

**NEIGHBORHOOD SOCIAL AND ECONOMIC CHANGE, FOOD
ENVIRONMENT CHANGE AND DIABETES INCIDENCE IN MADRID (SPAIN)**

by

USAMA BILAL

A dissertation submitted to Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March, 2017

©2017 USAMA BILAL

All Rights Reserved

ABSTRACT

Background

The dynamic nature of residential environments is an understudied macro-social determinant of health.

Objectives

The aims of this dissertation were to measure neighborhood social and economic change, and to evaluate its association with changes in the retail food environment and diabetes incidence.

Methods

We collected area-level data from multiple administrative sources in the city of Madrid (Spain) from 2005 to 2016. For Aim 1, we computed measures of change in indicators related to residential mobility, socioeconomic and sociodemographic characteristics, and housing construction and renovations. We used a finite mixture model to measure types of areas according to how they change, revealing four types of neighborhood social and economic change. For Aim 2, we geocoded and categorized retail food stores into 4 categories: a) any food store, b) supermarkets, c) small specialized stores and d) fruit and vegetable stores. We used a multinomial logistic regression model to evaluate the association between neighborhood change type and food environment change. For Aim 3, we used data from the HeartHealthyHoods Retrospective Study that included electronic health records on every individual registered in a health center of four

districts of Madrid. We used a Cox proportional hazards model to estimate the association between neighborhood change type and diabetes incidence.

Results

Our discrete measurement model identified four types of neighborhoods (census sections) according to their change: (Type 1) areas with an increasing proportion of foreign-born migrants and a relative worsening in SES markers; (Type 2) areas with high residential mobility and housing constructions, relative reduction in average age and increase in total population; (Type 3) areas with a relative improvement in SES markers and increases in housing renovations; and (Type 4) areas with low residential mobility, and a relative increase in average age, reduction in foreign-born migrants and total population. Type 3 areas were associated with an increase in the number of small specialized stores, a decrease in the number of supermarkets and an increased incidence of diabetes. Type 1 and 4 areas were associated with an increase in the number of supermarkets and decrease in the incidence of diabetes.

Conclusions

Measuring a complex exposure such as neighborhood social and economic change is a challenging endeavor. Further study of this association with food environment changes should include consideration of opening hours. If the finding of an association between neighborhood type and diabetes incidence can be replicated, policy-based diabetes prevention programs may be developed and tested.

ADVISERS AND READERS

David D. Celentano, ScD, MHS (Academic Advisor)

Professor of Epidemiology

Department of Epidemiology

Johns Hopkins Bloomberg School of Public Health

Thomas A. Glass, PhD (Co-Advisor, Thesis reader)

Professor of Epidemiology, Retired

Department of Epidemiology

Johns Hopkins Bloomberg School of Public Health

Manuel Franco, MD, PhD (Thesis reader)

Associate Professor, Adjunct

Department of Epidemiology

Johns Hopkins Bloomberg School of Public Health

Robert S. Lawrence, MD (Thesis reader, Chair)

Professor, Emeritus

Department of Environmental Health Sciences

Johns Hopkins Bloomberg School of Public Health

Joel Gittelsohn, PhD, MSc (Thesis reader)

Professor

Department of International Health

Johns Hopkins Bloomberg School of Public Health

Elizabeth Selvin, PhD, MPH (Alternate Thesis reader)

Professor

Department of Epidemiology

Johns Hopkins Bloomberg School of Public Health

Lawrence J. Appel, MD, MPH (Alternate Thesis reader)

Professor

Department of Medicine

Johns Hopkins Medical Institutions

ACKNOWLEDGEMENTS

There are so many people I am thankful for in during this trip. It is almost as if I need to write another entire chapter just for this purpose.

First, to my advisor during these last few months, David Celentano, thanks for graciously taking on the charge of making sure I finished all this in a timely manner. And thanks for the continuous support of Social Epidemiology in the department. As I will say below, being part of the social epidemiology effort here was one of my main inspirations during this period.

Second, to my advisor during most of the program, Tom Glass. I do not really have the words to describe how much I have learned next to you. I have tried to do my best in condensing some of our discussions in Chapter 6 of this dissertation, but I feel like I would need another entire new dissertation just to percolate on all that knowledge. When someone asks me how my experience was, I always tell them the same thing: I could not have had a better and more suited advisor. Dealing with my permanent state of iteration (lack of focus?) between topics and ideas must not be the easiest task for an advisor. But here we are, almost done. I will always bring thumbtacks. Beyond the academics, I am also very grateful for the true and honest friendship. We will continue to see each other in my next (academic) life.

Third, to my advisor Emeritus, Manuel Franco. Around 9 years ago, I applied to do six weeks of research in some crazy thing about psychosocial factors in myocardial infarction. Epidemiology was not something I had been

interested in, and here we are now. I owe so much to your mentoring and perseverance in keeping me on track during these last 9 years. And I feel like we will continue doing good work and having fun while doing it. This also extends to our shared teaching efforts, which I hope we will continue engaging in in the following years.

To the rest of the “Madrid Team”, the HeartHealthyHoods team I am also really grateful. For always keeping the doors open every time I visited. For dealing with my crazy “gringo” manners every time I go back to Spain. For always finding the time to go out for a beer. Especially to those that were there quite since the beginning: Julia, being the muscle and the force behind an enormous project, being a friend to confess things to, and always keeping a good smile in the face of scientific adversity; and to Pedro, for showing that hope is not lost in the medical establishment (some of us were not brave enough to persevere there). I also want to thank the Primary Care Research Unit in Madrid (especially Luis Sanchez Perruca and Isabel del Cura) for providing the data needed for my third aim. Moreover, none of this would’ve been possible without the help of the people in the Statistical Data Unit of the Madrid City Government (Antonio Bermejo and Remedios Perez). Before I “leave” Spain, I also want to thank the Obra Social La Caixa, which funded the first two years of my PhD and that put me in touch with such an amazing group of La Caixa Fellows. These I also want to thank, especially those I met here in Baltimore and whom I have kept seeing.

Back to Baltimore, I need to thank everyone at the Epidemiology department for making this experience a blast for me. Especially the academic support core. Fran takes care of us and protects us from any academic hardship that may get into our way, always receptive, always helpful; Matt has the strongest work ethic I have seen in my life and really cares about our material means for survival; for everyone else in W6508, including Ayesha, Julie, Sheila, Ebony and Jon, thanks a lot for all your help during this time. I want to also thank the rest of the people in the School that made this possible, including those that are always there but not always seen. This includes the house staff, the multimedia and IT staff, and the daily grind staff.

I also want to thank other people at JHMI that greatly helped me go through this process. Interestingly, doing a PhD in a public health field and taking care of your own health is not an easy task; so, thanks to the Cooley center personnel for always being so cheerful and supportive, and thanks to the UHS staff (especially Dr. Chaudhry) for keeping my health needs on track. As a “Rosean”, I believe in the power of upstream factors in determining weight, but I also acknowledge the importance of individual interventions when it comes to an specific individual’s health.

At the school, I am also extremely thankful to the Welch Center. This is a unique environment to thrive in; a model that must be copied by other epidemiology departments in their multidisciplinary collaborations. So much is done for students and trainees in general, for our growth and for improving our

experiences. I want to be especially thankful to Larry Appel for his leadership there and for his help; to Liz Selvin and Joe Coresh for believing in me as a part of the CVD track; and to all the staff that make the center such a good place to be a trainee in.

My second center is one that I believe will at some point save the world. It certainly saved me by believing in my project and naming me a Center for a Livable Future (CLF)-Lerner Fellow. The type of research and the efforts to reach out and advocate for their findings is an amazing and unique feature of the CLF. I need to emphasize the leadership and kindness of Bob Lawrence, and I am especially thankful to him for serving in both my thesis' and readers' committee. To the rest of the center (especially to Meg), thank you for helping me in these last few years. Thanks also to my "fellow Fellows", from whom I have learned so much and whom I respect so much.

My third home in the school was the Social Epidemiology Student Organization. My faith and optimism in the future of epidemiology is maintained through the people I have met while directing this organization. I want to be especially and extremely thankful to Emily for all her help in these last three years. We "created" the best journal club in the school, feed people the best journal club food and organized the best lectureships ever. All while having fun. We will keep on doing this wherever we continue in our next academic lives. I also want to thank the other Social Epi students, including those in the "social epi

closet". Our field may not produce the fanciest results, but they sure look important to us.

I also want to thank my fellow Epi students, especially those in my cohort. It seems like it was yesterday when it took us two hours to introduce each other in that endless session in August 2013. I only wish the Department had more established structures to allow us to see each other more frequently, because I feel like I could have learned so much from you all. Especial props to Jim, for all the time we spent together during the first two years and to Rene, for your trust and friendship, for all those tacos and biostats questions.

Speaking of learning, I cannot stress how much I have learned from some amazing faculty in this school. A big thanks to people like Alvaro Munoz, who has created a great angst in me: I now know for a fact that I will never be able to find a greater instructor in my life. I look up to you, and once aim at just being able to transmit half as much as you have been able to give to your students. I also want to thank instructors outside of Epi for giving me perspective and insights into public health from a completely different point of view. In particular, I have gotten a great part of my "sociological imagination" through the classes by Kate Smith, and my "political imagination" through Vicente Navarro's classes. If any Epi student ever reads this section, please go take their classes. They will open a new world to you. Before "leaving" the JHSPH I also want to give special thanks to my other research mentors: Bryan Lau and Geetanjali Chander. What started as a way for me to experiment with other parts of epi (and earn some money)

ended up developing into an entire new collaboration, one from which I have learned so much. Thanks for all the support, mentoring, trust and patience.

In my academic world and learning experience I want to be especially thankful to the Santa Fe Institute and the people at the Santa Fe Complex Systems Summer School. The staff and fellow attendees made that month-long experience and unforgettable one. I have completely shifted my view on science, have acquired the inspiration to pursue a different way of looking at my own field, and have gained the confidence to look beyond simple things and into chaos itself.

Onto my life outside of academia (yes, fellow student reading this: there is one!). I want to thank my friends in Baltimore, those that filled my heart during these last few years. The Spaniards in general (especially Paloma, that was already introducing me to Baltimore before I even came here; and Miguelin, for all of those “walking meetings” during your time here), whom endure life far away from home with enthusiasm; the other Hispanics (especially Cristian, whom I met randomly at a bar and we ended up sharing so many IPAs and experiences), with whom we share so much; and the Gringos, that welcomed me so well to their culture (especially Molly). I want to emphasize three people with whom I have shared so much of my time these last three years, and whose friendship I especially treasure and hold close to my heart. Ivanin, Alex and Esther: thanks for accepting me the way I am, for all the time we have spent together, and for sharing so much of your lives.

Circling back to Spain, I want to thank my friends there, that have endured my isolation here and with whom I expect to reconnect at some point. To my small “Watchmen” group (they asked to be referred to as the “dystopian political debate whatsapp group”; I oblige) that kept me connected to the political reality at home; to my “preventivos” whom I was so lucky to share an MPH with; and to my unclassifiable friends, especially to Natalia in London and to Nacho (wherever in the world you are when I print this). I also want to extend my thanks to everyone that contributed to my education. From my 1st grade teacher, who had the firm belief that the best way to keep this restless student quiet was not by reprimanding him but by giving him extra hard math problems to solve; to the teaching nuns in my high school, who told me all there is to know about the IMF, global capitalism and the revolution in Chiapas; to some of my med school professors, especially Manolo, who instilled in me a deep respect for research and scientific inquiry.

In Spain, I also want to give an especial and heartfelt thanks to my family for their support. Especially my parents. To my Dad, who always challenges my ideas and preconceived notions of anything, and who instilled in me the ambition to become more, and to learn more, and to thrive and push; and to my Mom, who gave the best she had to raise me, who insisted on me learning what “empathy” really meant, and who once told me “even this trashcan is the State”; it’s hard to believe how much that simple statement has shaped my ideas over time. To my family in Syria, that has had to endure one of the most horrible wars in the

modern history of humanity while the rest of the world looks passively, for whom I will always be the “دكتور”, and who have always transmitted me their continued love and compassion.

Last (but definitely not least), I want to thank Mariana: for your love, for your patience, for your generosity. For your capacity to deal with the kind of overachieving nutjob I am; with my lack of tolerance for frustration, and for my impatience with my own failures. You have always continued to support me, and that has meant the world for me. I am so proud of you.

TABLE OF CONTENTS

ABSTRACT	ii
ADVISERS AND READERS	iv
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	xiv
LIST OF TABLES	xxi
LIST OF FIGURES	xxiii
CHAPTER 1: INTRODUCTION	1
Overview	2
Cardiovascular Disease and Diabetes	2
Epidemiology of CVD	2
Cardiovascular Risk Factors	3
Epidemiology of Diabetes	4
Bending the Curve: Rose’s Approach to Prevention	5
Motivation for studying mass influences	6
Motivation for studying cities and neighborhoods as mass influences..	7
Challenges in the Study of Neighborhoods	8
Neighborhood Change	8
Overview of theories	8
Neighborhood Selection theories	9
External Shock theories	10
Drivers vs. Passengers: a neighborhood approach	11
The Local Food Environment	12
Distribution of local food environments	12
Neighborhood Change and the Food Environment.....	13
Evidence Gaps in the Literature	17

Specific Aims and Hypothesis.....	19
Organization of the Dissertation	20
Figures	21
References.....	22
CHAPTER 2: MATERIALS AND METHODS	28
Introduction and Overview.....	29
Spatial Units of Analysis	29
The Hierarchical Structure of Administrative Divisions in Spain	30
Administrative Spatial Organization of Madrid	30
The “Fluid Census Section” problem.....	31
The spatially explicit organization of the health system	35
The Overlapping Nature of the Administrative and Health organization of Madrid	36
Conclusion	37
Detailed Materials and Methods and Exploratory Data Analyses for Aim 1: Measuring Neighborhood Social and Economic Change.....	37
Data Sources for the Indicators of Neighborhood Social and Economic Change.....	37
Indicator operationalization	41
Exploratory Data Analysis for Aim 1	44
Detailed Materials and Methods and Exploratory Data Analyses for Aim 2: Association of Neighborhood Social and Economic Change with Food Environment Changes	46
Data Sources	46
Classification of Food Stores	47
Classification of Economic Activities in Spain.....	47
Healthy food availability as approximated by food store type	48
Classification of Food Stores	49
Exploratory Data Analysis for Aim 2.....	50
Detailed Materials and Methods and Exploratory Data Analyses for Aim 3: Association of Neighborhood Social and Economic Change with Diabetes Incidence	51
Parent Study	52

The Spanish Health Care System	52
The Importance of Primary Care in Spain.....	53
Healthcare access: universalization and back	54
The Healthcare Card.....	54
Electronic Health Records in Madrid	55
The HeartHealthyHoods Retrospective Study	56
Exploratory Data Analysis for Aim 3.....	58
Conclusion	60
Tables.....	61
Figures	69
References.....	87
CHAPTER 3: MEASURING NEIGHBORHOOD SOCIAL AND ECONOMIC	
CHANGE FOR HEALTH STUDIES	88
Introduction	91
Justifying the study of Change	91
Neighborhood Selection Processes	91
External Shock Processes	92
The Temporal Nature of Neighborhood Change.....	93
The difference between turnover and trajectories.....	94
Steps in building a measurement model	95
Methods	96
Study setting	96
Data Sources	96
Latent variable Model.....	98
Candidate Indicators	100
Model building and diagnosis.....	101
Representation of results	103
Model validation	104
Analytic Procedures	104
Results	105
Neighborhood Social and Economic Change Final Model.....	105

Neighborhood Social and Economic Change Types.....	105
Spatial Distribution of the 4 Types of Neighborhood Change	108
Transitions between Types of Neighborhood Change	109
Construct validity: Neighborhood Social and Economic Change and Socioeconomic Status.....	110
Discussion.....	112
Strengths and limitations.....	114
Conclusions.....	116
Tables.....	117
Figures	120
References.....	124
CHAPTER 4: NEIGHBORHOOD SOCIAL AND ECONOMIC CHANGE AND RETAIL FOOD ENVIRONMENT CHANGES IN MADRID (SPAIN).....	126
Introduction.....	129
The Local Food Environment is a Mass-Influence on Diet.....	129
Neighborhood Change is Understudied	129
A Measurement Model of Neighborhood Social and Economic Change	130
Neighborhood Change and Food Environment Change: Mechanisms	131
Objective	131
Methods	132
Study setting	132
Neighborhood Social and Economic Change	132
Retail Food Environment Changes	133
Data Analysis	134
Results	137
Retail Food Environment Changes	137
Association of Neighborhood Change and Retail Food Environment Changes.....	138
Discussion.....	141
Limitations and Strengths.....	144

Conclusion	145
Tables.....	147
Figures	149
References.....	152
CHAPTER 5: NEIGHBORHOOD SOCIAL AND ECONOMIC CHANGE AND	
DIABETES INDICENCE OVER 6 YEARS IN MADRID, SPAIN.....	155
Introduction.....	158
Diabetes Prevention Through the Study of Mass Influences	158
Neighborhood Characteristics and Diabetes Burden	158
Neighborhood Relocation and Diabetes	159
Neighborhood Change and Diabetes	159
Potential Mechanisms for the effect of Neighborhood Change on Diabetes	160
Measuring Neighborhood Change	160
Objective	161
Methods	162
Study setting	162
Neighborhood Social and Economic Change	162
Electronic Health Records.....	163
Diabetes Data	163
Statistical Analysis	164
Results	166
Study Population	166
Diabetes Incidence.....	167
Association of Neighborhood Social and Economic Change with Diabetes Incidence.....	167
Sensitivity analyses	168
Discussion.....	169
Conclusion	174
Tables.....	175
Figures	178

References.....	180
CHAPTER 6: CONCLUSION.....	184
Summary of Findings	185
Neighborhood Social and Economic Change	185
Neighborhood Change and Food Environment Changes	186
Neighborhood Change and Diabetes Incidence	186
Challenges.....	187
Trajectories, Stages and the MAUP/MTUP.....	187
Results from our measurement model.....	188
The Signal vs. Noise problem.....	189
Modifiable Areal Unit Problem (MAUP).....	190
Theory-driven vs. Data-driven decisions.....	191
Data-driven approach	191
Theory-driven approach.....	192
Data Availability and Data-Driven Decisions.....	192
Implications and Next Steps.....	195
Lesson Learned: Theory Development (Popper goes to Santa Fe)..	195
Epidemiology, Deduction and Induction.....	195
The Theory Sandbox	197
Developing a Theory.....	198
The Surprise: The Study of Dynamic Relationships and Throughput	201
Throughput and the City	201
Throughput and Change.....	202
Putative Social Mechanisms of Increased Throughput.....	203
Policy Implications	204
Controlling Throughput.....	204
The Consequences of Removing Land-use Regulations.....	205
The Consequences of Removing Restrictions on Opening Hours....	206
Conclusion	208
Figures.....	209
References.....	210

