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda.', *Obesity reviews : an official journal of the International Association for the Study of Obesity*. England, 12(3), pp. 217–230. doi: 10.1111/j.1467-789X.2010.00726.x.

Lee, A. C. K. and Maheswaran, R. (2011) 'The health benefits of urban green spaces: a review of the evidence', *Journal of Public Health*, 33(2), pp. 212–222. Available at: http://dx.doi.org/10.1093/pubmed/fdq068.

Lee, I. M. *et al.* (2012) 'Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy', *The Lancet*. doi: 10.1016/S0140-6736(12)61031-9.

Lee, J. G. *et al.* (2015) 'A systematic review of neighborhood disparities in point-of-sale tobacco marketing', *American Journal of Public Health*. doi: 10.2105/AJPH.2015.302777.

Lenntorp, B. (1976) *Paths in space-time environments: A time-geographic study of movement possibilities of individuals*, *Dept. of Geography*. Lund: Royal University of Lund.

Lenntorp, B. (1999) 'Time-geography - At the end of its beginning', *GeoJournal*. doi: 10.1023/A:1007067322523.

Lenormand, M. *et al.* (2014) 'Cross-checking different sources of mobility information', *PLoS ONE*. doi: 10.1371/journal.pone.0105184.

LeSage, J. P. (1999) 'The theory and practice of spatial econometrics'. Toledo, Ohio: University of Toledo, 28, p. 11.

Liese, A. D. *et al.* (2010) 'Validation of 3 food outlet databases: Completeness and geospatial accuracy in rural and urban food environments', *American Journal of Epidemiology*. doi: 10.1093/aje/kwq292.

Liese, A. D. *et al.* (2013) 'Characterizing the food retail environment: Impact of count, type, and geospatial error in 2 secondary data sources', *Journal of Nutrition Education and Behavior*. doi: 10.1016/j.jneb.2013.01.021.

Lim, S. S. *et al.* (2012) 'A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010', *The Lancet*. doi: 10.1016/S0140-6736(12)61766-8.

Longley, P. A. *et al.* (2005) *Geographic Information Systems and Science*, *John Wiley & Sons*. doi: 978-0-470-87002-0.

Lopez, A. D. *et al.* (2006) *Global Burden of Disease and Risk Factors*, *Global Burden of Disease and Risk Factors*. doi: 10.1596/978-0-8213-6262-4.

Luo, F. *et al.* (2016) 'Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago', *Applied Geography*. doi: 10.1016/j.apgeog.2016.03.001.

Lytle, L. A. (2009) 'Measuring the Food Environment. State of the Science', *American Journal of Preventive Medicine*. doi: 10.1016/j.amepre.2009.01.018.

Lytle, L. A. and Sokol, R. L. (2017) 'Measures of the food environment: A systematic review of the field, 2007–2015', *Health and Place*. doi: 10.1016/j.healthplace.2016.12.007.

Maantay, J. (2002) 'Mapping environmental injustices: Pitfalls and potential of geographic information systems in assessing environmental health and equity', *Environmental Health Perspectives*. doi: 10.2307/3455050.

Macdonald, L. *et al.* (2018) 'Do "environmental bads" such as alcohol, fast food, tobacco, and gambling outlets cluster and co-locate in more deprived areas in Glasgow City, Scotland?', *Health and Place*. doi: 10.1016/j.healthplace.2018.04.008.

Macdonald, S. *et al.* (2013) 'Alcohol consumption and injury', in *Alcohol: Science, Policy and Public Health*. Oxford University Press Oxford, pp. 171–178.

Macintyre, S. (2007) 'Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity?', *The international journal of behavioral nutrition and physical activity*. BioMed Central, 4, p. 32. doi: 10.1186/1479-5868-4-32.

Macintyre, S. and Ellaway, A. (2003) 'Neighborhoods and Health: An Overview', in Kawachi, I. and Berkman, L. (eds) *Neighbourhoods and health*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195138382.003.0002.

Macintyre, S., Ellaway, A. and Cummins, S. (2002) 'Place effects on health: How can we conceptualise, operationalise and measure them?', in *Social Science and Medicine*. doi: 10.1016/S0277-9536(01)00214-3.

Mackenbach, J. P. *et al.* (2003) 'Widening socioeconomic inequalities in mortality in six Western European countries', *International Journal of Epidemiology*, 32(5), pp. 830–837. Available at: http://dx.doi.org/10.1093/ije/dyg209.

Malambo, P. *et al.* (2016) 'Built environment, selected risk factors and major cardiovascular disease outcomes: A systematic review', *PLoS ONE*. doi: 10.1371/journal.pone.0166846.

Marmot, M. *et al.* (1997) 'Social inequalities in health: Next questions and converging evidence', *Social Science and Medicine*. doi: 10.1016/S0277-9536(96)00194-3.

Marmot, M. (2005) 'Social determinants of health inequalities', *The Lancet*. Elsevier, 365(9464), pp. 1099–1104. doi: 10.1016/S0140-6736(05)71146-6.