LIST OF APPENDICES213

LIST OF TABLES

Table 2.1. Characteristics by January 1st 2011 of administrative units overall, for Aim 1 and Aim 2 and for Aim 3.	61
Table 2.2. Changes in Census Sections over time in Madrid.....	62
Table 2.3: Full list of the 60 proposed indicators of neighborhood change	63
Table 2.4. Classification of Food Retailing Commercial Spaces Economic Activities under the CNAE1993 and CNAE2009 systems.....	68
Table 3.1. Indicators of Neighborhood Social and Economic Change.....	117
Table 3.2: Transitions Between Types of Neighborhood Change	119
Table 4.1: Number of food stores by type for the entire city of Madrid.	147
Table 4.2: Neighborhood Change (averaged from 2007 to 2011) and gain/loss in food stores (2012 to 2016) adjusted for baseline number of stores and socioeconomic status at baseline (2006)	148
Table 5.1. Description of study sample by neighborhood change type.....	175
Table 5.2. Association (HR, 95% CI) of Neighborhood Social and Economic Change and Diabetes Incidence	177
Appendix 3.1: Detailed results of the final measurement model of neighborhood social and economic change: Means and Probabilities	214
Appendix 3.2: Detailed results of the final measurement model of neighborhood social and economic change: Variances	215

Appendix 3.3: Detailed results of the final measurement model of neighborhood social and economic change: Covariance structure (expressed as correlations).....	216
Appendix 3.4: Detailed results of the final measurement model of neighborhood social and economic change: Type Membership.....	217
Appendix 4.1: Distribution of Neighborhood Change Types over time	218
Appendix 4.2: proportion of census sections that lost, gained or were stable in the number of food stores in each category by year.	219
Appendix 4.3: Neighborhood Change (averaged from 2007 to 2011) and gain/loss in food stores (2012 to 2016), unadjusted results	220
Appendix 4.4: Neighborhood Change (averaged from 2007 to 2011) and gain/loss in food stores (2012 to 2016), adjusted for baseline # of stores	221
Appendix 5.2. Type Membership after the application of the classification algorithm (averaged from 2006-2009) vs actual modal type membership from 2006 to 2009	223
Appendix 5.3. Association (OR, 95% CI) of Neighborhood Social and Economic Change and Diabetes Prevalence.....	224
Appendix 5.4. Association (HR, 95% CI) of Baseline Neighborhood Social and Economic Change and Diabetes Incidence (< 10 events excluded) ...	225

Appendix 5.5. Association (HR, 95% CI) of Baseline Neighborhood Social and Economic Change and Diabetes Incidence excluding people that changed health centers	226
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------

LIST OF FIGURES

Figure 1.1: Framework for the Study of Neighborhood Change and its association with the Food Environment and Chronic Diseases.....	21
Figure 2.1. Administrative Hierarchy of the Spanish Census	69
Figure 2.2. Exploratory Data Analysis for Aim 1: Trends in Change Indicators (raw value) from 2005 to 2015 in Madrid (Spain).....	70
Figure 2.3. Exploratory Data Analysis for Aim 1: Trends in Change Indicators (Delta operationalization) from 2005 to 2015 in Madrid (Spain)..	71
Figure 2.4. Exploratory Data Analysis for Aim 1: Comparison of Changes in Unemployment (left) vs Standardized Changes in Unemployment (Right) .	72
Figure 2.5. Exploratory Data Analysis for Aim 1: Trends in Raw Counts of Housing Variables from 2005 to 2015 in Madrid (Spain)	73
Figure 2.6. Exploratory Data Analysis for Aim 1: Trends in Mobility Variables from 2005 to 2015 in Madrid (Spain).....	74
Figure 2.7. Exploratory Data Analysis for Aim 1: Correlation Plot between All Change Indicators (no Delta).....	75

Figure 2.8. Exploratory Data Analysis for Aim 1: Correlation Plot between All Change Indicators (Delta)	76
Figure 2.9. Exploratory Data Analysis for Aim 1: Correlation Plot between All Mobility Indicators	77
Figure 2.10. Classification of all Commercial Spaces based on Economic Activity	78
Figure 2.11. Algorithm to classify food stores based on declared CNAE activities and store name.	79
Figure 2.12. Exploratory Data Analysis for Aim 2: Trends in the Number of Stores by Type.....	80
Figure 2.13. Exploratory Data Analysis for Aim 2: Trends in the Proportion of Food Stores by Type.....	81
Figure 2.14. Exploratory Data Analysis for Aim 2: Trends in the Average Number of Food Stores by Type per census section	82
Figure 2.15. Exploratory Data Analysis for Aim 2: Trends in the Median Number of Food Stores by Type per census section	83
Figure 2.16. Exploratory Data Analysis for Aim 3: Map of the Study Area within Madrid and the Health Centers included in the Study.....	84
Figure 2.17. Exploratory Data Analysis for Aim 3: Distribution of key sociodemographic and socioeconomic variables in the four districts as compared to the entire city of Madrid	85

Figure 2.18. Exploratory Data Analysis for Aim 3: Map of the empirical catchment areas of the centers in 2013	86
Figure 3.1: Characteristics of each type of neighborhood social and economic change.	120
Figure 3.2: Spatial Distribution of Neighborhood Social and Economic Change Types by Epoch (2005-2009 and 2010-2015)	121
Figure 3.3. Association of Neighborhood Social and Economic Change Types and Current Levels of Socioeconomic Status, Average Age, Proportion Foreign-Born and Population Density.	122
Figure 4.1: Description of Neighborhood Social and Economic Change Types	149
Figure 4.2. Algorithm to classify food stores based on declared CNAE activities and store name.	150
Figure 4.3: Trends in the Average number and proportion of each food store type by change type	151
Figure 5.1: Characteristics of each type of neighborhood social and economic change.	178
Figure 5.2. Kaplan-Meier Survival Curve of Diabetes Incidence by Neighborhood Social and Economic Change Type.	179
Figure 6.1: Theory Development.....	209

**Appendix 5.1: Comparison of the four study districts vs. the rest of Madrid
in education, country of origin, age, unemployment and property value..222**

CHAPTER 1: INTRODUCTION

Overview

This dissertation describes the process of developing a measure of neighborhood social and economic change and then determining its associations with changes in the food environment and diabetes incidence. This chapter introduces the importance of studying the macrosocial determinants of diabetes incidence, the challenges involved in studying dynamic urban environments as hypothesized determinants, and the role that the food environment plays in these associations.

Cardiovascular Disease and Diabetes

Epidemiology of CVD

Cardiovascular disease is the leading cause of death worldwide and one of the major causes of disability (Vos et al., 2016). This group of diseases includes coronary heart disease (acute myocardial infarction and other coronary syndromes, and congestive heart failure), cerebrovascular disease (stroke, including its hemorrhagic and ischemic types, and others), peripheral vascular disease, and other conditions such as cardiomyopathies, valvular heart disease and congenital heart disease (Marmot and Elliott, 2005). The burden of CVD is predicted to increase due to (1) an increase in prevalence originated from a decrease in the case lethality in industrialized countries, and (2) an increase in the incidence in non-industrialized and emerging economies (Beaglehole and Bonita, 2008).

After a consistent rise during the first half of the 20th century, cardiovascular disease mortality began a steep decline in the 1960s, 1970s and 1980s in most industrialized countries (Cooper et al., 1978). Nonetheless, CVD is still the leading cause of death in Europe overall (Nichols et al., 2014) and Spain (Franco et al., 2011a). Despite decreasing incidence rates, CVD is still the most prevalent cause of death in most developed countries and an enormous burden on health systems due to increased prevalence (because of decreased case fatality) and costs associated with it (Franco et al., 2011b).

Cardiovascular Risk Factors

Atherosclerosis is the main pathological finding in most cardiovascular diseases (namely coronary heart disease, ischemic stroke and peripheral vascular disease) (Marmot and Elliott, 2005). The determinants of atherosclerosis are therefore the causes of cardiovascular diseases. We have an incomplete understanding of the atherosclerotic process that leads to cardiovascular disease, but risk factors have been identified in the last 60 years (Blackburn and Pyorala, 2006), starting with the inception of the Framingham Heart Study. Those associated with behaviors (often referred to as ‘modifiable risk factors’) include high blood pressure (both systolic and diastolic blood pressure, grouped under ‘hypertension’), dyslipidemia (including high Low-density lipoproteins [LDL] and low high-density lipoproteins [HDL]), cigarette smoking and dysglycemia (or, in its most severe form, diabetes). Most of these risk factors are associated with obesity, some dietary patterns, and physical

inactivity. More recently identified risk factors include psychosocial conditions (such as depression or anxiety) (Everson-Rose and Lewis, 2005), inflammation (Pearson et al., 2003), and excessive alcohol consumption (Corrao et al., 2004). The importance of each risk factor varies, but previous studies have found that changes in average systolic blood pressure and hypertension control have accounted for a large proportion of the decline in cardiovascular disease in countries like Spain (Flores-Mateo et al., 2011). However, increases in diabetes prevalence may slow or even reverse the decreasing trend in cardiovascular disease mortality (Flores-Mateo et al., 2011).

Epidemiology of Diabetes

In the last forty years, the obesity epidemic has led to an increase in diabetes prevalence (West, 1978). In particular, diabetes trends in the US in the last 30 years are worrisome (Menke et al., 2015; Menke et al., 2014; Selvin et al., 2014). Estimates vary by method and definition of diabetes, but some methodologically sound reports predict an increase from a prevalence of 5.8% in the period 1998-1994 to a prevalence of 12.4% in the period 2005-2010 (Selvin et al., 2014). Few studies have explored these trends in Spain. In fact, to our knowledge, only one study has looked at national trends using directly measured markers of hyperglycemia: a very high prevalence of diabetes (around 13.8%) was reported, with approximately half of the cases undiagnosed (Soriguer et al., 2012).

These trends in diabetes are strongly linked to the obesity epidemic (Menke et al., 2014) and may be explained solely by increases in obesity (Menke et al., 2014). Tackling the determinants of obesity seems, therefore, the key to controlling diabetes. Studies such as the Diabetes Prevention Program (DPP) have shown that lifestyle modification leading to weight loss decreased diabetes incidence (Group, 2002). A natural experiment conducted in Cuba has also shown that a population-wide decrease in body weight led to a decrease in diabetes incidence, which was reversed following population-level weight gain (Franco et al., 2013). Individual and population level weight control through improvements in diet and physical activity are key to controlling diabetes.

Bending the Curve: Rose's Approach to Prevention

In order to understand the large differences over time and the wide variations in rates across countries, regions, or neighborhoods in cardiovascular disease or diabetes burden, we must first understand a crucial difference: the causes of individual diabetes cases may be different from the determinants of the incidence rate in a population (Rose, 1985). As Geoffrey Rose wrote, what makes two individual London civil servants differ in their systolic blood pressure (e.g., different behaviors related to physical activity) may be different than what makes average systolic blood pressure in London higher than the distribution in Kenyan nomads (e.g., different transportation options). For, "what distinguishes the two groups is nothing to do with the characteristics of individuals, it is rather a shift in the whole distribution – a mass influence acting in the population as a

whole. To find the determinants of prevalence and incidence rates, we need to study characteristics of populations, not characteristics of individuals” (Rose, 1985).

Motivation for studying mass influences

If we are to understand how to reduce the rates of cardiovascular disease or diabetes in countries like the US, UK, or Spain down to the level of Japan, France or Switzerland, we need to understand what makes these countries different from each other. If we are to bring the diabetes burden of poor neighborhoods closer to that of wealthier ones, we need to understand what factors drive these prevalence or incidence rates. If we are to prevent diabetes rates from continuing to increase in western economies or starting to increase in emerging economies, we need to understand what changed in the last few decades in these countries.

Cardiovascular diseases are not randomly distributed within populations. During the first half of the 20th century, they tended to be clustered among higher socioeconomic strata, probably due to smoking and increased availability of saturated fats (Marmot and Elliott, 2005; Rose et al., 2008). However, during the second half of the 20th century, an inverse social gradient started to emerge (Marmot et al., 1984; Rose and Marmot, 1981). Cardiovascular diseases and their determinants follow a clear social gradient in all developed nations: the lower the socioeconomic status of the individual, the higher the probability of

having a risk factor for or developing CVD (Marmot and Elliott, 2005; Rose et al., 2008).

As with other cardiovascular risk factors, there are also large disparities in diabetes burden by some social factors. In particular, individuals of lower socioeconomic status have higher rates of diabetes and lower rates of diabetes control (Agardh et al., 2011; Espelt et al., 2008). Moreover, there are large disparities in diabetes prevalence by country (Espelt et al., 2008), a region within a country (Espelt et al., 2008), and by neighborhood (Gary-Webb et al., 2013). In general, neighborhoods of lower socioeconomic status have been found to have higher rates of diabetes prevalence and incidence, even after adjusting for individual level socioeconomic status (Gary-Webb et al., 2013).

Motivation for studying cities and neighborhoods as mass influences

Cities and neighborhoods have many opportunities for the prevention of cardiovascular diseases and diabetes (Franco et al., 2015). The nature of urban areas, with a high density of people, services, and social relationships, creates the perfect environment for policy development, implementation, and evaluation (Franco et al., 2015). If we are to understand the inequities in diabetes incidence, prevalence, and control, we need to understand what makes certain neighborhoods have much higher rates than others. These mass influences across areas or time are the macrosocial determinants of health. According to Galea, these are “factors, such as culture, political systems, economics, and

processes of migration or urbanization, that are beyond the individual and are explicitly a function of population systems” (Galea, 2010).

Challenges in the Study of Neighborhoods

The study of inequalities in chronic diseases (including diabetes) in urban environments has grown exponentially in the last two decades (Diez Roux et al., 2016). This growth has been accompanied by the emergence of theoretical and empirical challenges in the study of the effects of neighborhoods on chronic diseases (Cummins et al., 2007; Diez Roux, 2007; Diez-Roux, 2003; Glass and Bilal, 2016). This dissertation focuses on one of these challenges: the dynamic nature of residential environments.

Neighborhood Change

Overview of theories

The formalization of theories to study neighborhood change took off during the second half of the 20th century. Grigsby (1987) laid the foundations for the study of neighborhood changes, based on metropolitan housing dynamics and theories of succession (similar to those being studied at the time in ecology). More recently, with the advent of new methodology and data sources, more comprehensive characterization and theoretical frameworks of neighborhood change have been developed (van Ham et al., 2012). These theories (pictured in Figure 1.1) include neighborhood selection theories (based on selective migration and residential mobility); demographic and socioeconomic change

theories (based on changes affecting residents of the neighborhood); and external shock theories (structural changes in the labor and housing markets or changes in urban policies). According to van Ham, the relationships predicted by these theories exist in neighborhoods at the same time and may help to explain the different phenomena associated with neighborhood change. Studies that rely exclusively on cross-sectional data may be naïve to these drivers of neighborhood change.

Neighborhood Selection theories

The most well-studied theories of neighborhood change rely on neighborhood selection as the main driver. In a recent review of the literature on neighborhood selection, Bailey et al. (2013) divided forces leading to selection into two broad categories: residential mobility (relating to overall flows of people in and out of neighborhoods) and selective migration (focusing on differential patterns of mobility by demographic or socioeconomic group). Regarding residential mobility, Bailey et al. (2013) find that the best predictors of mobility decisions are neighborhood perceptions (both satisfaction and subjective characteristics of the area; Bailey et al. explicitly state that objective neighborhood characteristics are very poor predictors of mobility decisions) and neighborhood change (which is a better predictor than current neighborhood composition). Regarding selective migration, where the literature is more scarce, Bailey et al. conclude that the main driver of differential behaviors of mobility is the life-course stage of the individual, and explicitly lay out a theory of

“demographic conveyors” where age distributes individuals across the city (i.e., young adults to poor neighborhoods and middle-aged adults to richer ones, while the elderly tends to be more stationary). Importantly, they highlight the importance of changes in non-movers (stayers) in determining neighborhood change, through natural growth (differential fertility) and socioeconomic change (e.g., changes in employment status). The most relevant conclusion of the review is that researchers should pay greater attention to “neighborhood dynamics and flows. Not only do people flow through places receiving varying durations of dose, places change too and change through different processes or flows” (Bailey et al., 2013).

External Shock theories

Aalbers (2013) developed a theory of neighborhood change based on theoretical notions by Henri Lefebvre and others. Of importance here is the differentiation between social space (that which acquires meaning through the interactions between neighbors) and abstract space (that which is used as a tool to reproduce a given societal structure). Aalbers focuses on two specific practices of the abstract space plane: redlining and predatory lending. He argues that these two practices, while opposed regarding their nature (one excludes people from the mortgage market, the other over includes people in very disadvantaged conditions), are two historical phases of the same manipulation of abstract space. Aalbers argues against “natural neighborhood change” theories

(which mostly are composed of neighborhood selection theories) and argues for measuring the impact of abstract space makers.

Drivers vs. Passengers: a neighborhood approach

This detailed elaboration of theories follows a specific purpose: to identify the 'drivers' and 'passengers' of neighborhood socioeconomic change. The reasons behind this differentiation are clear: interventions on passengers do shift system behavior in the desired direction, given the lack of importance of passengers in deciding the direction or speed of neighborhood change. On the other hand, policy interventions on drivers may shift the distribution of risk factors in a healthier direction more efficiently and more robustly. This idea of separating drivers vs. passengers is a driving force of this dissertation.

Chapter 3 of this dissertation elaborates on how these different theories impacted the selection of data sources, indicators and statistical models. Specifically, neighborhood selection theories, both through residential mobility (movers) and immobility (stayers), highlight the role that socioeconomic and sociodemographic characteristics play in determining who moves and who stays. External shocks theories highlight the importance of looking beyond residential mobility/immobility in understanding change, and point towards the role of real estate developers and local/regional governments in effecting neighborhood change.

The Local Food Environment

This dissertation focuses on a specific social mechanism through which place-based stratification (neighborhood inequalities) affects diet and diabetes: changes in the food environment following changes in neighborhoods. Population dietary patterns are shaped by mass-influences that differ across populations or within the same population over time (Díez et al., 2016; Rose, 1985). As a contextual factor, the Local Food Environment (LFE) affects everyone living in an area and is not a quality of a single individual, and therefore qualifies as a potential mass-influence on diet (Rose, 1985). The LFE is defined as the set of contextual aspects of the local environment that have the potential to influence dietary behaviors (Franco et al., 2016). The components of the local food environment include the location and accessibility of food stores and the availability of healthy foods within them (Glanz et al., 2005, 2007). Changes in these factors have the potential to affect population dietary patterns so understanding what causes changes in food stores (and their content) may be a feasible way to improve diet (Story et al., 2008).

Distribution of local food environments

The shape and distribution of local food environments in Europe are different from those in the US (Black et al., 2014), where socioeconomic gradients are more evident in cross-sectional studies. In Spain, according to preliminary research, the presence of stores carrying healthy products (fresh produce) is more common than in the US and does not seem to follow the same

cross-sectional socioeconomic gradient in their distribution (Bilal et al., 2016; Díez et al., 2016). Moreover, two qualitative studies have found that residents in different areas of Madrid (Spain) value small food stores above supermarkets to buy healthy foods (Bilal et al., 2016; Díez et al., 2017). In particular, residents of a medium SES area of Madrid highlighted the social role that small food stores play, as food retailers are places where social bonds are created, and trust in food retailing is maintained (Bilal et al., 2016). Residents of a lower SES area viewed small stores, as opposed to supermarkets, as positive places in their perceived food environment (Díez et al., 2017). These preliminary findings highlight the differences between previous research in the US and potential future research in Spain. In particular, research conducted in the UK has already highlighted these differential patterns in the distribution of the LFE, as it seems that this distribution in the UK differs widely from that of the US (Cummins, 2007; Cummins and Macintyre, 2006).

Neighborhood Change and the Food Environment

A component that is usually missing from food environment studies is an understanding of its dynamics. Most research focuses exclusively on cross-sectional associations or on longitudinal studies with time-fixed food environment features. There is, therefore, a need for studies that look into food environment dynamics, determinants and its consequences. There is strong evidence for a difference in how food environments change by levels of baseline neighborhood characteristics (Cobb et al., 2015; Rummo et al., 2016a; Rummo et al., 2016b). It

is, therefore, to be expected that changes in neighborhood characteristics may induce changes in the food environment. In fact, some research examining changes in socioeconomic characteristics in the CARDIA study has already shown hints of this (Richardson et al., 2014). Nonetheless, this research lacks a study of neighborhood characteristics beyond socioeconomic factors and race. For example, no consideration is given to the role of new housing, changes in housing or changes in zoning regulations. In essence, from the neighborhood change theories outlined in the previous version, only neighborhood selection is considered in the public health literature of food environment changes, and very few research looks at the consequences of external shocks.

The theories of neighborhood change outlined above predict different effects on the distribution and type of food stores. For example, related to neighborhood selection theories, studies have found that availability and affordability of healthy foods can be decoupled in neighborhoods that are undergoing gentrification, forming “Food Mirages” (Breyer and Voss-Andreae, 2013): areas where healthy food is plentiful but not affordable for some neighborhood residents (the non-movers). Urban policies can also have large effects on the food environment (as predicted by the external shock theory): a study in Canada found that the most predictive factors of food environment disparities were urban policies in place in different neighborhoods (Black et al., 2011).

No research, to our knowledge, has looked at the effect of neighborhood change on the local food environment in Spain. Spanish local food environments have unique characteristics shared by other countries in Southern Europe (Italy, France, and Portugal) that differ from those where most research on food environments has been done (mostly US, Australia and the United Kingdom). First, Spanish (and other Southern European) cities are more compact (González Pérez, 2007), with reduced levels of urban sprawl compared to Anglo-Saxon cities. This urban form creates more dense neighborhoods regarding residents and business, increasing food availability and making comparisons with those other countries harder (González Pérez, 2007). Recent changes to Spanish zoning regulations (González Pérez, 2007) have increased the levels of urban sprawl by creating suburban residential areas (in the image of US or UK suburbs) where food availability may be reduced (Munoz, 2003).

Second, Spanish (and other Southern European) local food environments, compared to Northern and Central European countries, are dominated by small retailers (Flavián et al., 2002). The number of outlets per resident is three times higher in Spain, Italy, and Portugal than in the UK, Finland, Denmark, and Belgium (Flavián et al., 2002). Compared to Northern and Central Europe, the market share of the top retailers is reduced in Southern European Countries (Flavián et al., 2002) and so is the average number of supermarket or shopping malls per resident. In the case of Spain, this is related to two factors: (a) the availability of a transportation network that is especially dense in dense

neighborhoods and city centers (as opposed to suburbs where large food retailers may open) (Castillo-Manzano and López-Valpuesta, 2009); and (b) the presence of small business owners' lobbies that have guaranteed protective regulatory mechanisms related to the opening of large food retailers (Flavián et al., 2002).

In summary, Spanish neighborhoods are denser and have a food environment dominated by small businesses, compared to other countries. Moreover, as shown in our qualitative research (Bilal et al., 2016; Díez et al., 2017), neighbors tend to prefer food acquisition in small stores rooted in the neighborhood history (the "lifetime store"). However, this situation is different for new developments in Madrid, characterized by lower residential density (González Pérez, 2007) and without an established network of small food retailers (Cornejo Nieto et al., 2010). Therefore, it seems that neighborhoods growing in terms of housing may have reduced food availability compared to older, denser neighborhoods.

A second aspect that may affect the presence of food stores is the availability of rental space. Fruit and vegetable stores, small fishmongers, butcheries, and bakeries make up a large percent of the market share of food retail in Spain (Puelles et al., 2011). As mentioned above, these stores are small businesses with no link to large retailers.

Small business with no link to large retailers have been shown before to be the most heavily affected by changes in property value (Zukin et al., 2009). As

property values go up, large retailers tend to take over, and small business owners are driven out of the area (Zukin et al., 2009). Given differential increases in property values in Madrid over the study period, it follows that the effect of these increases on small stores carrying healthy products will be disproportionate compared to large food retailers. Therefore, neighborhoods with large housing growth (because of increased demand and increased property value) may see the diminished availability of healthy foods.

Evidence Gaps in the Literature

A few studies have examined neighborhood social and economic change and CVD or its risk factors. As mentioned above, some studies in MESA (Hirsch et al., 2014a; Hirsch et al., 2016a; Hirsch et al., 2016b; Hirsch et al., 2014b; Hirsch et al., 2014c) and CARDIA (Hirsch et al., 2016b; Richardson et al., 2014; Richardson et al., 2015; Rummo et al., 2016a; Rummo et al., 2016b) have looked change in some specific socioeconomic or sociodemographic characteristics and changes in metabolic risk factors or its contextual determinants. The gentrification literature has also some examples of metabolic outcomes associated with the specific process of gentrification. For example, a study looking into type 1 and type 2 diabetes in Chicago (Grigsby-Toussaint et al., 2010), found that “non-type 1 Diabetes” risk was increased in “Emerging High Income” neighborhoods, but this association was only significant in Hispanic children. A study in Portland (Oregon), found that neighborhoods undergoing gentrification are at risk of becoming “food mirages” where the population is

alienated from its food environment making healthy foods unaffordable for its residents (Breyer and Voss-Andreae, 2013). A qualitative analysis in Chicago found that lower-income groups and minorities are heavily dependent on community fabric (and its subsequent social capital), disrupted during gentrification processes (Betancur, 2011). A qualitative analysis in Montreal (Burns et al., 2012) showed that gentrification might be linked to social exclusion in older adults, causing disconnectedness and loss of influence. Moreover, a study in New York City has shown that the effects of gentrification on health (in this case, preterm birth) may be different for minorities and may not be harmful to privileged majorities (Huynh and Maroko, 2014). Few studies have looked into gentrification, neighborhood economic change and health in Europe (Bacque et al., 2011; Ward et al., 2010) where the dynamics of neighborhood change may differ widely from American cities (Kazepov, 2005).

There is a lack of studies looking at neighborhood social and economic change as a process emerging from both neighborhood selection phenomena and the actions of external shocks. The role of new housing, housing renovations and property value changes must be explicitly studied, especially as they relate to other components of neighborhood change. Among the consequences of neighborhood change, as detailed above, may be negative or positive changes in the food environment and its downstream dietary consequences, such as diabetes. This dissertation tries to fill these gaps and offer a more integral understanding of neighborhood change and its consequences.

Specific Aims and Hypothesis

With the concerns and ideas outlined above in mind, we set to measure neighborhood social and economic change and study its associations with food environment changes and diabetes incidence. Specifically, the aims of this dissertation were to:

AIM 1: To create a measurement model of neighborhood social and economic change in the city of Madrid (Spain) from 2005 to 2015

AIM 2: To evaluate the association between neighborhood social and economic change with changes in the retail food environment in the city of Madrid (Spain).

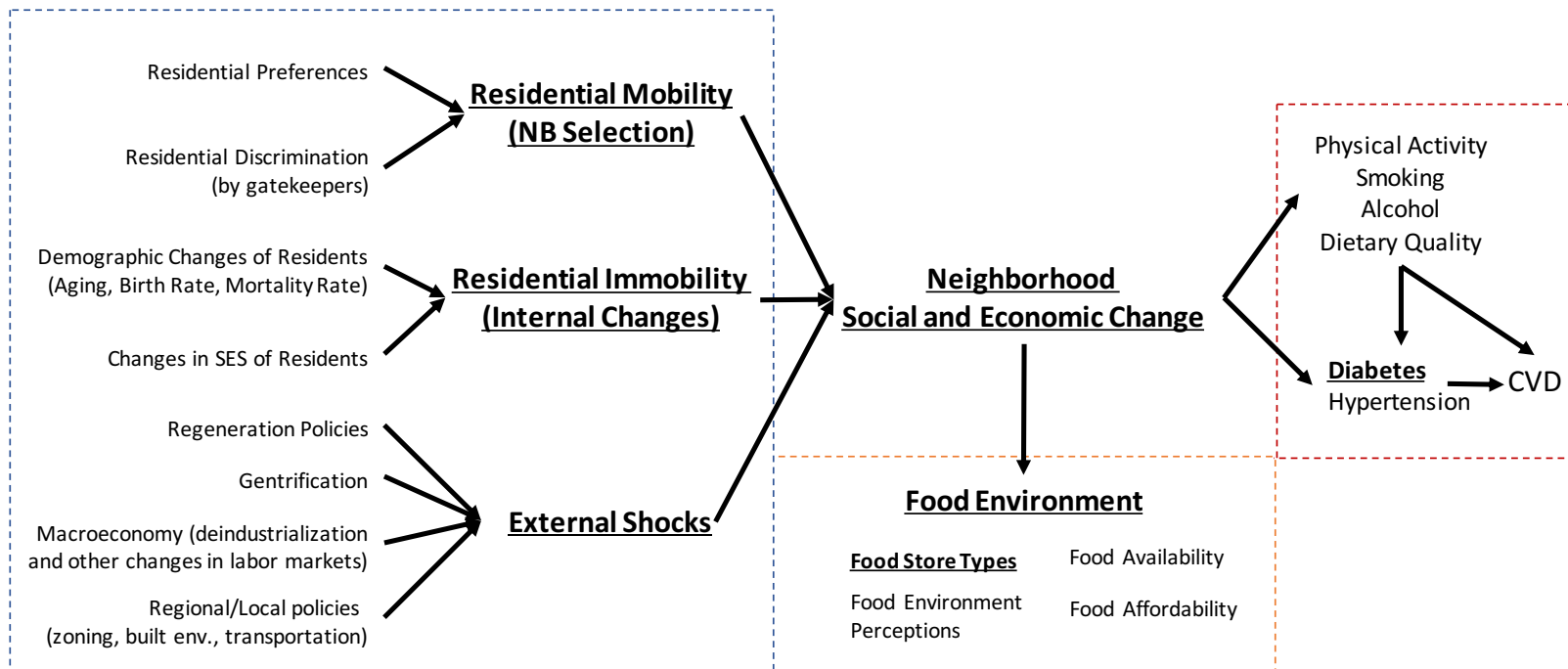
AIM 3: To evaluate the association between neighborhood social and economic change and diabetes incidence in an area of the city of Madrid (Spain).

Organization of the Dissertation

This dissertation is organized into six chapters: Chapter 1 (this chapter) summarizes the justification for the study of neighborhood change in social epidemiology; Chapter 2 offers a more detailed explanation of several methodological challenges inherent to the study aims and summarizes some of the exploratory data analysis conducted prior to each of the following 3 aims; Chapter 3 describes the measurement model of neighborhood social and economic change; Chapter 4 explores the association between neighborhood social and economic change and food environment changes; Chapter 5 studies the association between neighborhood social and economic change and diabetes incidence; Chapter 6 provides an integrative summary of the findings, a discussion of the challenges encountered in the conduct of this dissertation and paths forward.

Figures

Figure 1.1: Framework for the Study of Neighborhood Change and its association with the Food Environment and Chronic Diseases



References

- Aalbers, M.B. (2013). How Do Mortgage Lenders Influence Neighbourhood Dynamics? Redlining and Predatory Lending. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research* pp. 63-85). Dordrecht: Springer Netherlands.
- Agardh, E., Allebeck, P., Hallqvist, J., Moradi, T., Sidorchuk, A. (2011). Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. *International Journal of Epidemiology*. 40 (3), 804-818.
- Bacque, M.H., Fijalkow, Y., Launay, L., Vermeersch, S. (2011). Social mix policies in Paris: discourses, policies and social effects. *International journal of urban and regional research*. 35 (2), 256-273.
- Bailey, N., Barnes, H., Livingston, M., Mclennan, D. (2013). Understanding Neighbourhood Population Dynamics for Neighbourhood Effects Research: A Review of Recent Evidence and Data Source Developments. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research* pp. 23-41). Dordrecht: Springer Netherlands.
- Beaglehole, R., Bonita, R. (2008). Global public health: a scorecard. *Lancet*. 372 (9654), 1988-1996.
- Betancur, J. (2011). Gentrification and Community Fabric in Chicago. *Urban Studies*. 48 (2), 383-406.
- Bilal, U., Diez, J., Alfayate, S., et al. (2016). Population cardiovascular health and urban environments: the Heart Healthy Hoods exploratory study in Madrid, Spain. *BMC medical research methodology*. 16, 104.
- Black, C., Moon, G., Baird, J. (2014). Dietary inequalities: what is the evidence for the effect of the neighbourhood food environment? *Health & place*. 27, 229-242.
- Black, J.L., Carpiano, R.M., Fleming, S., Lauster, N. (2011). Exploring the distribution of food stores in British Columbia: Associations with neighbourhood socio-demographic factors and urban form. *Health & Place*. 17 (4), 961-970.
- Blackburn, H., Pyorala, K. (2006). *Preventing Heart Attacks and Stroke: A History of the Research*: Oxford University Press.
- Breyer, B., Voss-Andreae, A. (2013). Food mirages: geographic and economic barriers to healthful food access in Portland, Oregon. *Health Place*. 24, 131-139.
- Burns, V.F., Lavoie, J.P., Rose, D. (2012). Revisiting the role of neighbourhood change in social exclusion and inclusion of older people. *Journal of aging research*. 2012, 148287.

- Castillo-Manzano, J.I., López-Valpueda, L. (2009). Urban retail fabric and the metro: A complex relationship. Lessons from middle-sized Spanish cities. *Cities*. 26 (3), 141-147.
- Cobb, L.K., Anderson, C.A., Appel, L., et al. (2015). Baltimore City Stores Increased The Availability Of Healthy Food After WIC Policy Change. *Health Affairs*. 34 (11), 1849-1857.
- Cooper, R., Stamler, J., Dyer, A., Garside, D. (1978). The decline in mortality from Coronary Heart Disease, U.S.A., 1968–1975. *Journal of Chronic Diseases*. 31 (12), 709-720.
- Cornejo Nieto, C., Moran Saez, J., Prada Trigo, J. (2010). *Ciudad, territorio y paisaje*. Madrid: CSIC.
- Corrao, G., Bagnardi, V., Zambon, A., La Vecchia, C. (2004). A meta-analysis of alcohol consumption and the risk of 15 diseases. *Preventive medicine*. 38 (5), 613-619.
- Cummins, S. (2007). Commentary: Investigating neighbourhood effects on health—avoiding the ‘Local Trap’. *International Journal of Epidemiology*. 36 (2), 355-357.
- Cummins, S., Curtis, S., Diez-Roux, A.V., Macintyre, S. (2007). Understanding and representing 'place' in health research: a relational approach. *Soc Sci Med*. 65 (9), 1825-1838.
- Cummins, S., Macintyre, S. (2006). Food environments and obesity - neighborhood or nation? *Int J Epidemiol*. 25,
- Díez, J., Bilal, U., Cebrecos, A., 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., 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.
- Diez Roux, A.V. (2007). Neighborhoods and health: where are we and where do we go from here? *Revue d'Épidémiologie et de Santé Publique*. 55 (1), 13-21.
- Diez Roux, A.V., Mujahid, M.S., Hirsch, J.A., Moore, K., Moore, L.V. (2016). The Impact of Neighborhoods on CV Risk. *Global Heart*. 11 (3), 353-363.
- Diez-Roux, A.V. (2003). Residential environments and cardiovascular risk. *Journal of Urban Health*. 80 (4), 569-589.
- Espelt, A., Borrell, C., Roskam, A.J., et al. (2008). Socioeconomic inequalities in diabetes mellitus across Europe at the beginning of the 21st century. *Diabetologia*. 51 (11), 1971.
- Everson-Rose, S.A., Lewis, T.T. (2005). Psychosocial factors and cardiovascular diseases. *Annu. Rev. Public Health*. 26, 469-500.
- Flavián, C., Haberberg, A., Polo, Y. (2002). Food retailing strategies in the European Union. A comparative analysis in the UK and Spain. *Journal of Retailing and Consumer Services*. 9 (3), 125-138.

- Flores-Mateo, G., Grau, M., O'Flaherty, M., et al. (2011). Analyzing the Coronary Heart Disease Mortality Decline in a Mediterranean Population: Spain 1988-2005. *Revista Española de Cardiología (English Version)*. 64 (11), 988-996.
- Franco, M., Bilal, U., Díez, J. (2016). Food Environment. In B. Caballero, P.M. Finglas, & F. Toldrá (Eds.), *Encyclopedia of Food and Health* pp. 22-26). Oxford: Academic Press.
- Franco, M., Bilal, U., Díez-Roux, A.V. (2015). Preventing non-communicable diseases through structural changes in urban environments. *J Epidemiol Community Health*. 69 (6), 509-511.
- Franco, M., Bilal, U., Ordunez, P., et al. (2013). Population-wide weight loss and regain in relation to diabetes burden and cardiovascular mortality in Cuba 1980-2010: repeated cross sectional surveys and ecological comparison of secular trends. *Bmj*. 346, f1515.
- Franco, M., Cooper, R., Bilal, U., Fuster, V. (2011a). Control of coronary heart disease risk factors and evidence-based therapies: joint efforts for coronary heart disease prevention in Spain. *Revista española de cardiología*. 64 (11), 962-964.
- Franco, M., Cooper, R.S., Bilal, U., Fuster, V. (2011b). Challenges and opportunities for cardiovascular disease prevention. *The American journal of medicine*. 124 (2), 95-102.
- Galea, S. (2010). *Macrosocial Determinants of Population Health*: Springer New York.
- Gary-Webb, T.L., Suglia, S.F., Tehranifar, P. (2013). Social Epidemiology of Diabetes and Associated Conditions. *Current Diabetes Reports*. 13 (6), 850-859.
- Glanz, K., Sallis, J.F., Saelens, B.E., Frank, L.D. (2005). Healthy nutrition environments: concepts and measures. *Am J Health Promot*. 19,
- Glanz, K., Sallis, J.F., Saelens, B.E., Frank, L.D. (2007). Nutrition Environment Measures Survey in stores (NEMS-S): development and evaluation. *American journal of preventive medicine*. 32,
- Glass, T.A., Bilal, U. (2016). Are neighborhoods causal? Complications arising from the 'stickiness' of ZNA. *Soc Sci Med*,
- González Pérez, J.M. (2007). Urban Planning System in Contemporary Spain. *European Planning Studies*. 15 (1), 29-50.
- Grigsby, W.G. (1987). *The Dynamics of Neighborhood Change and Decline*: Pergamon Journals.
- Grigsby-Toussaint, D.S., Lipton, R., Chavez, N., Handler, A., Johnson, T.P., Kubo, J. (2010). Neighborhood Socioeconomic Change and Diabetes Risk: Findings from the Chicago Childhood Diabetes Registry. *Diabetes Care*. 33 (5), 1065-1068.
- Group, D.P.P.R. (2002). Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin. *New England Journal of Medicine*. 346 (6), 393-403.