Marmot, M. (2006) 'Health in an unequal world', *Lancet*. doi: 10.1016/S0140-6736(06)69746-8.

Matthews, S. A., Moudon, A. V. and Daniel, M. (2009) 'Work Group II: Using Geographic Information Systems for Enhancing Research Relevant to Policy on Diet, Physical Activity, and Weight', *American Journal of Preventive Medicine*. doi: 10.1016/j.amepre.2009.01.011.

McCormack, G. R. and Shiell, A. (2011) 'In search of causality: a systematic review of the relationship between the built environment and physical activity among adults', *The International Journal of Behavioral Nutrition and Physical Activity*. doi: 1479-5868-8-125 [pii] 10.1186/1479-5868-8-125.

Meyer, K. A. *et al.* (2015) 'Combined measure of neighborhood food and physical activity environments and weight-related outcomes: The CARDIA study', *Health and Place*. doi: 10.1016/j.healthplace.2015.01.004.

Middelweerd, A. *et al.* (2014) 'Apps to promote physical activity among adults: A review and content analysis', *International Journal of Behavioral Nutrition and Physical Activity*. doi: 10.1186/s12966-014-0097-9.

Moore, L. V *et al.* (2008) 'Associations of the local food environment with diet quality--a comparison of assessments based on surveys and geographic information systems: the multi-ethnic study of atherosclerosis.', *American journal of epidemiology*. doi: 10.1093/aje/kwm394.

Nau, C. *et al.* (2015) 'Exploring the forest instead of the trees: An innovative method for defining obesogenic and obesoprotective environments', *Health and Place*. doi: 10.1016/j.healthplace.2015.08.002.

Nelson, J. K. and Brewer, C. A. (2017) 'Evaluating data stability in aggregation structures across spatial scales: revisiting the modifiable areal unit problem', *Cartography and Geographic Information Science*. doi: 10.1080/15230406.2015.1093431.

Nelson, M. C. *et al.* (2006) 'Built and Social Environments. Associations with Adolescent Overweight and Activity', *American Journal of Preventive Medicine*. doi: 10.1016/j.amepre.2006.03.026.

Nguyen, Q. C. *et al.* (2016) 'Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity', *Applied Geography*. doi: 10.1016/j.apgeog.2016.06.003.

Nguyen, Q. C. *et al.* (2017) 'Geotagged US tweets as predictors of county-level health outcomes, 2015-2016', *American Journal of Public Health*. doi: 10.2105/AJPH.2017.303993.

Van Oers, J. A. M. *et al.* (1999) 'Alcohol consumption, alcohol-related problems, problem drinking, and socioeconomic status', *Alcohol and Alcoholism*. doi: 10.1093/alcalc/34.1.78.

Office for National Statistics (2017) *Health state life expectancies, UK: 2014 to 2016*, *Office for National Statistics*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/healthstatelifeexpectanciesuk/2014to2016#life-expectancy-has-improved-fastest-in-the-london-borough-of-camden (Accessed: 31 January 2019).

Openshaw, S. (1983) 'The modifiable area unit problem', *Concepts and Techniques in Modern Geography*. doi: 10.1177/1077558707312501.

Openshaw, S. (1984) 'Modifiable Areal Unit Problem', in *SpringerReference*. doi: 10.1177/1077558707312501.

Papas, M. A. *et al.* (2007) 'The built environment and obesity.', *Epidemiologic reviews*. United States, 29, pp. 129–143. doi: 10.1093/epirev/mxm009.

Pearce (2000) 'The ecological fallacy strikes back', *Journal of Epidemiology and Community Health*. doi: 10.1136/jech.54.5.326.

Pearce, J. R. *et al.* (2010) 'Environmental justice and health: the implications of the socio-spatial distribution of multiple environmental deprivation for health inequalities in the United Kingdom', *Transactions of the Institute of British Geographers*. John Wiley & Sons, Ltd, 35(4), pp. 522–539. doi: 10.1111/j.1475-5661.2010.00399.x.

Perchoux, C. *et al.* (2013) 'Conceptualization and measurement of environmental exposure in epidemiology: Accounting for activity space related to daily mobility', *Health and Place*. doi: 10.1016/j.healthplace.2013.01.005.

Peters, D. H. *et al.* (2008) 'Poverty and access to health care in developing countries', *Annals of the New York Academy of Sciences*. doi: 10.1196/annals.1425.011.

Pikora, T. *et al.* (2003) 'Developing a framework for assessment of the environmental determinants of walking and cycling', *Social Science and Medicine*. doi: 10.1016/S0277-9536(02)00163-6.

Pikora, T. J. *et al.* (2006) 'Neighborhood environmental factors correlated with walking near home: Using SPACES', *Medicine and Science in Sports and Exercise*. doi: 10.1249/01.mss.0000210189.64458.f3.

Popova, S. *et al.* (2009) 'Hours and days of sale and density of alcohol outlets: Impacts on alcohol consumption and damage: A systematic review', *Alcohol and Alcoholism*. doi: 10.1093/alcalc/agp054.