- Hirsch, J.A., Diez Roux, A.V., Moore, K.A., Evenson, K.R., Rodriguez, D.A. (2014a). Change in Walking and Body Mass Index Following Residential Relocation: The Multi-Ethnic Study of Atherosclerosis. *American journal of public health*. 104 (3), e49-e56.
- Hirsch, J.A., Grengs, J., Schulz, A., et al. (2016a). How much are built environments changing, and where?: Patterns of change by neighborhood sociodemographic characteristics across seven U.S. metropolitan areas. *Social Science & Medicine*. 169, 97-105.
- Hirsch, J.A., Meyer, K.A., Peterson, M., et al. (2016b). Obtaining Longitudinal Built Environment Data Retrospectively across 25 years in Four US Cities. *Frontiers in Public Health*. 4 (65),
- Hirsch, J.A., Moore, K.A., Barrientos-Gutierrez, T., et al. (2014b). Built environment change and change in BMI and waist circumference: Multi-ethnic Study of Atherosclerosis. *Obesity*. 22 (11), 2450-2457.
- Hirsch, J.A., Moore, K.A., Clarke, P.J., et al. (2014c). Changes in the Built Environment and Changes in the Amount of Walking Over Time: Longitudinal Results From the Multi-Ethnic Study of Atherosclerosis. *American Journal of Epidemiology*. 180 (8), 799-809.
- Huynh, M., Maroko, A.R. (2014). Gentrification and preterm birth in New York City, 2008-2010. *J Urban Health*. 91 (1), 211-220.
- Kazepov, Y. (2005). *Cities of Europe: Changing Contexts, Local Arrangement and the Challenge to Urban Cohesion*: Wiley.
- Marmot, M.G., Elliott, P. (2005). *Coronary Heart Disease Epidemiology: From Aetiology to Public Health*: Oxford University Press.
- Marmot, M.G., Shipley, M.J., Rose, G. (1984). Inequalities in Death - Specific Explanations of a General Pattern? *The Lancet*. 323 (8384), 1003-1006.
- Menke, A., Casagrande, S., Geiss, L., Cowie, C.C. (2015). Prevalence of and trends in diabetes among adults in the united states, 1988-2012. *JAMA : the journal of the American Medical Association*. 314 (10), 1021-1029.
- Menke, A., Rust, K.F., Fradkin, J., Cheng, Y.J., Cowie, C.C. (2014). Associations between trends in race/ethnicity, aging, and body mass index with diabetes prevalence in the united states: A series of cross-sectional studies. *Annals of Internal Medicine*. 161 (5), 328-335.
- Munoz, F. (2003). Lock living: Urban sprawl in Mediterranean cities. *Cities*. 20 (6), 381-385.
- Nichols, M., Townsend, N., Scarborough, P., Rayner, M. (2014). Cardiovascular disease in Europe 2014: epidemiological update. *Eur Heart J*,
- Pearson, T.A., Mensah, G.A., Alexander, R.W., et al. (2003). Markers of Inflammation and Cardiovascular Disease. Application to Clinical and Public Health Practice: A Statement for Healthcare Professionals From the Centers for Disease Control and Prevention and the American Heart Association. 107 (3), 499-511.

- Puelles, M., Puelles, J., Romero, S. (2011). The Retail Industry in Spain. In D. Morschett, T. Rudolph, P. Schnedlitz, H. Schramm-Klein, & B. Swoboda (Eds.), *European Retail Research* pp. 129-166): Gabler Verlag.
- Richardson, A.S., Meyer, K.A., Howard, A.G., et al. (2014). Neighborhood socioeconomic status and food environment: A 20-year longitudinal latent class analysis among CARDIA participants. *Health & Place*. 30, 145-153.
- Richardson, A.S., Meyer, K.A., Howard, A.G., et al. (2015). Multiple pathways from the neighborhood food environment to increased body mass index through dietary behaviors: A structural equation-based analysis in the CARDIA study. *Health & Place*. 36, 74-87.
- Rose, G. (1985). Sick individuals and sick populations. *International journal of epidemiology*. 14 (1), 32-38.
- Rose, G., Marmot, M.G. (1981). Social class and coronary heart disease. *British Heart Journal*. 45 (1), 13-19.
- Rose, G.A., Khaw, K.T., Marmot, M.G. (2008). *Rose's Strategy of Preventive Medicine: The Complete Original Text*: Oxford University Press.
- Rummo, P.E., Guilkey, D.K., Shikany, J.M., Reis, J.P., Gordon-Larsen, P. (2016a). How do individual-level sociodemographics and neighbourhood-level characteristics influence residential location behaviour in the context of the food and built environment? Findings from 25 years of follow-up in the CARDIA Study. *Journal of Epidemiology and Community Health*,
- Rummo, P.E., Hirsch, J.A., Howard, A.G., Gordon-Larsen, P. (2016b). In Which Neighborhoods Are Older Adult Populations Expanding? Sociodemographic and Built Environment Characteristics Across Neighborhood Trajectory Classes of Older Adult Populations in Four U.S. Cities Over 30 Years. *Gerontology and Geriatric Medicine*. 2,
- Selvin, E., Parrinello, C.M., Sacks, D.B., Coresh, J. (2014). Trends in prevalence and control of diabetes in the united states, 1988–1994 and 1999–2010. *Annals of Internal Medicine*. 160 (8), 517-525.
- Soriguer, F., Goday, A., Bosch-Comas, A., et al. (2012). Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Di@bet.es Study. *Diabetologia*. 55 (1), 88-93.
- Story, M., Kaphingst, K.M., Robinson -O'Brien, R., Glanz, K. (2008). Creating healthy food and eating environments: policy and environmental approaches. *Annual Reviews of Public Health*. 29,
- van Ham, M., Manley, D., Bailey, N., Simpson, L., Maclennan, D. (2012). *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research*: Springer Netherlands.
- Vos, T., Allen, C., Arora, M., et al. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*. 388 (10053), 1545-1602.

- Ward, K., Fagan, C., McDowell, L., Perrons, D., Ray, K. (2010). Class Transformation and Work-Life Balance in Urban Britain: The Case of Manchester. *Urban Studies*. 47 (11), 2259-2278.
- West, K.M. (1978). *Epidemiology of diabetes and its vascular lesions*: Elsevier.
- Zukin, S., Trujillo, V., Frase, P., Jackson, D., Recuber, T., Walker, A. (2009). New Retail Capital and Neighborhood Change: Boutiques and Gentrification in New York City. *City & Community*. 8 (1), 47-64.

CHAPTER 2: MATERIALS AND METHODS

Introduction and Overview

The purpose of this chapter is to provide details regarding the methods employed in this dissertation. This chapter is divided into four sections. The first section discusses the overarching challenge of selecting the most appropriate spatial unit of analysis. The following three sections describe detailed methods and exploratory data analysis for each of the three aims of this dissertation. The details included in this chapter are of importance to completely replicate the analysis conducted in Chapters 3, 4 and 5, but the lack of inclusion of these details in each chapter should not jeopardize the reader's capability to understand them independently. Moreover, the details included in this chapter do not require the reader to be acquainted with each of the following chapters before reading this chapter.

Spatial Units of Analysis

This section will discuss three specific methodological challenges related to problems related to the selection of, operationalization and measurement related to the spatial unit of analysis of this dissertation: (1) the hierarchical structure of administrative divisions in Spain; (2) the fluid census section problem; and (3) the spatially explicit organization of the healthcare system in Spain.

The Hierarchical Structure of Administrative Divisions in Spain

The modifiable areal unit problem (MAUP) is a well-known issue in all studies involving spatial data (Fotheringham and Wong, 1991). This problem emerges when the inferences drawn from a study are not robust to the selection of a spatial level of analysis. For example, if we are to study phenomena occurring at the smallest geographical level, and instead we use data on larger city areas much larger than those neighborhoods (such as neighborhood clusters, police jurisdictions or other administrative boundaries), our inferences may be biased compared to using small areas. Apart from the solution of having very precise and granular spatial data (e.g., precise geocoded data down to an individual or household level), an approach to deal with the MAUP is to understand how administrative areas are constructed and to have a sense of the variability within them (Eagleson et al., 2003, 2002).

Administrative Spatial Organization of Madrid

Figure 2.1 shows the spatial organization of the Spanish census. The Spanish census divides the country into several hierarchical levels. Starting with the Autonomous Region (*Comunidad Autonoma*), where legislative and executive power resides at the regional level; followed by the province (*Provincia*), where some specific executive powers are (Note: some Autonomous Regions like Madrid are mono-provincial, that is, are made of only one province and therefore this level does not exist); followed by the municipality (*Municipio*) where all local power resides. Municipalities are then divided into Districts

(*Distritos*) and these are in turn divided into the basic census area, called the census section (*seccion censal*). The census section is equivalent in population to the US Census Block Group, and hosts around 1500 people each. Most census data and other administrative data are available at the census section level.

The organization of Madrid differs slightly, in that the 21 Districts have some executive autonomy through the local district councils (*Juntas Municipales de Distrito*). These districts are then divided into 128 neighborhoods (*Barrios*), which in turn are divided into a number of census sections that varies year to year (~2400 from 2005 to 2015). While most data are available at the census section level, some data sources only release information at the neighborhood or district level. To re-emphasize, these three levels (district, neighborhood, census section) follow a strict, nested hierarchical structure. Census sections belong to a single neighborhood that in turn belongs to a single district. Last, in the context of health care delivery, it is important to mention that census sections are also nested into a spatially explicit Basic Health Area (which, in turn, is not perfectly nested into neighborhoods or districts). See Figure 2.1 for an intuitive hierarchy of the Spanish administrative divisions, and Table 2.1 for details on the size and population of each division.

The “Fluid Census Section” problem

Census sections are designed and delineated for electoral purposes. The main objective of the spatial organization through census sections is to assign

individuals to the local electoral colleges where people vote in each election, regardless of the level (European, National, Regional or Municipal). Each city government maintains a continuous census of the entire population in a municipal registry called *Padron*. This registry is then compiled by the National Bureau of Statistics (*INE* or *Instituto Nacional de Estadística*). The INE then uses this information to delineate and release a new set of census sections every year. This new delineation includes every street and street number and street side allowing local urbanism offices to generate census section borders.

The method through which the INE delineates census sections is complex, but the goal is to keep changes to a minimum and maintain the population at around 1500 individuals of all ages per census section. The most common changes are as follows:

- Census Section Merging: when census sections lose population, two adjacent census sections may be merged into one. The ID of the new census section corresponds to one of the previous census sections. For example, if census section 2807915001 merges with census section 2807915002, the new census section may adopt the ID 2807915001 (or 002).
- Census section Splitting: when census sections gain too much population they may split into two (or more) census sections. One of the resulting census sections will keep the ID of the original census section. For example, if census section 2807915010 splits

in two census sections, one of them will keep the 2807915010 ID while the other resulting section will be assigned a unique ID not used before.

Between 2005 and 2015 these two changes occurred in approximately 1% of the census sections (see Table 2.2 for more details). In addition, there may also be small adjustments to census section borders, including small changes to the street numbers included in each census section. These changes make the use of areal densities complicated, since a change in the area can lead to a spurious change in the density (due to a change in the denominator).

Ignoring these issues would lead to measurement error if using longitudinal data. For example, if a census section is split in 2006, and this split results in two very different areas, it may seem like one of the census section has intensely changed in a year, while this would be just the result of an administrative change. Several techniques are available to deal with this issue:

1. Ignoring changing census section: keep only census sections that do not change. While for some applications this approach may be appropriate, for this dissertation in which the objective is to understand the effect of neighborhood change on health, these census sections with changes are of great relevance. These sections have undergone the most intense change and removing them would lead to a severe underestimation of neighborhood change intensity in Madrid.

2. Rasterization: this option would involve moving to a different spatial paradigm, where spatial attributes would be assigned to fixed areas through a raster instead of vectorial areas (polygons). A problem with this approach lies in deciding how to best assign attributes to cells based on larger units, which can then lead to a geographical ecological fallacy.
3. Generating common ancestors or descendants of census sections: this approach assigns each split census sections the ID of the ancestor census section, effectively collapsing the two sections into one and re-calculating all attributes based on this new ID. The same procedure applies to merged census sections, but backwards in time, as past census sections that will be merged in the future get assigned the id of their future descendant. While this approach may wash down some changes, especially if the resulting census sections derived from splits are very different, it does so in a much less severe way than just ignoring changing sections. Compared to rasterizing the study area, this approach is simpler.

We used the latter approach (3) and generated a common ancestor for all census sections from 2005 to 2015. The number of “common” census sections was therefore reduced to 2272 (from between 2360 to 2420 regular census sections).

The spatially explicit organization of the health system

As mentioned above, there is a second complexity in dealing with Spanish Health Care data, as this kind of data has its own spatial organization. The Spanish Health Care System is organized into regional administrations. These have autonomy over their own organization, and may include a single geographical area or divisions into certain geographical boundaries, which are then divided into Basic Health Areas (*Zonas Basicas de Salud*). In theory, according to the 1986 LGS (*Ley General de Salud*, the Law that created the National Health System in Spain), each of these Basic Health Areas represent the catchment area for a single Primary Care Center (*Centro de Salud*). They are also the basic unit of organization of the health care delivery system, and may include other services beyond primary care doctors and nurses (e.g., social workers, occupational therapy, etc.).

In Madrid, from 2002 through 2010, the *Servicio Madrileño de Salud* (*SERMAS*) was organized in 11 Health Areas (*Areas de Salud*) for administrative purposes. Each of these 11 Health Areas, in turn, were organized (according to the 1986 LGS) in 242 Basic Health Areas (*Zonas Basicas de Salud*) and each person was assigned to the Primary Care Center whose catchment area overlapped with their residence. From 2010 onwards, the 11 health areas were merged into a single area (*Area Unica de Salud*), still divided into Basic Health Areas, but allowing each individual to freely decide his/her doctor and primary care center. In spite of these changes, some features of the regional health

system remain; of particular relevance to this dissertation the Electronic Health Records system has continued with the same structure. Nonetheless, some of those former Health Areas implemented Electronic Health Records at different stages (times) during the last decades. As Chapter 5 will show, we will use data from a specific Health Area (Area 4) which was among the first to implement and standardize data collection in Electronic Health Records.

The Overlapping Nature of the Administrative and Health organization of Madrid

The two systems described above are not hierarchically congruent with each other, but share some commonalities. While the health region level is analogous (the Region of Madrid corresponds with the Regional Health System), the city level is different. In the now obsolete Health Areas (still used for some purposes), Madrid city was part of Areas 1, 2, 4, 5, 6, 7 and 11. Of these, only Areas 4 and 7 were exclusive to the city of Madrid. Within these Health Areas are, as stated above, Basic Health Areas. While these are similar in size to the Neighborhood, they are non-overlapping (see Figure 2.1 and Table 2.1). While in most cases they belong to a single district, there are a few exceptions that place the same Basic Health Areas in two different districts. However, the census section level follows a hierarchical structure with Basic Health Areas and there is no census section in two Basic Health Areas.

Conclusion

The consequence of this spatial organization is that, in order to do analysis at the Basic Health Area level, data need to be available at either this level (but this would only be available for data coming from the Health Care System) or at the census section level (that then can be aggregated to the Basic Health Area). For these reasons, all analyses of this dissertation were conducted at the census section level, in order for analysis of health to be feasible.

Detailed Materials and Methods and Exploratory Data Analyses for Aim 1: Measuring Neighborhood Social and Economic Change

This section summarizes the data sources, indicators operationalization and exploratory data analysis for Aim 1 (Chapter 3).

Data Sources for the Indicators of Neighborhood Social and Economic Change

The indicators of neighborhood social and economic change used in aim 1 to build the measurement model were obtained from a diverse set of data sources. The following is a brief description of each data source, along with its limitation, strengths and analytic considerations for their use.

1. *Padron*: the *Padron* is a municipal registry in effect in Spain since 1986. The main purpose of this registry is to help administering several social services that have a spatially explicit catchment area. For example, and as described above, assignment to a given Primary Care

Health Center used to be strictly linked to residential location which was obtained from the *Padron*. The *Padron* population estimates are also used to delineate census sections for electoral purposes. The Spanish law mandates (but does not enforce) notification of residential relocations through the filling of a *Padron* form. This form includes data on education level, country of origin, age and sex. It should be emphasized that data on education level is only updated when a person moves, so people that obtain a degree and do not move again have an outdated education level. This phenomenon does not apply from 2014 onwards, when the National Institute of Statistics sends local government a list of all conferred degrees for the updating of this variable. The *Padron* dataset is composed of two sub-datasets

- a. Cross-sectional cuts: these are point estimates at one given date (usually January 1st or July 1st of a given year).
- b. Mobility Statistics: given the centralized nature of *Padron* through the National Institute of Statistics, data are available on census section of origin and destination for every move within a single city. Moreover, for moves between cities, data are available for census section of origin (for people moving out of the city) or destination (for people moving in). These data also include information on age, education level and country of origin.

2. Cadaster (*Catastro*): the cadaster is a tax registry for all properties in the entire Spanish territory. This registry helps local governments collect property taxes based on area and land use. The data are available on two formats:
 - a. A regular dataset with coordinates comprised of a row for every property, with information on area, year of construction, whether it has been renovated, and its tax value. This last piece of information is not publicly available and was not requested for the purposes of this dissertation.
 - b. A shapefile format, that includes a polygon for each building that ever existed in an area, with data on dates of construction or demolition (or change of features).
3. Data compiled from several sources by the Department of Statistics at the Madrid Local Government. These include data collected by the local government and by other institutions (public or private) but organized, managed and distributed by the Department of Statistics through two different systems and mostly freely available online: *Datos Abiertos Madrid* (Madrid Open Data) and *Banco de Datos de Madrid* (Madrid Data Bank).
 - a. Unemployment data collected by the *Servicio de Empleo Publico Estatal* (National Employment Service). This includes data on people registered as job seekers and those receiving

unemployment benefits. Given that individuals are not required to notify unemployment to the Unemployment Services (unless the individual wants to receive unemployment benefits), this registry may be incomplete and lead to underreporting.

Importantly, data on the active population is not available at the same spatial scales, so it is proxied by age (everyone aged 16 to 64).

- b. Occupation data collected by the Social Security Administration. This includes data on occupational class, industry, part-time and temporal jobs. Available from 2010 onwards for each neighborhood.
- c. Vehicle data, with type of vehicle, for each neighborhood, available from 2004 to 2009.
- d. Idealista Report: This is a report conducted by the biggest real estate corporation in Spain (*Idealista*), with data on average sales price of all residential properties sold through them. From 2002 to 2015 this report has been released with average sales price per m² by neighborhood in Madrid. While these data may not be entirely precise (since not all properties are sold through this company and the sales price may not be entirely accurate) it tracks very well with other SES indicators and major trends in property values over time.

Indicator operationalization

Using the data sources above we operationalized indicators in several ways (see Table 2.3 for more details):

- Change in Raw Counts: this was used for indicators that would otherwise rely on area densities, like population density. We did not use population density (and similar indicators) because census section borders may change (even with the common set of census sections), leading to spurious changes in area densities. For these indicators we used change in raw counts (e.g.: total population at time t – total population at time $t-1$)
- Change in Proportions: we used these for sociodemographic and socioeconomic indicators (e.g., change in proportion of foreign-born). The proportion was usually relative to total population, but some of these indicators relied on denominators different from total population, like people aged 25 or above (for education) or all workers (for workers data).
- Change in Average Values: we used these for ordinal sociodemographic characteristics (education and age) that could be summarized into a single indicator, in order to reduce the number of parameters in the measurement model. A potential disadvantage is that this type of operationalization may ignore changes in the

margins, so we included also proportions in each category as candidate indicators.

- Change in Diversity: we calculated Shannon's entropy (Shannon, 2001) for proportion indicators (age, education, country of origin). Shannon's entropy provides a measure of diversity that is independent of the scale of the indicator. Importantly, it provides a way of measuring differences between proportions without including all proportions into the model. That is, an area with 50% highly educated people and 50% people with secondary education will have a lower diversity than one with 50% highly educated people, 20% secondary education, 20% primary education and 10% no studies. It is calculated as shown below (where i is a variable [e.g., education], j is each category [e.g., people with primary education], and P_{ij} is the portion in each category j of variable i). A higher value indicates higher diversity.

$$\text{Entropy}_i = \sum_{j=1}^J \ln(P_{ij}) * P_{ij}$$

- Change in other per capita indicators: These include indicators that are not, *per se*, proportions (the numerator is not included in the denominator), but are operationalized relative to the population in the area (e.g., vehicles per capita).

- **Mobility Flows:** all mobility statistics are inherently change statistics, as they represent the number of people moving to/from an area in the previous year. To standardize these indicators (since areas with a higher number of people will have a higher number of people moving in or out), we divided the inflows and outflows by the total number of people in the area at time $t-1$.
- **Mobility Balance:** this is the difference between the two flows (inflow-outflow), and represents the population change due to mobility. If inflows are larger than outflows, there is a positive net gain in population through mobility. Note that the change in total population may differ from this number due to other sources of population change (deaths, births).
- **Mobility Throughput:** this is the sum of the two flows, and represents the total number of people involved in mobility.
- **Mobility Potentials:** these are the difference between the characteristics of the people moving into an area (incoming potential) or out from an area (outgoing potential) and are calculated as the proportion of a given characteristic (e.g.: < age 25) in the inflow over the proportion of the same characteristic in the area at time $t-1$.

- Mobility Potential Balance: this is the difference between the incoming potential and the outgoing potential, representing how the area is changing in a given characteristic due to mobility.
- Mobility Potential Throughput: this is the sum of the incoming and the outgoing potentials, representing how different are overall the mobility flows from the current area characteristics.
- Categorized indicators: some indicators were extremely skewed and were therefore categorized. These include new housing or new housing renovations, which was dichotomized into any new house or any housing renovations, and change in housing space per person (a very high proportion of 0), which was categorized into increasing, decreasing and stable.

Exploratory Data Analysis for Aim 1

We conducted a series of exploratory data analysis for Aim 1 data in order to understand the best operationalization of indicators for the measurement of neighborhood social and economic change. We also explored the results of the measurement models, as detailed in Chapter 3. What follows is a summary of these analyses and how they influenced decisions.

Figures 2.2 to 2.5 show the trends in candidate indicators. Figure 2.2 shows the prevalence in the change indicators operationalized as raw values instead of delta (t minus $t-1$). This shows a wide range of values in most

indicators. Importantly, several long-trend patterns become evident: an increase in the average level of education and the proportion of people with university education or above; and a decrease in property value, a decrease followed by a marked increase in unemployment from 2009 onwards.

Figure 2.3 shows the delta operationalization of the change indicators. Here the issues with unemployment and property value are apparent, as, for example, all changes in property value are positive until 2009 and negative from there on. This led us to standardize all indicators to remove these long-trends trends (see Chapter 3 for more details). Figure 2.4 shows an example of this for unemployment. As the right panel shows, the trend is completely eliminated after standardizing the indicator every year.

Figure 2.5 shows the housing indicators that had to be categorized due to a very low proportion (<1%) of census sections having a value above 2. Figure 2.6 shows the mobility indicators where, again, some trends are apparent, including a higher overall mobility during the housing crash years (2009-2010).

Figure 2.7 shows a correlation plot between all change indicators. The lower off-diagonal shows the scatter plots between each pair of indicators while the upper off-diagonal shows the spearman correlation between them. The diagonal shows the distributions of each variable. Figure 2.8 is also a correlation plot of the change indicators but with a delta operationalization ($t - t-1$). Some very high correlations are apparent here too, including correlations between average age or education and proportion in each age or education category (as

expected), between diversity indicators of a characteristic and the proportion of people in categories of such characteristic, and the available housing space and proportion of surface area built. Figure 2.9 shows a correlation plot for mobility indicators. These plots were useful to identify indicators that given the very strong correlation could demonstrate poor performance in the measurement model (e.g., adding change in education proportions to a model with change in average education did not improve model performance).

For more details on the use of these indicators to build a model of neighborhood social and economic change, along with the challenges associated with using this model, please see Chapter 3.

Detailed Materials and Methods and Exploratory Data Analyses for Aim 2: Association of Neighborhood Social and Economic Change with Food Environment Changes

This section summarizes the data sources, indicators operationalization and exploratory data analysis for Aim 2 (Chapter 4).

Data Sources

1. Active Economic Units Directory: this is a directory of businesses addresses. The directory is maintained by the Regional Institute of Statistics of Madrid. Data on the address refers to the business location and may not reflect data on the actual store. Data are collected through a mixture of registries and field work.

It was updated yearly from 2006 through 2010. Areas with less than 5 units are censored, meaning that data is only effectively available at the neighborhood level.

2. Commercial Spaces Census: this is an exhaustive census of all commercial spaces in the city of Madrid, curated by the Department of Statistics of the Local Government of Madrid. Spatial data refers to the actual location of the space, and includes coordinates and census section. This is available from 2012 through 2016, with a yearly to monthly update schedule.

Classification of Food Stores

Classifying food stores by type is a commonly used method to approximate healthy food availability, affordability and accessibility (Glanz et al., 2015). What follows is a more detailed description (as compared to the summary shown in Chapter 4) of the method we used to categorize food stores in Madrid.

Classification of Economic Activities in Spain

All Spanish business are classified by the CNAE (*Clasificación Nacional de Actividades Económicas*), either using its 1993 or 2009 version. The CNAE is an exhaustive four-tiered classification system. The four tiers are: 1) sections (e.g., wholesale and retail; accommodation and food service activities); 2) divisions (e.g., retail trade, except of motor vehicles and motorcycles; food and beverage service activities); 3) groups (e.g., retail sale of food, beverages and tobacco in specialized stores; restaurant and food stands); and 4) classes (retail

trade of fruits and vegetables in specialized establishments; bars). See Figure 2.10 and Table 2.4 for a description of this hierarchical system.

Healthy food availability as approximated by food store type

Our categorization of food stores relies heavily on detecting small specialized stores, especially fruit and vegetable stores. The presence of small Fruit and Vegetable Specialty Stores is abundant in Spain; they are widely used and appreciated by neighbors (Bilal et al., 2016). Compared to other Western countries, the Spanish food environment is dominated by small retailers (Flavián et al., 2002). In Spain, FV Stores have lower scores on standard measures of healthy food availability (such as the HFAI) compared to supermarkets because they lack some healthy foods such as whole grain breads or low-fat milk (Bilal et al., 2016; Díez et al., 2016). We focused on them because they lack healthier products, such as ultra-processed foods (Bilal et al., 2016; Díez et al., 2016). Other small specialty stores also carry less unhealthy products and focus on specific categories of food, such as meat (butcheries), seafood (fishmongers) and baked products such as bread (bakeries) (Bilal et al., 2016; Díez et al., 2016). We also opted to categorize supermarkets as they represent a common food retailing place, especially one that is heavily studied in the food environment literature (Glanz et al., 2015).

In summary, our objective was to classify all food stores into supermarkets and small specialty stores, and within the latter, detect fruit and vegetable stores.

Classification of Food Stores

To classify each store, we created a classification algorithm that uses the CNAE economic activity code provided in the Commercial Spaces Census. We started from data obtained from ground truthing of 12 contiguous census sections of Madrid (Bilal et al., 2016; Díez et al., 2016). Once we developed the first algorithm, we further trained and tested it in 3 census sections with a high number of food stores (>100). The algorithm presented in Figure 2.11 is the final version. A validation study will be conducted in 48 census sections, where the results of ground truthing will be compared to the food store classification that emerges from this algorithm.

The algorithm classifies unspecialized stores into convenience stores, supermarkets or small grocery stores. Convenience stores have their own code (4711.03) and are classified into a convenience store if they have that code. To differentiate between supermarkets and small grocery stores we used name recognition (e.g., a store with the name *Carrefour* was categorized as a supermarket, while a store with the name *Alimentacion Perez* was categorized as a non-supermarket) using a list of 60 supermarket names obtained from the yellow pages. We also applied the same procedure to stores with more than one specialized store code (e.g.: butcher [code 4722] and fruit store [code 4721]).

We classified stores as specialized stores if they had just one specialized store code (e.g.: Fruit and Vegetable Specialty stores were those with a single 4721 code). The specialized stores category was created by summing all fruit

and vegetable stores, butcheries, fishmongers and bakeries. All food stores were the sum of convenience stores, supermarkets, small groceries, specialized stores and other stores (which may include herbal products stores, frozen goods stores, etc.).

Exploratory Data Analysis for Aim 2

We conducted a series of exploratory analysis of Aim 2 data in order to get an overview of the food environment changes, how the classification algorithm worked over time and how variable were store numbers over time. What follows is a summary of these analyses and how they influenced decisions.

Figure 2.12 shows the trends from 2012 to 2016 in the total number of commercial spaces, open commercial spaces, those devoted to retail and those specifically retailing food. On the right of the figure are food stores subcategorized using our algorithm. As this figure shows, the overall number of commercial spaces and open commercial spaces rose over time, fueled especially by an increase in the number of retail stores.

Within food stores, the most evident trend was an increase in unspecialized stores, especially supermarkets. Figure 2.13 shows the proportion of food stores by type and highlights the increase in the proportion of supermarkets, from around 7% in 2012 to 11% in 2016; and a decrease in the number of specialized food stores (from 33% in 2012 to 19% in 2016). The proportion of unclassified stores in our algorithm stayed low over the entire period, at around 3% in 2012 and 3% in 2016.

Figures 2.14 and 2.15 show the average and median number of commercial spaces and food stores by type per census section. A census section has on average around 65 commercial spaces (41 of those open for business, 18 on retail, 7 on food). Again, on average, there is around 1 supermarket and fruit and vegetable store per census section. The median numbers are lower, with around 50 commercial spaces (35 of them open for business, 10 on retail and 4 on food). The median number of fruit and vegetable stores and supermarkets is around 0 (meaning at least half of the areas had no store of that type), which combined with the data on the figure with averages (with a mean of 0.8 and 0.6 fruit and vegetable stores and supermarkets, respectively), indicates that there's a high dispersion in the number of these stores (SD=1.9 and 1.1, respectively), as more than half census sections have no supermarkets or fruit and vegetable stores. For this reason (high number of 0s), we operationalized food stores as increases, decreases or stability in the number of stores.

For more details on the use of this data to study food environment changes and its association with neighborhood social and economic change see Chapter 4.

Detailed Materials and Methods and Exploratory Data Analyses for Aim 3: Association of Neighborhood Social and Economic Change with Diabetes Incidence

This section summarizes the data sources, indicators operationalization and exploratory data analysis for Aim 3 (Chapter 5).

Parent Study

Data for Aim 3 was obtained from an ancillary study of the HeartHealthyHoods Study (www.hhhproject.eu). The parent study aims to understand the relationship between urban environments and cardiovascular disease and risk factors, with a special emphasis on diet, physical activity, alcohol and tobacco. Data are being collected prospectively now, including a cohort study and Electronic Health Records (EHR)-based cohort of the entire city of Madrid.

Data for this dissertation came from the HHH Retrospective Ancillary Study. This ancillary study used a retrospective dataset of the whole population registered in the health centers of four districts of Madrid, from January 1st 2009 to December 31st 2014. We now follow with a description of the Spanish Health Care System, how this affects our data collection processes and details about the data used in this dissertation.

The Spanish Health Care System

The Spanish Health Care system was restructured and revamped in 1986 through the General Healthcare Act (*Ley General de Sanidad, 14/1986*), LGS from now on. The LGS introduced several changes to the Spanish Healthcare System that used to be organized around a German-inspired social insurance model (Rodriguez et al., 1999). This law transformed the system into a British-inspired National Health Service model, funded through general taxes revenue and organized at the Autonomous Region model (Rodriguez et al., 1999). From

the implementation of the Act until 2002, only the so-called *Historical Regions* (i.e.: Catalonia, the Basque Country, Andalusia, Galicia, the Valencian Country, Navarre and the Canary Islands) organized their Health System (Rodriguez et al., 1999). From 2002 on, all Spanish Autonomous Regions took control of their Health System and the National Healthcare Institute (*INSALUD*) was made obsolete. Essentially this means that from 2002 onwards, the Madrid Region had autonomous power to manage the Madrid Health System under the basic regulations that the LGS provided.

The Importance of Primary Care in Spain

The re-design of the Spanish healthcare System based on the British NHS not only brought universalization but also emphasized the importance of primary care doctors as *gatekeepers* of the system. Users of the healthcare system in Spain do not have direct access to specialist care in hospitals or outpatient settings but rather have to go through their assigned primary care doctor, who, if needed, will then refer to a specialist within the system. Primary care doctors may diagnose and treat some conditions or may refer users to a specialist. If the specialist makes a diagnosis and starts a chronic treatment (e.g., diabetes or hypertension), the patient has to come back to the primary care doctor in order to obtain future prescriptions. What this essentially means is that all chronic conditions must be registered at the Primary Care level, or else prescriptions will not be available.

Healthcare access: universalization and back

With subsequent regulations from 1986 through 2011, the Spanish Health System has almost universal coverage, free at the point of access. The regulation of access was complex, as it was a hybrid of a National Health System (everyone has the right to access given certain criteria are met) and a Social Insurance System (where access is linked to job structures), inherited from the pre-1986 organization. Through different regulations, where people without resources, unemployed, etc., all gained access, coverage was around 99% by 2011 (Cuadra, 2011).

In 2011, the newly enacted *Public Health General Act (Ley General de Salud Publica, 33/2011)*, finalized the transition towards a complete National Health Service where everyone was guaranteed coverage, regardless of legal or work status (Cuadra, 2011). This was reversed in 2012, with the Executive Order 16/2012, that excluded undocumented immigrants from the healthcare system. Some regions have continued providing access to undocumented migrants (in the case of Madrid, from August 2015). This means that there may be a potential gap in health insurance from 2012 to 2015 in Madrid, and these individuals may not show in our database from 2012 onwards (since last follow-up time is December 31st 2014).

The Healthcare Card

The main instrument to determine eligibility is the Healthcare Card (*tarjeta sanitaria*). These are granted by each Regional Health System but allow access

to the entire system in Spain. In order to obtain it, people must be registered in a municipal registry (the *Padron*, that is used also for education and other services) and then request it at a Primary Care Center. These will determine eligibility, grant a National Healthcare ID number (in the case of the first registration) and a Regional Healthcare ID number (in the case of the first registration in Madrid), and assign the users to a doctor/nurse team at the Primary Care Center. Individuals can then switch doctors or even Primary Care Centers, with no restriction.

Electronic Health Records in Madrid

The Electronic Health Records (EHR) system implementation in Spain has been patchy. The main reason is the different organizational structures by region. To begin with, the Regional Health Systems EHR systems are mostly non-commensurable with each other, so analysis must always be restricted to the regional level (e.g., entire region of Madrid) if data compatibility is desired. Within each Regional Health System, the implementation has also been patchy due to internal differences in organization. More importantly, EHR systems have often been implemented in a two-tier system: one system for each hospital and one system for the entire primary care system. Nonetheless, chronic conditions diagnosed at the hospital level will appear in the primary care EHR system given the necessity of prescriptions.

In Madrid, the EHR system was started in 2001 in the Area 4. By 2004, all primary care centers in this area shared the same system called *OMIap*. By

2009, all primary care centers of the area had all paper-based medical records translated to the OMIap system and all record collection methods were standardized. Other Health Areas of Madrid had more delayed timelines for implementation and only recently the entire system has become universal (*APMadrid*). Hence, all the analyses that include EHR data were restricted to the Health Area 4, which has had this system running for the longest and introduced standardization measures since 2009.

The HeartHealthyHoods Retrospective Study

With the above in mind, and in collaboration with the Research Unit and the Information Systems Unit of the Primary Care Directorate of the Madrid Regional Government we set up an EHR-based cohort of the entire Health Area 4 of Madrid city, with data for all people registered in a health center of this area from 2009 to 2014. This encompasses 4 districts with around 25% of the population of Madrid. The data from the EHR is divided into several files, linkable through the unique ID of each individual:

- Population dataset: these are six separate datasets (one for every year from 2009 to 2014) with a row for every person registered in an Area 4 center. This includes date of birth, sex, center, and for the 2013 and 2014 dataset, the census section the individual lives in. Moreover, the dataset also includes a flag for moving out of the area or dying, with the date of moving or death.

- Morbidity dataset: this is a single dataset with all recorded diagnoses and date of diagnosis for every individual registered in an Area 4 center. Due to the timeline of the implementation of the EHR in the area, only diagnoses dated after January 1st 2009 are considered precise in terms of date and allow for analyses of the incidence. Diagnoses before 2009 can only be considered prevalent by January 1st 2009. Diagnoses are classified using the ICPC-2 classification developed by WONCA (World Organization of Family Doctors) (2011).
- Clinical datasets: these are two datasets (one for 2013, one for 2014) with all laboratory values of lipids and HbA1c (%) for every individual registered in an Area 4 center.

Our interest in Aim 3 was to study incident diabetes from 2009 onwards in people aged 40 or above by baseline. The reason behind studying people aged 40 or above is the existence of a system in place to collect data for cardiovascular risk factors in people aged 40 or above in Madrid, providing more accuracy in the collection of diabetes data (Bilal et al., 2016). We created a dataset with individuals aged 40 or above in the population file of 2009, and matched it to the subsequent population files including only individuals that were present in 2009. We then matched these individuals to their diagnoses in the morbidity file, and excluded those that had a diabetes diagnose code dated before January 1st 2009. We then created a long dataset, where each row was

an individual-year of observation, with entry on January 1st and exit on December 31st, date of diabetes diagnosis (if any) or date of death/moving out (if any).

Exploratory Data Analysis for Aim 3

We conducted a series of exploratory analysis of Aim 3 data in order to assess the spatial distribution of the EHR data and how representative were the health centers in the four areas compared to the rest of the city. What follows is a summary of these analyses and how they influenced decisions.

Figure 2.16 shows a map of the entire city of Madrid (and surrounding areas) along with the location of all health centers (in red). The figure also shows, in the top-right part of Madrid, the four districts that make up the Health Area 4, our Study Area for Aim 3 (Ciudad Lineal, Hortaleza, San Blas-Canillejas and Barajas). The health centers where data was obtained from are highlighted in blue.

Figure 2.17 shows the distribution of key sociodemographic and socioeconomic variables in these districts and (on the left) in the entire city of Madrid. In terms of average education level, these areas have census sections in the entire spectrum. In particular, San Blas covers the lower end, while both Ciudad Lineal and Hortaleza have a high concentration of areas thorough the spectrum and in the upper end. Barajas also has some of the areas with the highest average education level in the city. Regarding age, the districts in the study area cover the spectrum, with San Blas covering the higher and lower end.