Quigg, R. *et al.* (2010) 'Using accelerometers and GPS units to identify the proportion of daily physical activity located in parks with playgrounds in New Zealand children', *Preventive Medicine*. doi: 10.1016/j.ypmed.2010.02.002.

Ramis, R. *et al.* (2009) 'Study of non-Hodgkin's lymphoma mortality associated with industrial pollution in Spain, using Poisson models', *BMC Public Health*. doi: 10.1186/1471-2458-9-26.

Rao, M. *et al.* (2007) 'The built environment and health.', *Lancet (London, England)*. Elsevier, 370(9593), pp. 1111–3. doi: 10.1016/S0140-6736(07)61260-4.

Reed, D. D., Niileksela, C. R. and Kaplan, B. A. (2013) 'Behavioral economics: a tutorial for behavior analysts in practice', *Behavior analysis in practice*. doi: 10.1007/BF03391790.

Rey, G. *et al.* (2009) 'Ecological association between a deprivation index and mortality in France over the period 1997 - 2001: Variations with spatial scale, degree of urbanicity, age, gender and cause of death', *BMC Public Health*. doi: 10.1186/1471-2458-9-33.

Richardson, E. A. *et al.* (2014) 'Have regional inequalities in life expectancy widened within the European Union between 1991 and 2008?', *European Journal of Public Health*. doi: 10.1093/eurpub/ckt084.

Robert, S. A. (1998) 'Community-level socioeconomic status effects on adult health.', *Journal of health and social behavior*. United States, 39(1), pp. 18–37.

Roberts, E. M. (1997) 'Neighborhood social environments and the distribution of low birthweight in Chicago', *American journal of public health*, 87(4), pp. 597–603. Available at: https://www.ncbi.nlm.nih.gov/pubmed/9146438.

Robinson, W. S. (1950) 'Ecological correlations and the behavior of individuals.', *American Sociological Review*. US: American Sociological Assn, 15, pp. 351–357. doi: 10.2307/2087176.

Roerecke, M. and Rehm, J. (2012) 'Alcohol intake revisited: Risks and benefits', *Current Atherosclerosis Reports*. doi: 10.1007/s11883-012-0277-5.

Root, E. D. (2012) 'Moving Neighborhoods and Health Research Forward: Using Geographic Methods to Examine the Role of Spatial Scale in Neighborhood Effects on Health', *Annals of the Association of American Geographers*. doi: 10.1080/00045608.2012.659621.

Rose, G. (1985) 'Sick individuals and sick populations', *International Journal of Epidemiology*. doi: 10.1093/ije/30.3.427.

Roth, G. A. *et al.* (2017) 'Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015', *Journal of the American College of Cardiology*. doi: 10.1016/j.jacc.2017.04.052.

Rothman, K. J., Greenland, S. and Associate, T. L. L. (2014) *Modern Epidemiology, 3rd Edition*, *The Hastings Center report*. doi: 10.1002/hast.292.

Rundle, A. *et al.* (2009) 'Neighborhood food environment and walkability predict obesity in New York City', *Environmental Health Perspectives*. doi: 10.1289/ehp.11590.

Saelens, B. E. *et al.* (2012) 'Obesogenic neighborhood environments, child and parent obesity: The neighborhood impact on kids study', *American Journal of Preventive Medicine*. doi: 10.1016/j.amepre.2012.02.008.

Sallis, J. F. and Owen, N. (2015) 'Ecological models of health behavior.', in *Health behavior: Theory, research, and practice*. doi: 10.7326/0003-4819-116-4-350_1.

Sánchez-Díaz, G. *et al.* (2018) 'Geographic Analysis of Motor Neuron Disease Mortality and Heavy Metals Released to Rivers in Spain', *International journal of environmental research and public health*. MDPI, 15(11), p. 2522. doi: 10.3390/ijerph15112522.

Schneider, S. and Gruber, J. (2013) 'Neighbourhood deprivation and outlet density for tobacco, alcohol and fast food: First hints of obesogenic and addictive environments in Germany', *Public Health Nutrition*. doi: 10.1017/S1368980012003321.

Schoeppe, S. *et al.* (2017) 'Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: A review of quality, features and behaviour change techniques', *International Journal of Behavioral Nutrition and Physical Activity*. doi: 10.1186/s12966-017-0538-3.

Schuurman, N. *et al.* (2007) 'Deprivation indices, population health and geography: An evaluation of the spatial effectiveness of indices at multiple scales', *Journal of Urban Health*. doi: 10.1007/s11524-007-9193-3.

Shaw, M. *et al.* (2000) 'Increasing mortality differentials by residential area level of poverty: Britain 1981–1997', *Social Science & Medicine*. Pergamon, 51(1), pp. 151–153. doi: 10.1016/S0277-9536(99)00434-7.

Shelton, T., Poorthuis, A. and Zook, M. (2015) 'Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information', *Landscape and Urban Planning*. doi: 10.1016/j.landurbplan.2015.02.020.