There are a few outliers in the city (with average age above 60 or below 30) that are not represented in this area. Regarding country of birth, both OECD foreign-born and non-OECD foreign-born (as proxies for developed and developing countries) are covered, especially with the high levels of non-OECD migrants in San Blas (which includes a few of the city's outliers) and Ciudad Lineal, and the high levels of OECD migrants in Hortaleza and San Blas. Regarding unemployment and property value, the trend is similar as with the rest of the city: San Blas covers the lower end of the SES spectrum (high unemployment, low property value) while the other districts cover the higher end (low unemployment, high property value). The higher end of property value is not as well represented as the other variables, as only a few areas in Hortaleza go above 5000 EUR/m². Nonetheless, in summary, these areas seem to represent the city of Madrid well, especially from the lower end to the mid-higher end of the SES and demographic spectrums.

Figure 2.18 shows the catchment areas. As detailed in this chapter, health centers no longer have designated catchment areas as people have freedom of choice to change health centers. Nonetheless, as people are automatically registered in the health center of their area, these catchment areas tend to be reproduced if people do not exercise their freedom of choice. Hence, empirical catchment areas are defined as the health center that covers at least 50% of the population of the census section. This classified all census sections of the area

except for one, that was split between three centers. As the figure shows, these catchment areas are well defined spatially, and may cross district boundaries.

Conclusion

In this Chapter, we have provided a detailed exposition of several overarching methodological challenges, with a special focus on selecting the most appropriate spatial unit of analysis. We have also provided subsequent details on some of the data sources and methods employed in Chapters 3 through 5.

Tables

Table 2.1. Characteristics by January 1st 2011 of administrative units overall, for Aim 1 and Aim 2 and for Aim 3.

Unit	Description	Spain			Aim 1 and Aim 2			Aim 3		
		N	Area*	Population*	N	Area*	Pop.*	N	Area*	Pop.*
Autonomous Communities	Main Regional division of Spain	17	11073.55 [4995.87-93827.15]	2078.3 [319.4-8343.7]	+	+	+	+	+	+
Provinces	Main Regional Subdivision	50	9722.31 [1906.09-21792.47]	650 [93.2-6387.4]	+	+	+	+	+	+
Municipalities	Main Local division of Spain	8114	34.9 [0.03-1753.85]	0.6 [0-3186.6]	+	+	+	+	+	+
Census Districts	Main Local subdivision	10500	27.44 [0.02-1668.26]	1 [0-263.1]	21	14.07 [4.68-237.9]	143.7 [45.6-246]	4	24.84 [11.44-41.85]	163.7 [45.6-219.5]
Neighborhoods	Sub-divisions of districts (Madrid)	NA	NA	NA	128	1.34 [0.25-188.07]	22.8 [0.8-69.3]	28	1.74 [0.25-28.25]	18.4 [1.6-61.6]
Census Sections	Basic census area	35823	0.22 [0-1125.11]	1.2 [0-13.1]	2409	0.04 [0.01-94.56]	1.2 [0.1-4.5]	429	0.05 [0.01-28.25]	1.2 [0.4-4.3]
Common Census Sections	Common set of census sections from 2005 to 2015	#	#	#	2272	0.04 [0.01-94.56]	1.2 [0.2-34.4]	393	0.04 [0.01-28.25]	1.2 [0.4-30]

Footnote

*Area is in km², and shown as the Median [Range]; Population is in 1000s of residents, and shown as the Median [Range]

NA: Does not apply (neighborhoods are unique to Madrid City)

+ This unit is larger than the area under study

Common census sections were only constructed for Madrid City

Dashed line separates city level vs macro-level units.

Table 2.2. Changes in Census Sections over time in Madrid

Year	Number of Census Sections	Splits from Last Year	Merges from Last Year	% Changed (merges and splits)
2005	2363	8	1	0.4%
2006	2386	45	22	2.8%
2007	2381	10	15	1.0%
2008	2396	15	0	0.6%
2009	2398	17	15	1.3%
2010	2409	12	1	0.5%
2011	2409	9	9	0.7%
2012	2409	0	0	0.0%
2013	2412	15	12	1.1%
2014	2415	8	5	0.5%
2015	2420	9	4	0.5%

Table 2.3: Full list of the 60 proposed indicators of neighborhood change

Indicator	Operationalization*	Source	Final List of Indicators?	Final Model?
Δ Mean Age	% in each 5-year age bin in * Mid Point of age bin	Padron	Yes	Yes
Δ Proportion Aged <25	# Aged < 25 / Total Population	Padron	Yes	No
Δ Proportion Aged >64	# Aged > 64 / Total Population	Padron	Yes	No
Δ Proportion Born in Spain	# Born in Spain / Total population	Padron	No (Redundant)	No
Δ Proportion Foreign-Born in non-OECD	# Born in non-OECD Country / Total population	Padron	Yes	Yes
Δ Proportion Foreign-Born in OECD	# Born in OECD Country / Total population	Padron	Yes	Yes
Δ Mean Education Level	% in each education group (4 groups) * 1 to 4 (for no official studies, primary education, secondary education, university education)	Padron	Yes	Yes
Δ Proportion Low Education	# with primary education or below / # aged 25 or above	Padron	Yes	No
Δ Proportion Medium Education	# with secondary education / # aged 25 or above	Padron	No (redundant)	No
Δ Proportion High Education	# with university education or above / # aged 25 or above	Padron	Yes	No
Δ Property Value	Average price sale of all housing units sold in EUR/sqm.	Idealista Report	Yes	Yes
Δ Unemployment Rate	# Registered for Unemployment / Population 16 to 64	Unemployment Serv.	Yes	Yes
Δ Proportion Unskilled Occupational Class	# Workers in Unskilled Manual or Non-Manual Work (Classes I, II, III and IV) / # All Workers	Social Security	No (not available 2005-2009)	No
Δ Proportion Services/Construction	# Workers in Services, Retail, Construction and Home Work / # All Workers	Social Security	No (not available 2005-2009)	No
Δ Proportion Part-Time	# Workers in Part-Time Work / # All Workers	Social Security	No (not available 2005-2009)	No
Δ Proportion Temporal	# Workers in Temporal Work / # All Workers	Social Security	No (not available 2005-2009)	No
Δ Young Workers	# Workers in Below Age 25 / # All Workers	Social Security	No (not available 2005-2009)	No

Δ Vehicles per capita	# Vehicles (all types) / Total Population	City Hall Statistics	No (not available 2010-2015)	No
Δ Cars per capita	# Vehicles (cars) / Total Population	City Hall Statistics	No (not available 2010-2015)	No
Any Renovation	Any housing renovation conducted in that year	Cadastré	Yes	Yes
# Renovations	# housing renovations conducted in that year	Cadastré	No (extremely skewed, dichotomized)	No
Any Integral Renovations	Any integral housing renovation (complete renovation) conducted in that year	Cadastré	Yes	No
New Housing	Any new housing built in that year	Cadastré	Yes	Yes
# New Houses	# new housing built in that year	Cadastré	No (extremely skewed, dichotomized)	No
Δ Housing Space / Person	Total Housing Space / Total Population (increased/stable/decreased)	Cadastré and Padron	Yes	Yes
Δ Built Ground Density	Sum of all ground surface area built for residential purposes / Total Census Section Area	Cadastré	Yes	No
Δ Median Year of Construction	Median year of construction of all housing units in the area	Cadastré	No (extremely low variability over time)	No
Δ Maximum Number of Floors	Number of floors in the tallest building	Cadastré	No (extremely low variability over time)	No
Δ Total Population	Total Population	Padron	Yes	Yes
Δ Population Density	Total Population / Area	Padron	No (changed with census section boundary changes)	No
Mobility Inflows	# People Incoming to the area during time t-1 / Total People in the Area at time t-1	Padron	No (Redundant)	No
Mobility Outflows	# People Outgoing from the area during time t-1/ Total People in the Area at time t-1	Padron	No (Redundant)	No
Mobility Throughput	# People Incoming to the area during time t-1 / Total People in the Area at time t-1 +	Padron	Yes	Yes

	# People Outgoing from the area during time t-1/ Total People in the Area at time t-1			
Mobility Balance	# People Incoming to the area during time t-1 / Total People in the Area at time t-1 - # People Outgoing from the area during time t-1/ Total People in the Area at time t-1	Padron	Yes	No
Proportion Incoming from Out of Madrid	# People Incoming to the area from Out of Madrid during time t-1 / # People Incoming to the area during time t-1	Padron	Yes	No
Proportion Outgoing to Out of Madrid	# People Outgoing from the area to Out of Madrid during time t-1 / # People Outgoing from the area during time t-1	Padron	Yes	No
Mobility Throughput of people aged 25 or below	% below 25 years of age in incoming people during time t-1 / % below 25 years of age at time t-1 + % below 25 years of age in outgoing people during time t-1 / % below 25 years of age at time t-1	Padron	Yes	Yes
Mobility Balance of people aged 25 or below	% below 25 years of age in incoming people during time t-1 / % below 25 years of age at time t-1 - % below 25 years of age in outgoing people during time t-1 / % below 25 years of age at time t-1	Padron	Yes	No
Mobility Throughput of people aged 25 or above	% above 25 years of age in incoming people during time t-1 / % above 25 years of age at time t-1 + % above 25 years of age in outgoing people during time t-1 / % above 25 years of age at time t-1	Padron	No (Redundant)	No
Mobility Balance of people aged 25 or above	% above 25 years of age in incoming people during time t-1 / % above 25 years of age at time t-1 - % above 25 years of age in outgoing people during time t-1 / % above 25 years of age at time t-1	Padron	No (Redundant)	No
Mobility Throughput of Spain-Born	% Spain-Born in incoming people during time t-1 /% Spain-Born at time t-1 + % Spain-Born in outgoing people during time t-1 /% Spain-Born at time t-1	Padron	No (Redundant)	Yes
Mobility Balance of Spain-Born	% Spain-Born in incoming people during time t-1 /% Spain-Born at time t-1 - % Spain-Born in outgoing people during time t-1 /% Spain-Born at time t-1	Padron	No (Redundant)	No

Mobility Throughput of Foreign-Born (non-OECD)	% foreign-born (non-OECD) in incoming people during time t-1 / % foreign born (non-OECD) at time t-1 + % foreign-born (non-OECD) in outgoing people during time t-1 / % foreign born (non-OECD) at time t-1	Padron	Yes	Yes
Mobility Balance of Foreign-Born (non-OECD)	% foreign-born (non-OECD) in incoming people during time t-1 / % foreign born (non-OECD) at time t-1 - % foreign-born (non-OECD) in outgoing people during time t-1 / % foreign born (non-OECD) at time t-1	Padron	Yes	No
Mobility Throughput of Foreign-Born (OECD)	% foreign-born (OECD) in incoming people during time t-1 / % foreign born (OECD) at time t-1 + % foreign-born (OECD) in outgoing people during time t-1 / % foreign born (OECD) at time t-1	Padron	Yes	No
Mobility Balance of Foreign-Born (OECD)	% foreign-born (OECD) in incoming people during time t-1 / % foreign born (OECD) at time t-1 - % foreign-born (OECD) in outgoing people during time t-1 / % foreign born (OECD) at time t-1	Padron	Yes	No
Mobility Throughput Low Education	% with low education in incoming people during time t-1 / % with low education at time t-1 + % with low education in outgoing people during time t-1 / % with low education at time t-1	Padron	Yes	No
Mobility Balance Low Education	% with low education in incoming people during time t-1 / % with low education at time t-1 - % with low education in outgoing people during time t-1 / % with low education at time t-1	Padron	Yes	No
Mobility Throughput Medium Education	% with medium education in incoming people during time t-1 / % with medium education at time t-1 + % with medium education in outgoing people during time t-1 / % with medium education at time t-1	Padron	No (Redundant)	No
Mobility Balance Medium Education	% with medium education in incoming people during time t-1 / % with medium education at time t-1 - %	Padron	No (Redundant)	No

	% with medium education in outgoing people during time t-1 / % with medium education at time t-1			
Mobility Throughput High Education	% with high education in incoming people during time t-1 / % with high education at time t-1 + % with high education in outgoing people during time t-1 / % with high education at time t-1	Padron	Yes	No
Mobility Balance High Education	% with high education in incoming people during time t-1 / % with high education at time t-1 - % with high education in outgoing people during time t-1 / % with high education at time t-1	Padron	Yes	No
Δ Education Diversity	- sum of [ln(proportion in each education group) * proportion in each education group]	Padron	Yes	Yes
Δ Country of Origin Diversity	- sum of [ln(proportion in each country of origin group) * proportion in each country of origin group]	Padron	Yes	Yes
Δ Age Diversity	- sum of [ln(proportion in each age group) * proportion in each age group]	Padron	Yes	No
Δ # Business	Total number of business in the area	Business Registry	No (not available for 2011-2015)	No
Δ # Retail Business	Total number of retail business in the area	Business Registry	No (not available for 2011-2015)	No
Δ Proportion Open Commercial Spaces	# Open Commercial Spaces / # Total Commercial Spaces	Commercial Spaces Census	No (not available for 2005-2011)	No
Δ Proportion Retail	# Retail Commercial Spaces / # Total Commercial Spaces	Commercial Spaces Census	No (not available for 2005-2011)	No
Δ Proportion Hospitality	# Hospitality (bars, restaurants) Commercial Spaces / # Total Commercial Spaces	Commercial Spaces Census	No (not available for 2005-2011)	No

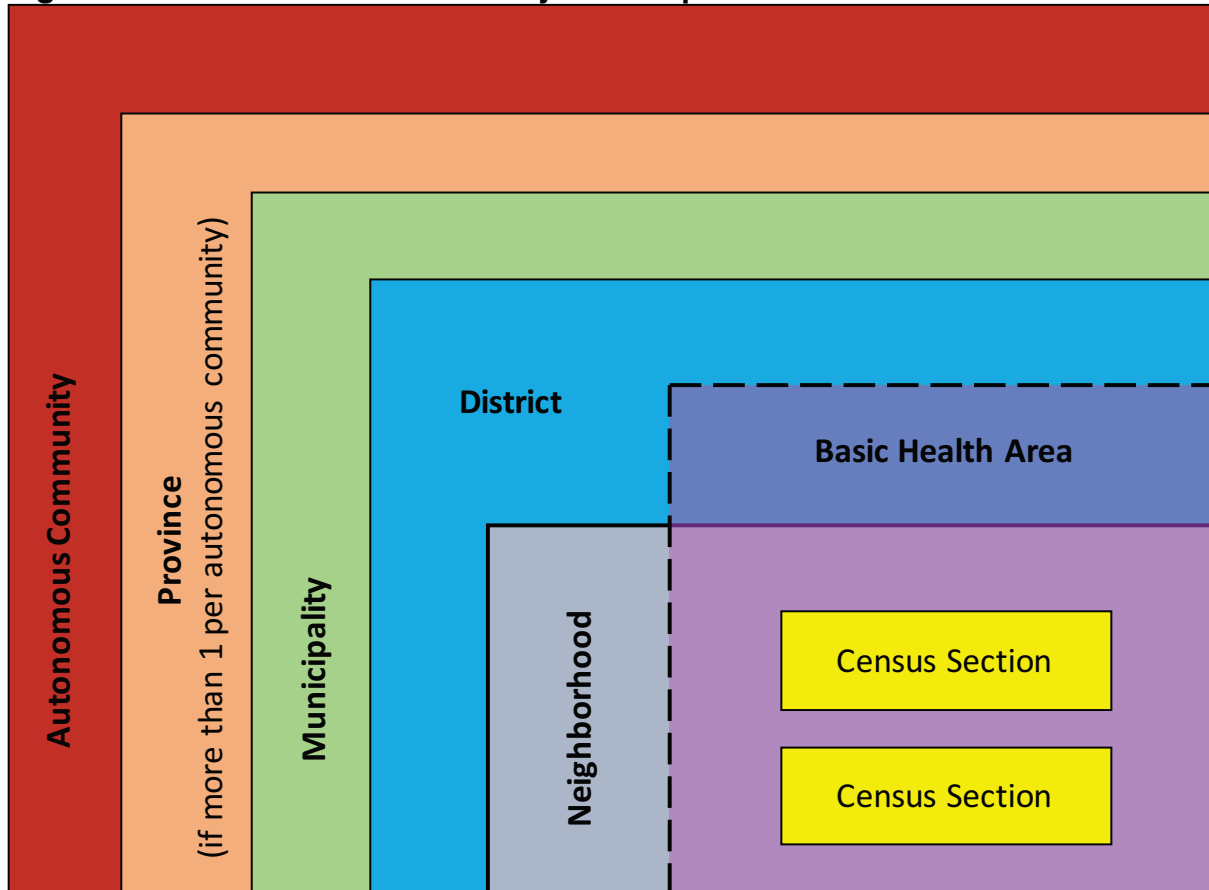
Footnote: all delta indicators (Δ) are operationalized as indicator at time t – indicator at time t-1; all non-delta indicators are operationalized as shown in the table

Table 2.4. Classification of Food Retailing Commercial Spaces Economic Activities under the CNAE1993 and CNAE2009 systems.