Shortt, N. K. *et al.* (2015) 'A cross-sectional analysis of the relationship between tobacco and alcohol outlet density and neighbourhood deprivation', *BMC Public Health*. doi: 10.1186/s12889-015-2321-1.

Stanaway, J. D. *et al.* (2018) 'Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Stu', *The Lancet*. doi: 10.1016/S0140-6736(18)32225-6.

Stone, A. A. and Shiffman, S. (1994) 'Ecological Momentary Assessment (Ema) in Behavioral

Medicine', *Annals of Behavioral Medicine*. doi: 10.1093/abm/16.3.199.

Stronks, K. *et al.* (2016) *Social justice and human rights as a framework for addressing social determinants of health*, *WHO Regional Office for Europe*. Available at: http://www.euro.who.int/__data/assets/pdf_file/0006/334356/HR-task-report.pdf?ua=1.

Sureda, X. *et al.* (2017) 'Development and evaluation of the OHCITIES instrument: Assessing alcohol urban environments in the Heart Healthy Hoods project', *BMJ Open*, 7(10). doi: 10.1136/bmjopen-2017-017362.

Takano, T., Nakamura, K. and Watanabe, M. (2002) 'Urban residential environments and senior citizens' longevity in megacity areas: The importance of walkable green spaces', *Journal of Epidemiology and Community Health*. doi: 10.1136/jech.56.12.913.

Thornton, L. E. *et al.* (2012) 'Does the choice of neighbourhood supermarket access measure influence associations with individual-level fruit and vegetable consumption? A case study from Glasgow', *International Journal of Health Geographics*. doi: 10.1186/1476-072X-11-29.

Tobler, W. R. (1970) 'A Computer Movie Simulating Urban Growth in the Detroit Region', *Economic Geography*. doi: 10.2307/143141.

Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and deprivation. Inequality and the North*. London: Routledge: Croom Helm.

Townshend, T. and Lake, A. (2016) 'Obesogenic environments: Current evidence of the built and food environments', *Perspectives in Public Health*. doi: 10.1177/1757913916679860.

UN DESA (2018) *World Urbanization Prospects: The 2018 Revision, Key Facts*, *United Nations, Department of Economic and Social Affairs, Pupulation Division*. doi: (ST/ESA/SER.A/366).

Valiente, R. *et al.* (2018) 'Regulating the local availability of tobacco retailing in Madrid, Spain: A GIS study to evaluate compliance', *Tobacco Control*. doi: 10.1136/tobaccocontrol-2018-054269.

Villar, F. *et al.* (2007) *Las enfermedades cardiovasculares y sus factores de riesgo en España: hechos y cifras*. Madrid: Asociación Española de Arteriosclerosis.

Waitzman, N. J. and Smith, K. R. (1998) 'Phantom of the area: Poverty-area residence and mortality in the United States', *American Journal of Public Health*. doi: 10.2105/AJPH.88.6.973.

Wall, M. M. *et al.* (2012) 'Patterns of obesogenic neighborhood features and adolescent weight: A comparison of statistical approaches', *American Journal of Preventive Medicine*. doi: 10.1016/j.amepre.2012.02.009.

Wen, C. P. and Wu, X. (2012) 'Stressing harms of physical inactivity to promote exercise', *The Lancet*. doi: 10.1016/S0140-6736(12)60954-4.

Weng, M. *et al.* (2017) 'Area Deprivation and Liver Cancer Prevalence in Shenzhen, China: A Spatial Approach Based on Social Indicators', *Social Indicators Research*, 133(1), pp. 317–332. doi: 10.1007/s11205-016-1358-6.

Wheeler, D. C. and Paez, A. (2010) *Geographically weighted regression*, *Handbook of applied spatial analysis: Software tools, methods and applications*. doi: 10.1007/978-3-642-03647-7.

WHO (2008) 'Our cities, our health, our future', *Report to the WHO Commission on Social Determinants of Health from the Knowledge Network on Urban Settings*, pp. 1–199. doi: 10.1007/978-1-4020-8496-6.

WHO (2010) *Global recommendations on physical activity for health*, *Journal of Chemical Information and Modeling*. doi: 10.1080/11026480410034349.

WHO (2014) 'Global status report on alcohol and health', *World Health Organization*. doi: /entity/substance_abuse/publications/global_alcohol_report/en/index.html.

WHO (2015) 'Physical activity strategy for the WHO European Region 2016-2025', *Organización Mundial de la Salud*. doi: 10.1177/1039856213506502.

WHO (2017) *WHO report on the global tobacco epidemic, 2017: monitoring tobacco use and prevention policies*. Geneva: World Health Organization. Available at: https://www.who.int/tobacco/global_report/2017/en/.

WHO (2018a) *Global status report on alcohol and health 2018*. doi: /entity/substance_abuse/publications/global_alcohol_report/en/index.html.

WHO (2018b) *World health statistics 2018: monitoring health for the SDGs, sustainable development goals*, *World Health Organization*. Geneva. Available at: https://www.who.int/gho/publications/world_health_statistics/2018/en/.

Widener, M. J. and Li, W. (2014) 'Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US', *Applied Geography*. doi: 10.1016/j.apgeog.2014.07.017.

Willett, W. C. *et al.* (1995) 'Mediterranean diet pyramid: A cultural model for healthy eating', *American Journal of Clinical Nutrition*. doi: 10.1093/ajcn/61.6.1402S.

Yang, W. and Mu, L. (2015) 'GIS analysis of depression among Twitter users', *Applied Geography*. doi: 10.1016/j.apgeog.2014.10.016.

Yin, J. *et al.* (2016) 'Exploring Multi-Scale Spatiotemporal Twitter User Mobility Patterns with a Visual-Analytics Approach', *ISPRS International Journal of Geo-Information*. doi: 10.3390/ijgi5100187.

ANNEX I. Scientific papers related to this dissertation

Rivera, J., Franco, M., Conde, P., Sandín, M., Gutiérrez, M., **Cebrecos, A.**, Sainz, A., Gittelsohn, J., 2018. Understanding Urban Health Inequalities: Methods and Design of the Heart Health Hoods Qualitative Project. Gac. Sanit. https://doi.org/10.1016/J.GACETA.2018.07.010

Gullón, P., Bilal, U., **Cebrecos, A.**, Badland, H.M., Galán, I., Franco, M., 2017. Intersection of neighborhood dynamics and socioeconomic status in small-area walkability: The Heart Healthy Hoods project. Int. J. Health Geogr. https://doi.org/10.1186/s12942-017-0095-7

Sureda, X., Espelt, A., Villalbí, J.R., **Cebrecos, A.**, Baranda, L., Pearce, J., Franco, M., 2017. Development and evaluation of the OHCITIES instrument: Assessing alcohol urban environments in the Heart Healthy Hoods project. BMJ Open 7. https://doi.org/10.1136/bmjopen-2017-017362

Díez, J., Bilal, U., **Cebrecos, A.**, Buczynski, A., Lawrence, R.., Glass, T., Escobar, F., Gittelsohn, J., Franco, M., 2016. Understanding differences in the local food environment across countries: A case study in Madrid (Spain) and Baltimore (USA). Prev. Med. (Baltim). 89. https://doi.org/10.1016/j.ypmed.2016.06.013

Gullón, P., Badland, H.M, Alfayate, S., Bilal, U., Escobar, F., **Cebrecos, A.**, Diez, J., Franco, M., 2015. Assessing Walking and Cycling Environments in the Streets of Madrid: Comparing On-Field and Virtual Audits. J. Urban Heal. 92. https://doi.org/10.1007/s11524-015-9982-z

ANNEX II. Communications in scientific conferences

XXXV Scientific Meeting of the Spanish Society of Epidemiology y XII Congresso da Associação Portuguesa de Epidemiologia (11-14 September 2018, Lisboa, Portugal)

- Díez, J., Francia, I., Bilal, U., **Cebrecos, A.**, Franco, M. Mercados municipales, ¿la opción saludable en nuestras ciudades? Proyecto Heart Healthy Hoods

- Marí dell'Olmo, M., Domínguez-Berjón, M.F., Duque, I., **Cebrecos, A.**, Prieto Salceda, M.D., Rodríguez Sanz, M., Esnaola, S., Plaza, I., Rodrigo, P., Calvo, M. Elaboración de un índice de privación socioeconómica en las secciones censales de España basado en el censo de 2011.

- **Cebrecos, A.**, Domínguez-Berjón, MF., Duque, I., Klein, O., Franco, M., Escobar, F. Estabilidad estadística y geográfica de índices de privación en distintas áreas administrativas

14th International Conference On Urban Health (26-29 September 2017, Coimbra, Portugal)

- **Cebrecos, A.**, Domínguez-Berjón, MF., Duque, I., Klein, O., Franco, M., Escobar, F. Issues arising from multi-scale analysis of urban deprivation índices

- Gullón, P., Bilal, U., **Cebrecos, A.**, Badland, H., Galán, I. and Franco, M. Intersection of Neighborhood Dynamics and Socioeconomic Status in Small-Area Walkability: The Heart Healthy Hoods Project

XXXV Scientific Meeting of the Spanish Society of Epidemiology y XII Congresso da Associação Portuguesa de Epidemiologia (6-8 September, Barcelona, Spain)

- Díez, J., Francia, I., Bilal, U., **Cebrecos, A.**, Franco, M. Mercados municipales, ¿la opción saludable en nuestras ciudades? Proyecto Heart Healthy Hoods

- Marí dell'Olmo, M., Domínguez-Berjón, M.F., Duque, I., **Cebrecos, A.**, Prieto Salceda, M.D., Rodríguez Sanz, M., Esnaola, S., Plaza, I., Rodrigo, P., Calvo, M. Elaboración de un índice de privación socioeconómica en las secciones censales de España basado en el censo de 2011.