CNAE93	CNAE2009	Definition
G	G	Wholesale and retail trade; repair of motor vehicles and motorcycles
52	47	Retail trade, except of motor vehicles and motorcycles
521	471	Retail sale in non-specialized stores
5211	4711	Retail sale in non-specialized stores focused on Food, Alcohol or Tobacco
5219	4719	Retail sale in non-specialized stores not focused on Food, Alcohol or Tobacco
522	472	Specialized Retailing of food, alcohol or tobacco
5221	4721	Specialized Retailing of Fruits and Vegetables
5222	4722	Specialized Retailing of Meat and Animal Products
5223	4723	Specialized Retailing of Fish and Seafood
5224	4724	Specialized Retailing of Bread and Baked Products
5225	4725	Specialized Retailing of Beverages (Liquor Stores)
5226	4726	Specialized Retailing of Tobacco (Tobacco Stores)
5227	4729	Other Specialized Retailing

Figures

Figure 2.1. Administrative Hierarchy of the Spanish Census



Footnote: neighborhoods are unique to Madrid; basic health areas are not part of the census divisions but are included here for illustrative purposes

Figure 2.3. Exploratory Data Analysis for Aim 1: Trends in Change Indicators (Delta operationalization) from 2005 to 2015 in Madrid (Spain)

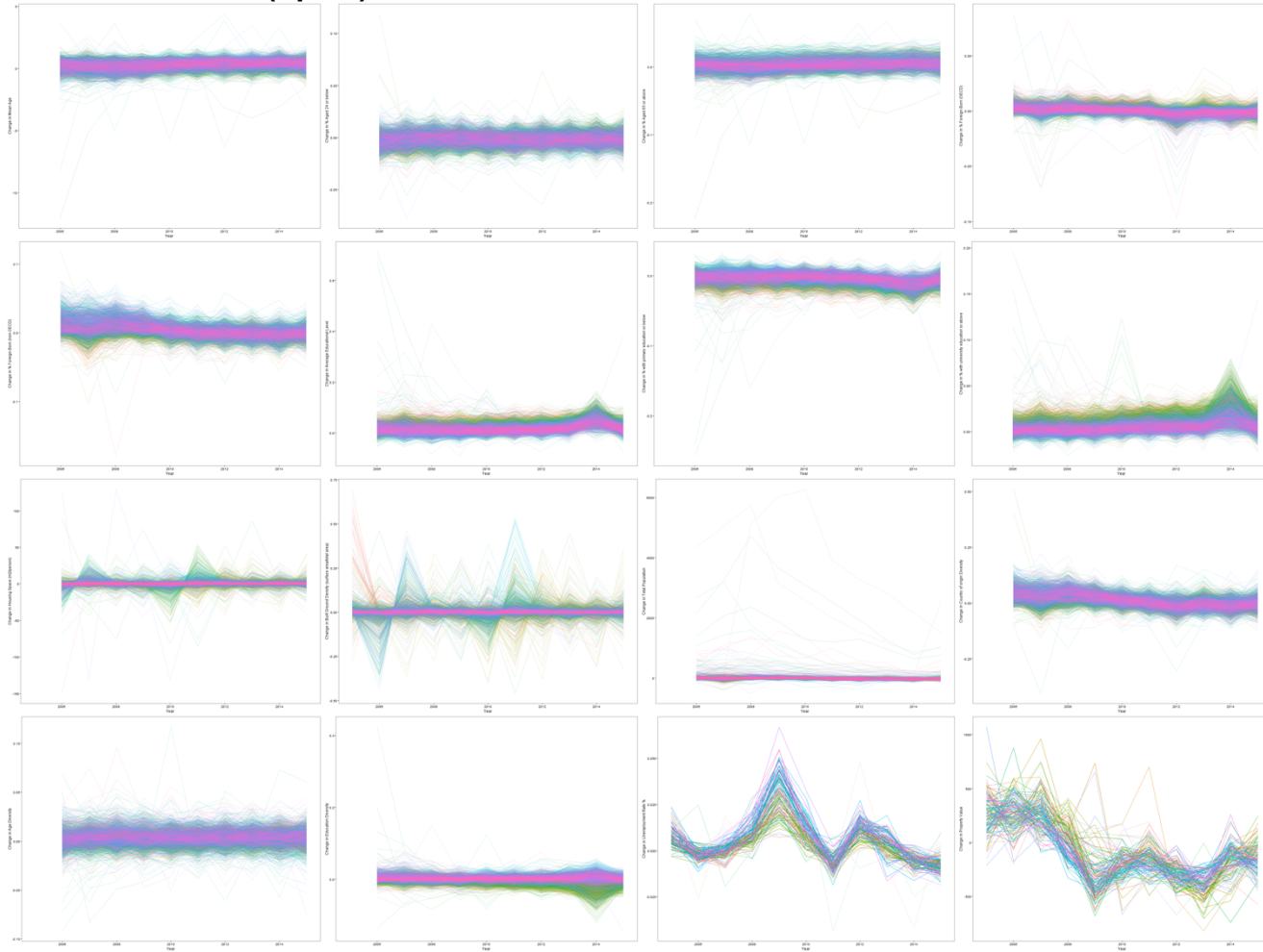
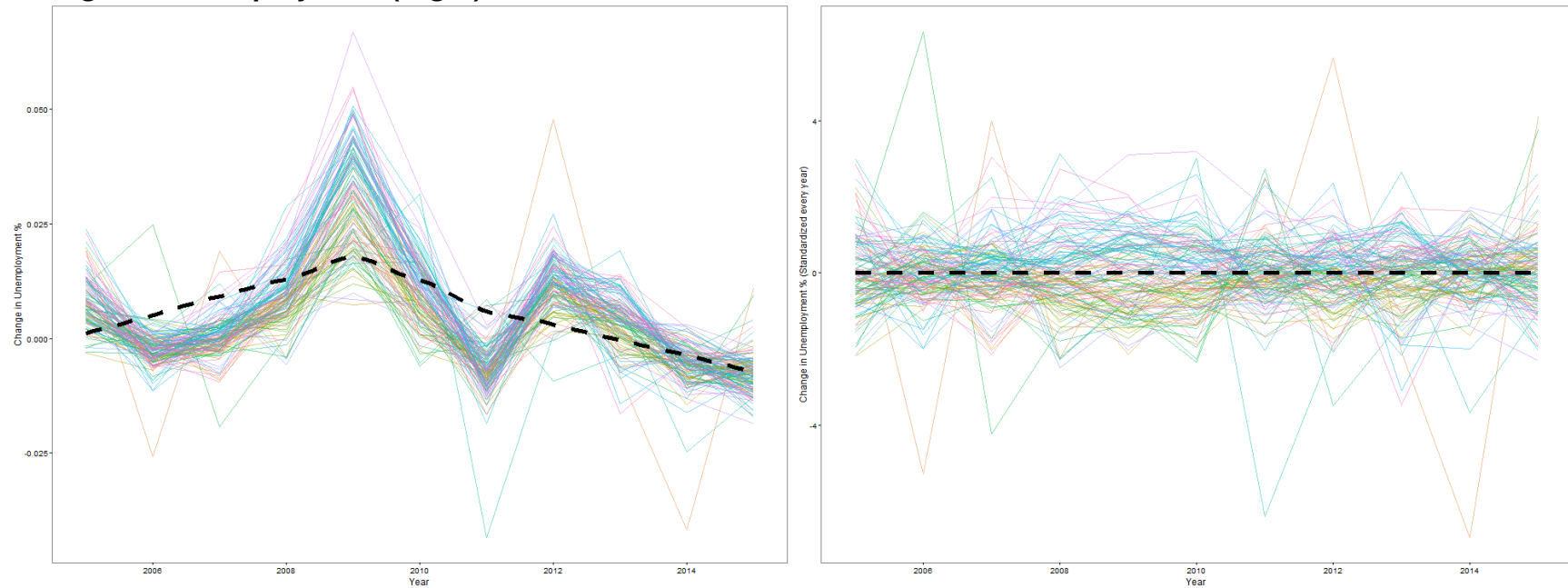


Figure 2.4. Exploratory Data Analysis for Aim 1: Comparison of Changes in Unemployment (left) vs Standardized Changes in Unemployment (Right)



Footnote: thick dashed black line is a lowess mean estimator over time.

Figure 2.5. Exploratory Data Analysis for Aim 1: Trends in Raw Counts of Housing Variables from 2005 to 2015 in Madrid (Spain)

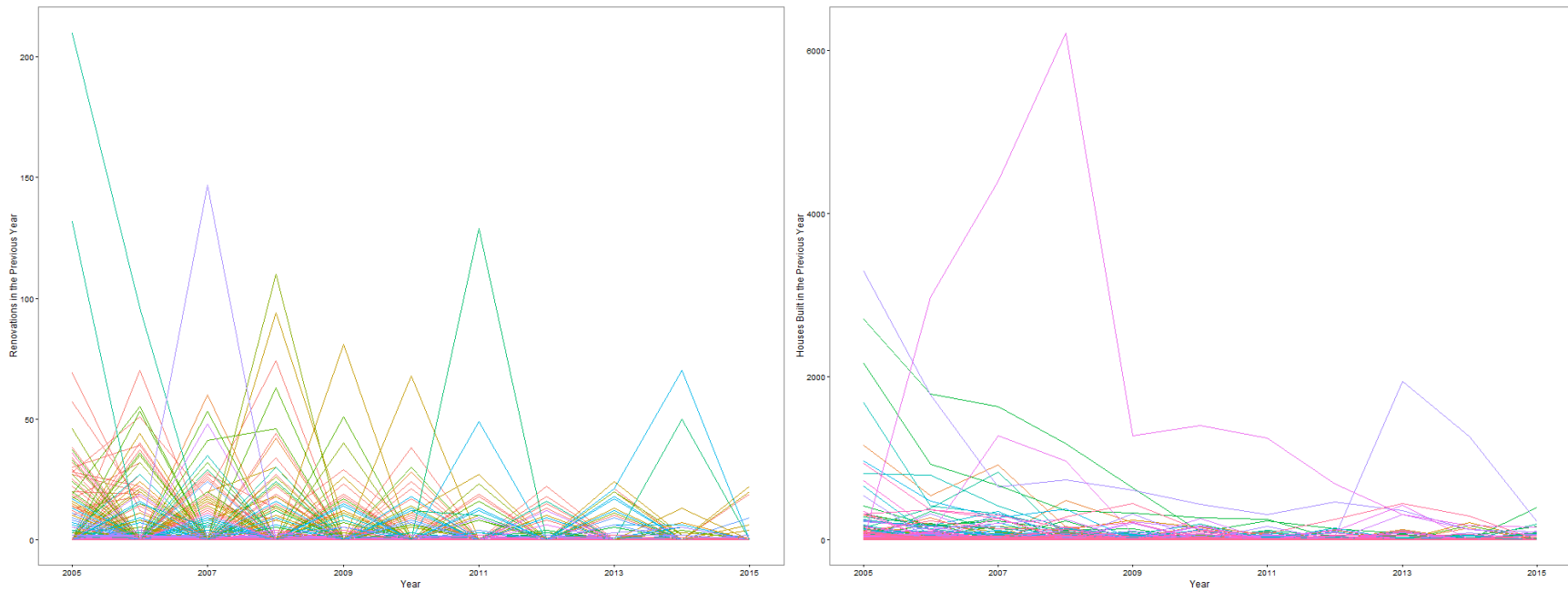


Figure 2.6. Exploratory Data Analysis for Aim 1: Trends in Mobility Variables from 2005 to 2015 in Madrid (Spain)

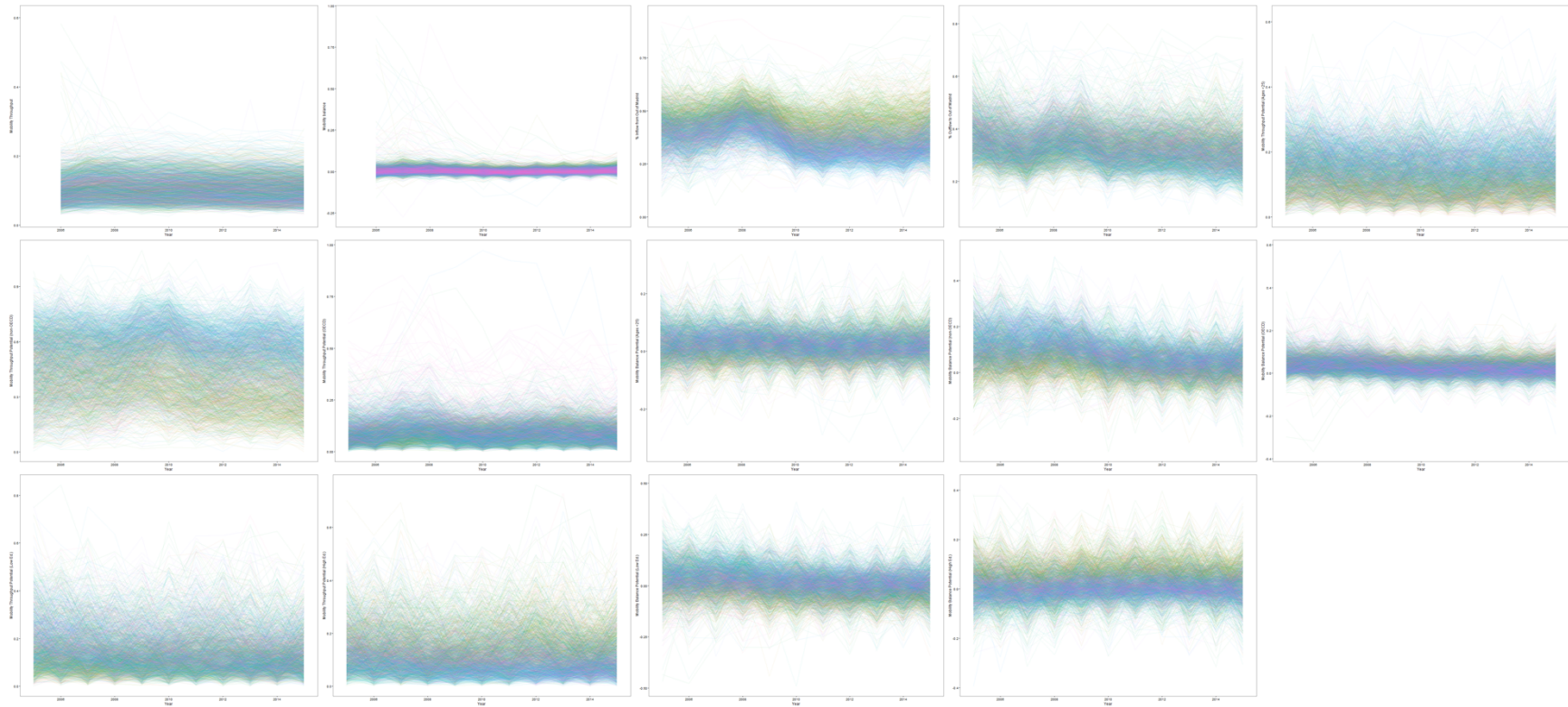


Figure 2.7. Exploratory Data Analysis for Aim 1: Correlation Plot between All Change Indicators (no Delta)

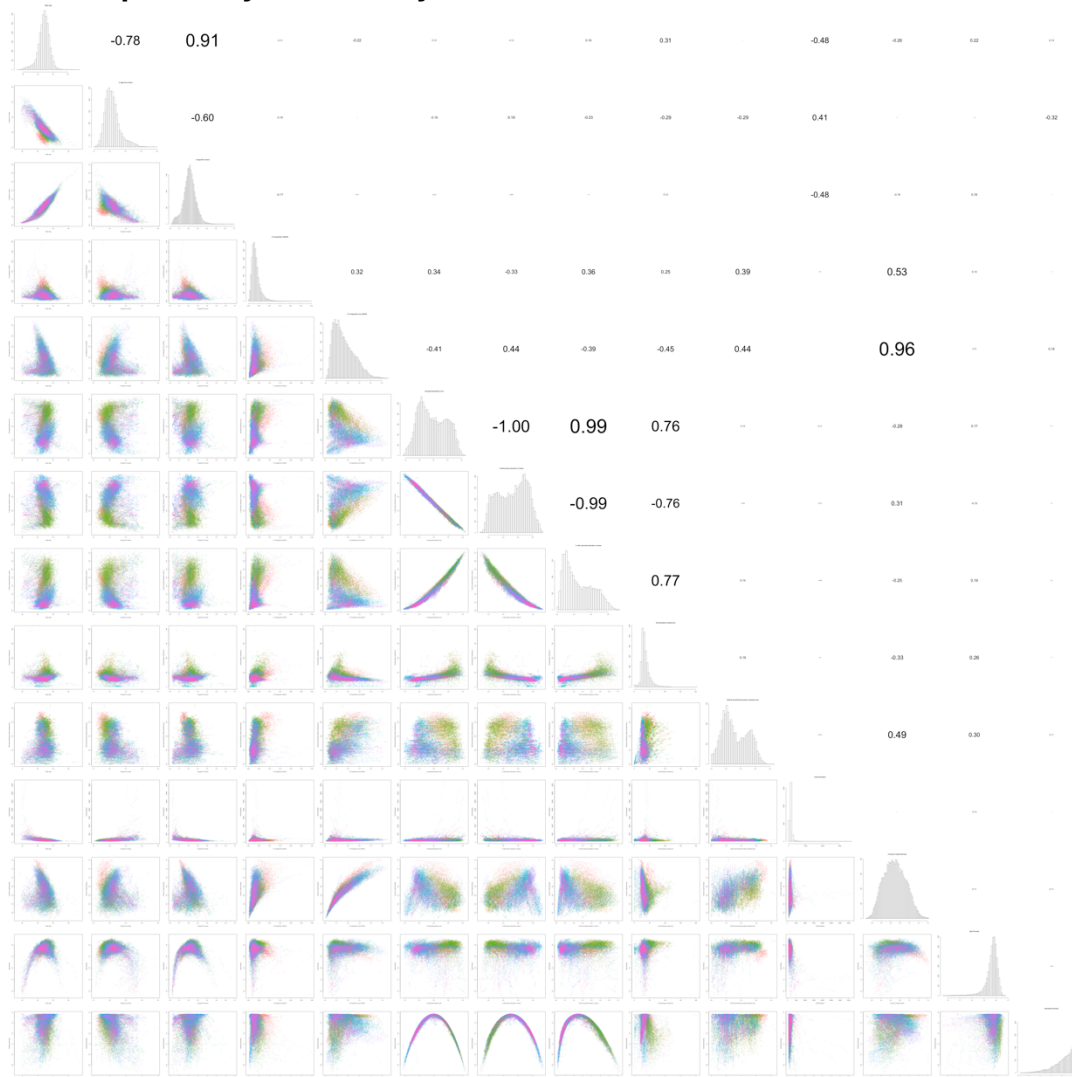


Figure 2.8. Exploratory Data Analysis for Aim 1: Correlation Plot between All Change Indicators (Delta)



Figure 2.9. Exploratory Data Analysis for Aim 1: Correlation Plot between All Mobility Indicators