- **Cebrecos, A.**, Domínguez-Berjón, MF., Duque, I., Klein, O., Franco, M., Escobar, F. Estabilidad estadística y geográfica de índices de privación en distintas áreas administrativas

XXXIV Scientific Meeting of the Spanish Society of Epidemiology (SEE) y XI Congresso da Associação Portuguesa de Epidemiologia (APE) (14-16 September 2016, Sevilla, Spain)

- **Cebrecos, A.**, Domínguez-Berjón, MF., Duque, I., Klein, O., Franco, M., Escobar, F. Estabilidad estadística y geográfica de índices de privación en distintas áreas administrativas.

- Domínguez-Berjón, MF., **Cebrecos, A.**, Duque, I., Calvo, M., Marí-Dell'Olmo, M., Plaza, I., Prieto, D., Rodriguez-Sanz, M. Indicadores de privación socioeconómica en España en 2011 por sección censal.

- Valiente R., **Cebrecos A.**, Franco M., Escobar F. Salud en los mapas: la aplicación Cartográfica del Proyecto Photovoice sobre Alimentación en Villaverde.

- Sureda X., Villalbí JR., Espelt A., Baranda L., **Cebrecos A.**, Moure L., Fuentes S., Franco M. Desarrollo de un instrumento para caracterizar el entorno urbano en relación con el alcohol en Madrid y Barcelona.
- Gullón P., Bilal U., **Cebrecos A.**, Díez J., Sureda X., Franco M. Comparación de la caminabilidad medida por sistemas de información geográfica y con instrumento de medición directa.

- **Cebrecos A.**, Diez J., Gullón P., Bilal U., Franco M., Escobar F. Un método SIG para caracterizar el entorno obesogénico en áreas urbanas: una propuesta multivariable.

- Diez J., Valiente R., **Cebrecos A.**, Olea A., García R., Ramos C., Franco M. Hacia un entorno alimentario saludable: barreras y oportunidades en Madrid usando métodos mixtos.

XVII National Congress of Geographic Information Technologies (June-July 2016, Málaga, Spain)

- **Cebrecos A.**, Diez J., Gullón P., Bilal U., Franco M., Escobar M. Un método SIG para caracterizar el entorno obesogénico en áreas urbanas: una propuesta multivariable

13th International Conference of Urban Health (March-April 2016, San Francisco, CA, US)

- **Cebrecos A.**, Diez J., Gullón P., Bilal U., Franco M., Escobar M. Healthy urban environment characterization focused on physical activity and food: A GIS-based method.

- Gullón P., **Cebrecos A.**, Sureda X., Bilal U., Diez J., Escobar F., Franco M. Does walkability differ by area sociodemographic profile? A study of Madrid City

- Diez J., Bilal U., **Cebrecos A.**, Buczynski A., Lawrence R., Glass T., Escobar F., Gittelsohn J., Franco M. Comparing local availability and accessibility to healthy foods across countries: A case study in Madrid (Spain) and Baltimore (US)

Reimagining health in cities Symposium (September 2015, Philadelphia, PA, US)

- Gullon, P., Badland,H., Alfayate, S., Bilal, U., Escobar, F., **Cebrecos, A.**, Diez, J., Franco, M. Assessing walking and cycling environments in the streets of Madrid: Comparing on-field and virtual audits.

- Diez, J., Bilal, U., **Cebrecos, A.**, Buczynski, A., Lawrence, R., Glass, T., Escobar, F., Gittelsohn, J., Franco, M. Access to food stores and healthy food availability: comparative study between Madrid (Spain) and Baltimore (US)

- Gullon, P., **Cebrecos, A.**, Sureda, X., Bilal, U., Diez, J., Escobar, F., Franco, M. Neighborhood sociodemographic status and walkability in Madrid

II Iberoamerican Congress of Epidemiology and Public Health (September 2015, Spain)

- Díez,J., Bilal, U., **Cebrecos, A**., Franco, M. Acceso a comercios de alimentación y disponibilidad de alimentos saludables: estudio comparativo entre Madrid (España) y Baltimore (EE.UU).

- Sureda, X., **Cebrecos, A**., Bilal, U., Gullón, P., Díez, J., Fuentes, S., Franco, M. Densidad de estancos y características sociodemográficas en los barrios de Madrid.

- Gullón, P., **Cebrecos A.**, Sureda X., Bilal U., Díez J., Escobar F., Franco M. Variación Vidas de barrio en transformación y posible impacto en la salud de la caminabilidad en áreas de diferente nivel sociodemográfico en Madrid.

ANNEX III. Characterizing physical activity and food urban environments: a GIS-based multicomponent proposal (Supplementary material)

S1: The Median Neighborhood Index Methodological Details

The Median Neighborhood Index

The Median Neighborhood Index (MNI) is the average Euclidean rank distance of each spatial unit of analysis to the median neighborhood in a series of variables. More specifically, this index uses four variables to represent the demographic and socioeconomic structure, segregation phenomena and urban form. The Euclidean rank distance in each variable is calculated by sorting all units of analysis (census sections) and computing how far in rank each unit is from the median neighborhood. The four distances are then averaged to the Median Neighborhood Index. A low value in this index represents a more average neighborhood in the four variables, while a higher value represents more extreme neighborhoods. Importantly, and since rank distances are all positive, these extreme neighborhoods may be on either tail of the distribution of social factors.

Variables

For the four variables, we used % population aged 65 or above as the demographic indicator, % people with college education or above as the socioeconomic indicator, % foreign-born as the segregation indicator, and population density (in sq. km) as the urban form indicator. The unit of analysis was the census section (around 1500 people).

Selecting Average Neighborhoods

To select average neighborhoods we look for clusters of spatial units of analysis of the desired size. For example, if the unit of analysis is the census sections and we seek an area of 15,000 people, we must seek clusters of a maximum of 12 census sections. We use Kulldorf's Spatial Scan Statistic (Kuldorff 1997). This method allows for the search of clusters of cases, normally distributed variables or other distributions. Given the normally distributed nature of the MNI, we looked for clusters of low MNI values. The Kulldorf's Spatial Scan Statistic also allows for the setting of a maximum cluster size. Given that this statistic requires for spatial point data to be used, we calculated the centroids of each spatial unit of analysis prior to the cluster search.

S2: Adapted NEMS-S Audit Tool

Store ID: ____________ Store #: ______	**Madrid Store Study** **Healthy Food Availability Index**	Data Collector: ______

Type: [ ] Public Market [ ] Supermarket [ ] Small Grocery [ ] Specialty Store [ ] Discount Store [ ] Corner Store [ ] Convenience Store [ ] Gas Station	Store Name:					CS:	Date:
	Store Address:					Neighborhood:	
	[ ] Confirmed [ ] New [ ] Absent	Ethnic: [ ] Yes [ ] No	# Registers:	# Aisles:	Prepared Food: [ ] Yes [ ] No	Parking Lot: [ ] Yes [ ] No	
	Comments:				Photo: [ ] Yes [ ] No	Refusal: [ ] Yes [ ] No	

Measure 1: MILK		Measure 3: FRUIT		Measure 4: VEGETABLES		Measure 5: MEATS	
Available:	[1] Yes [ ] No	Available:	[ ] Yes [ ] No	Available:	[ ] Yes [ ] No	Ground Meats Available:	[1] Yes [ ] No
Low Fat Option(s) Available:	[1] Yes [ ] No	Quality:	[ ] A [ ] UA	Quality:	[ ] A [ ] UA	Quality:	[ ] A [ ] UA
Measure 2: JUICE		Type(s) Available: ______	[ ] 0 [1] 1-3 [2] 4-6 [3] 7-10 [4] 11-25 [5] >25	Type(s) Available: ______	[ ] 0 [1] 1-3 [2] 4-6 [3] 7-10 [4] 11-25 [5] >25	Beef:	[1] A [ ] UA
100% Fruit Juice Available:	[1] Yes [ ] No					Other Options Available:	[ ] Yes [ ] No
Comments:		Comments: Whole ______ Cut ______		Comments: Potatoes ____ Onions ____ Whole ______ Cut ______		Comments:	

Measure 6: CHICKEN		Measure 8: FROZEN FOODS		Measure 9: PACKAGED FOODS		Measure 11: BREAD	
Available:	[1] Yes [ ] No	Meal(s) Available:	[ ] Yes [ ] No	Dried Beans Available:	[.5] Yes [ ] No	Available:	[1] Yes [ ] No
Quality:	[ ] A [ ] UA	Healthier Meal(s) Available:	[1] Yes [ ] No	Rice Available:	[.5] Yes [ ] No	100% Whole Wheat:	[1] Yes [ ] No
Measure 7: SEAFOOD				Pasta(s) Available:	[.5] Yes [ ] No	Corn Tortillas Available:	[1] Yes [ ] No
Available:	[1] Yes [ ] No	Fruits(s) Available:	[1] Yes [ ] No	**Measure 10: CANNED FOODS**		**Measure 12: CEREAL**	
Quality:	[ ] A [ ] UA	Vegetables(s) Available:	[1] Yes [ ] No	Soup(s) Available:	[ ] Yes [ ] No	Available:	[ ] Yes [ ] No
Option(s) Available:	[ ] Fresh [ ] Frozen [ ] Both	Comments:		Low-Sodium Soup(s) Available:	[1] Yes [ ] No	Low Sugar Options:	[1] Yes [ ] No
Comments:				Fruit(s) Available:	[.5] Yes [ ] No	# Healthy Varieties:	[ ] 0 [ ] 1 [1] 2+
				Vegetable(s) Available:	[1] Yes [ ] No	Comments:	
				Comments:			

Adaptations (from the Abridged NEMS version produced by the CLF [available at http://mdfoodsystemmap.org/):

- +1 For Ground beef available changed to +1 For any ground meat
- +1 for Lean ground beef changed to +1 for beef
- .5 for frozen fruits or vegetables changed to +1 for each
- +1 if any of dried beans, rice or pasta available, changed to +0.5 for each

S3: Density of Food, Alcohol and Tobacco Outlets in our Study Area, in the District of Ciudad Lineal and in the entire City of Madrid. Density units are in businesses per 10.000 people

TYPE OF BUSINESS	STUDY AREA	DISTRICT	MADRID
FOOD STORES			
GENERAL	6.6	4.2	5.4
FRUIT	17.1	6.1	6.3
MEAT/EGGS	26.9	8.2	9.2
FRUIT	13.1	2.5	2.4
BAKERY	16.4	7.1	8.4
FOOD SERVICES, ON-SALE ALCOHOL AND TOBACCO VENDING (BARS/RESTAURANTS)	39.4	41.1	49.9
ALCOHOL OFF-SALE	0.6	0.2	0.4
TOBACCO STORES (*ESTANCOS*)	2.6	1.5	1.9

ANNEX IV. A multicomponent method assessing healthy cardiovascular urban environments: The Heart Healthy Hoods Index (Supplementary material)

S1: Neighborhood deprivation indicators

Table 1. Operational definitions for area-based socioeconomic indicators of the 2011 census included in the construction of the area based deprivation index.

	Indicator	Operational definition
SOCIOECONOMIC GROUP	Manual workers ≥ 16 years old among the employed	Percentage of people aged 16 years or over employed in sectors services, agriculture, fishing, craftwork, skilled workers in manufacturing industries, construction, mining, installations operators, and non-skilled workers; with respect to the total employed population aged 16 years or over
	Manual workers ≥ 16 years old among the employed or unemployed who have worked before	Percentage of people aged 16 years or over, employed or unemployed who have worked before as manual workers (employed in sectors: services, agriculture, fishing, craftwork, skilled workers in manufacturing industries, construction, mining, installations operators, and non-skilled workers) with respect to the total employed population aged 16 years or over
	Unemployment ≥ 16 years old	Percentage of people aged 16 years or over without a job (unemployed and those seeking work for the first time), with respect to the total economically active population
	Temporary employees ≥ 16 years old among the occupied	Percentage of people aged 16 or older occupied as eventual employees.
	Temporary employees ≥ 16 years old among the occupied or unemployed who have worked before	Percentage of people aged 16 or older occupied or unemployed who have worked before as eventual employees.
	Insufficient instruction ≥ 16 years old	Percentage of people aged 16 years and over who, according to a list of the National Statistics Institute, cannot read or write; can read and write but have less than 5 years schooling; went to school for 5 years or more but did not complete basic compulsory education, with respect to the total population aged 16 years and over.
	Insufficient instruction 16-29 years old	Percentage of people aged between 16 and 29 years with low educational level, with respect to the total population aged from 16 to 29 years.
DEMOGRAPHIC GROUP	Aging	Percentage of population aged 65 or older respect to the total population.
	Born in low income countries	Percentage of population born in countries other than the following: Spain, Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Iceland, Italy, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, Andorra, Germany, San Marino, Vatican City, Sweden, Switzerland, Canada, United States of America, Japan, Australia and New Zealand; respect to the total population
	Born in low income countries arriving in Spain after 2006	Percentage of population born in countries other than the following: Spain, Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Iceland, Italy, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, Andorra, Germany, San Marino, Vatican City, Sweden, Switzerland, Canada, United States of America, Japan, Australia and New Zealand, and who arrived in Spain between 2007 and 2011; respect to the total population.
	Born in low income countries or born in Spain whose father or mother was born in low income countries	Percentage of population born or with their father or mother born in countries other than listed above, respect to the total population.
	Population in single-parent households	Population in households with a single-father who lives at least, with a child ≤ 25 years.

S2: CVD deaths attributable to the four main behavioral risk factors

Table 2: Percentage of death due to cardiovascular disease (CVD) attributable to the four main behavioral risk factors in Spain for the year 2016 obtained from Global Burden Disease (GBD).

Age	Cause of death	Risk factor	Measure	Value	Confidence interval
40 - 75	CVD	Dietary risks	% of total deaths	46.34%	(30.69 – 68.89)
40 - 75	CVD	Tobacco	% of total deaths	23.55%	(9.49– 50.59)
40 - 75	CVD	Low Physical Activity	% of total deaths	6.86%	(2.61 – 11.18)
40 - 75	CVD	Alcohol	% of total deaths	17.12%	(6.41– 19.91)

S3: Spatial analysis results

Table 3: Results of the global spatial correlation analysis (Moran's Index) for the index models, age-standardized rates of CVD and the deprivation index.

	Morans Index (p< 0.01)
Model 0	0.92
Model 1	0.92
CVD rate	0.28
Deprivation index	0.56

ANNEX V. Funding

This PhD dissertation will not be possible without the funding of the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Starting Grant HeartHealthyHoods Agreement n. 336893