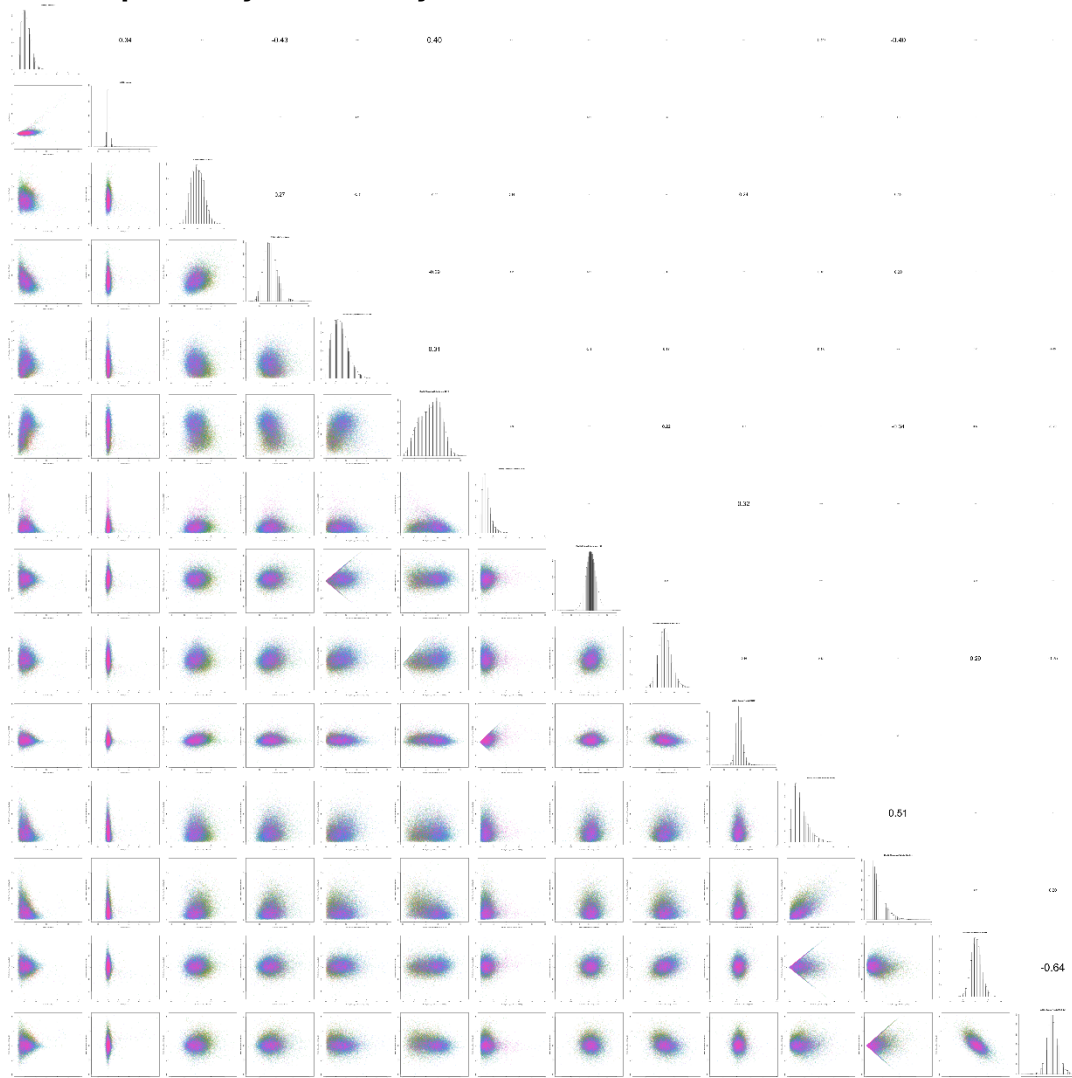
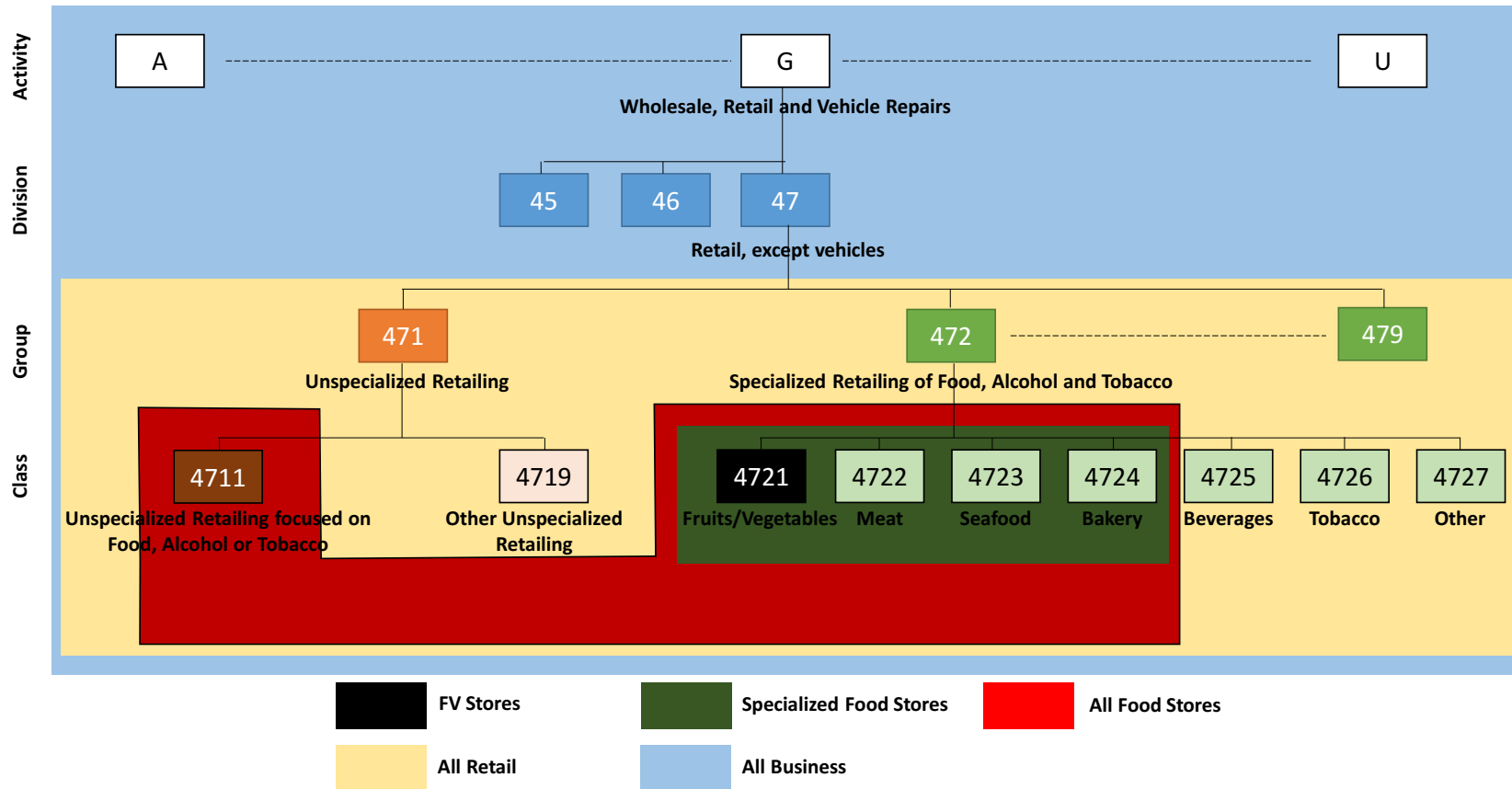


Figure 2.10. Classification of all Commercial Spaces based on Economic Activity



Footnote: Activities are represented by a letter (e.g., G is wholesale, retail and vehicle repairs; A is agriculture and fishing, etc.). Divisions are represented by two digits (e.g., 47 is retail except vehicles, 45 is vehicle retailing). Groups are represented by three digits (e.g., 471 is unspecialized retailing, 479 is other retailing not included in the previous categories). Classes are represented by four digits (e.g., 4721 are FV stores).

Figure 2.11. Algorithm to classify food stores based on declared CNAE activities and store name.

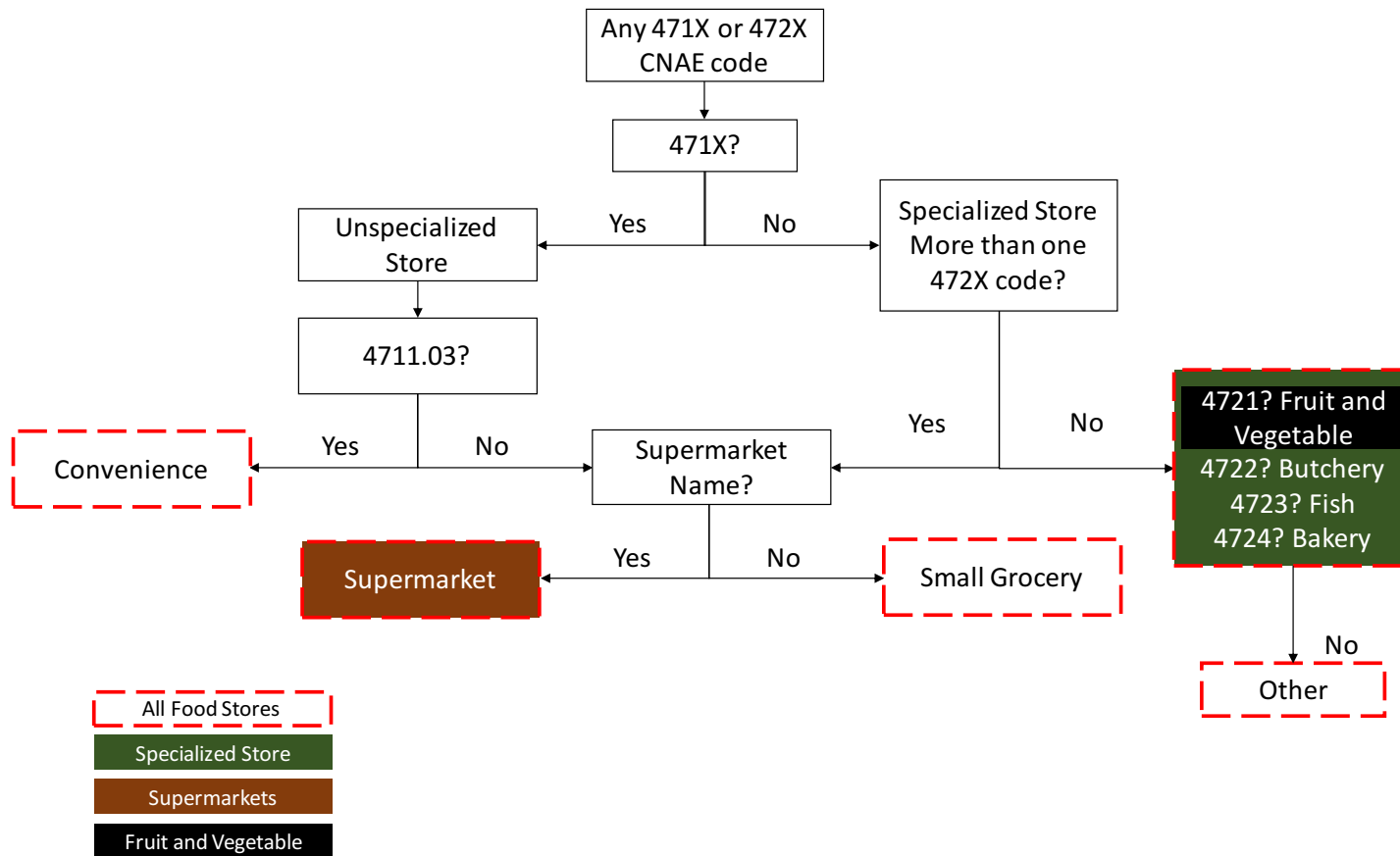


Figure 2.12. Exploratory Data Analysis for Aim 2: Trends in the Number of Stores by Type

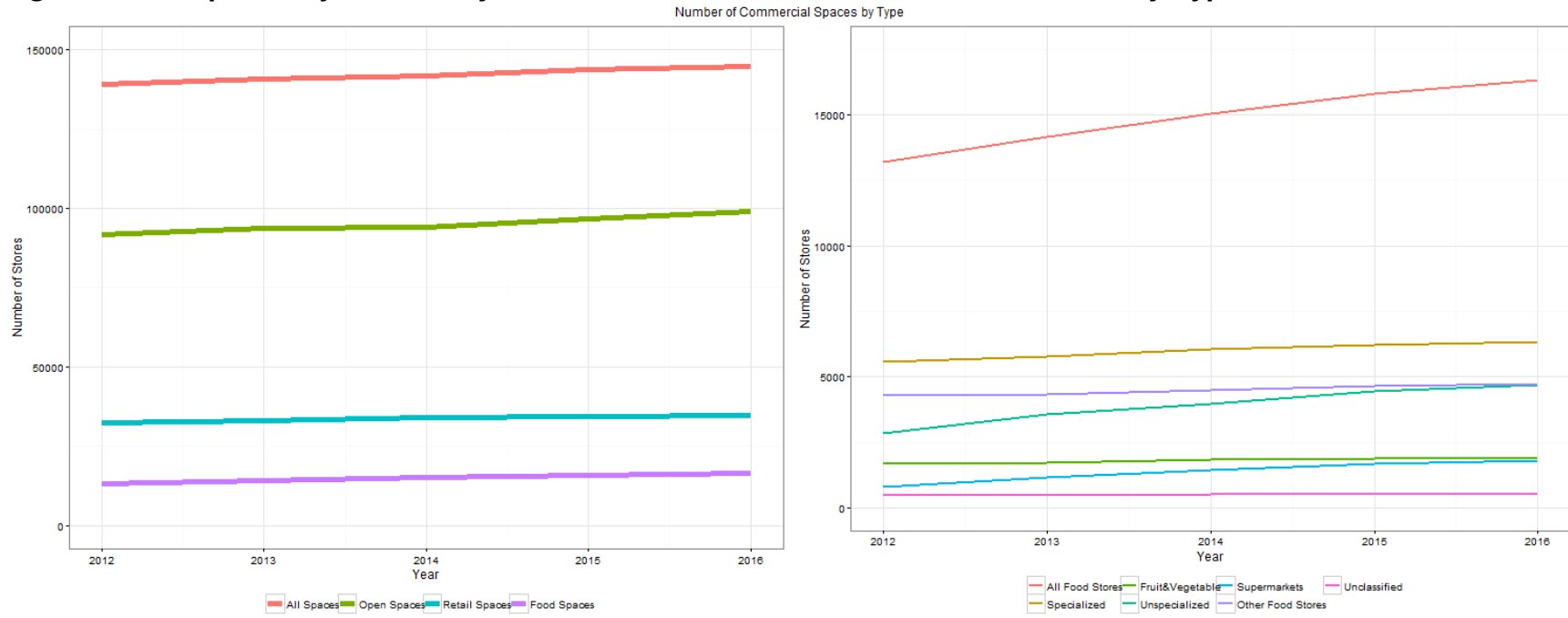


Figure 2.13. Exploratory Data Analysis for Aim 2: Trends in the Proportion of Food Stores by Type

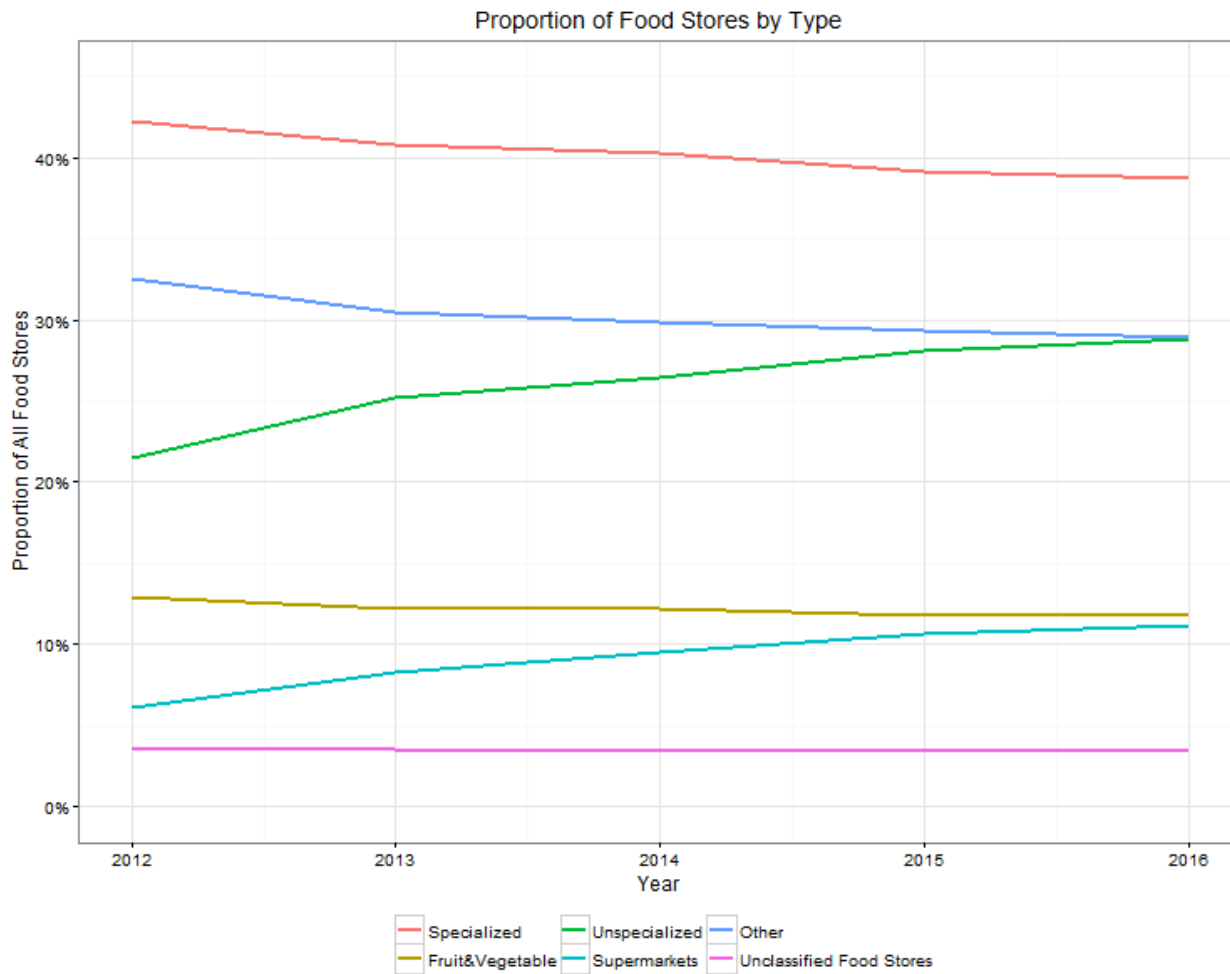


Figure 2.14. Exploratory Data Analysis for Aim 2: Trends in the Average Number of Food Stores by Type per census section

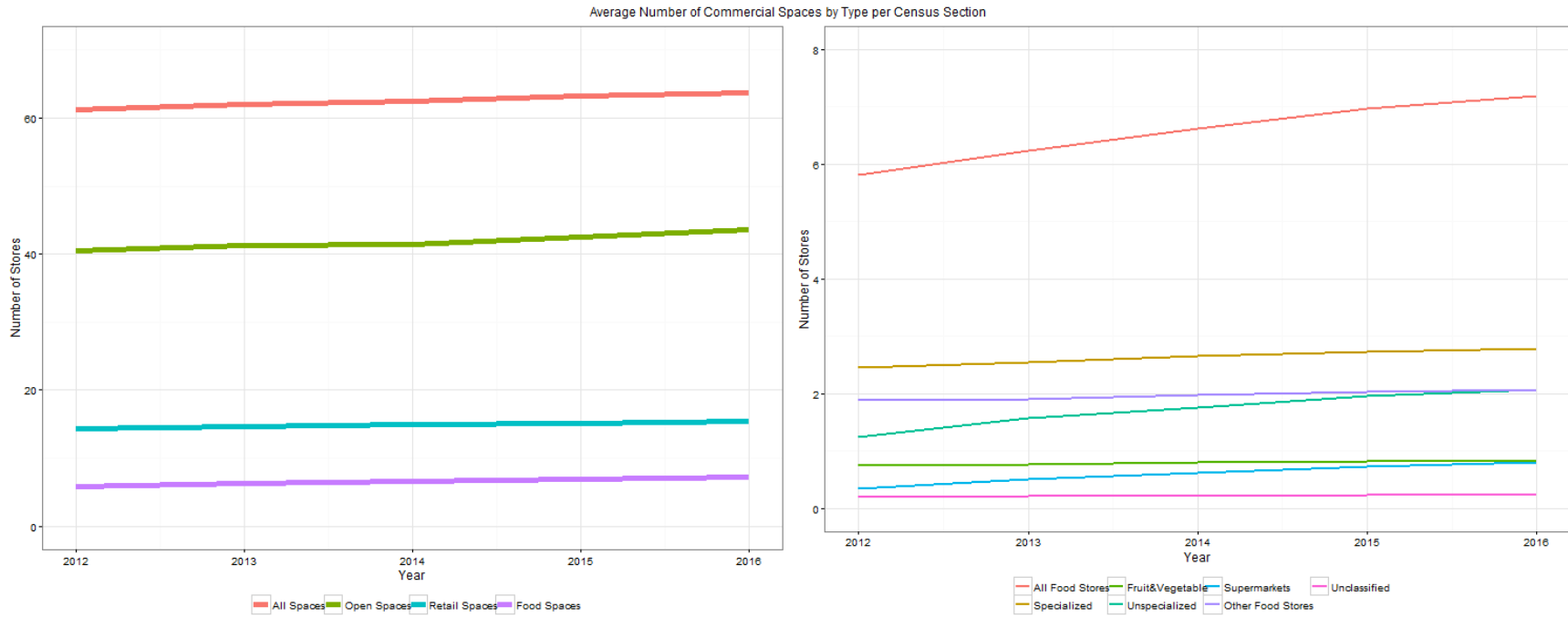


Figure 2.15. Exploratory Data Analysis for Aim 2: Trends in the Median Number of Food Stores by Type per census section

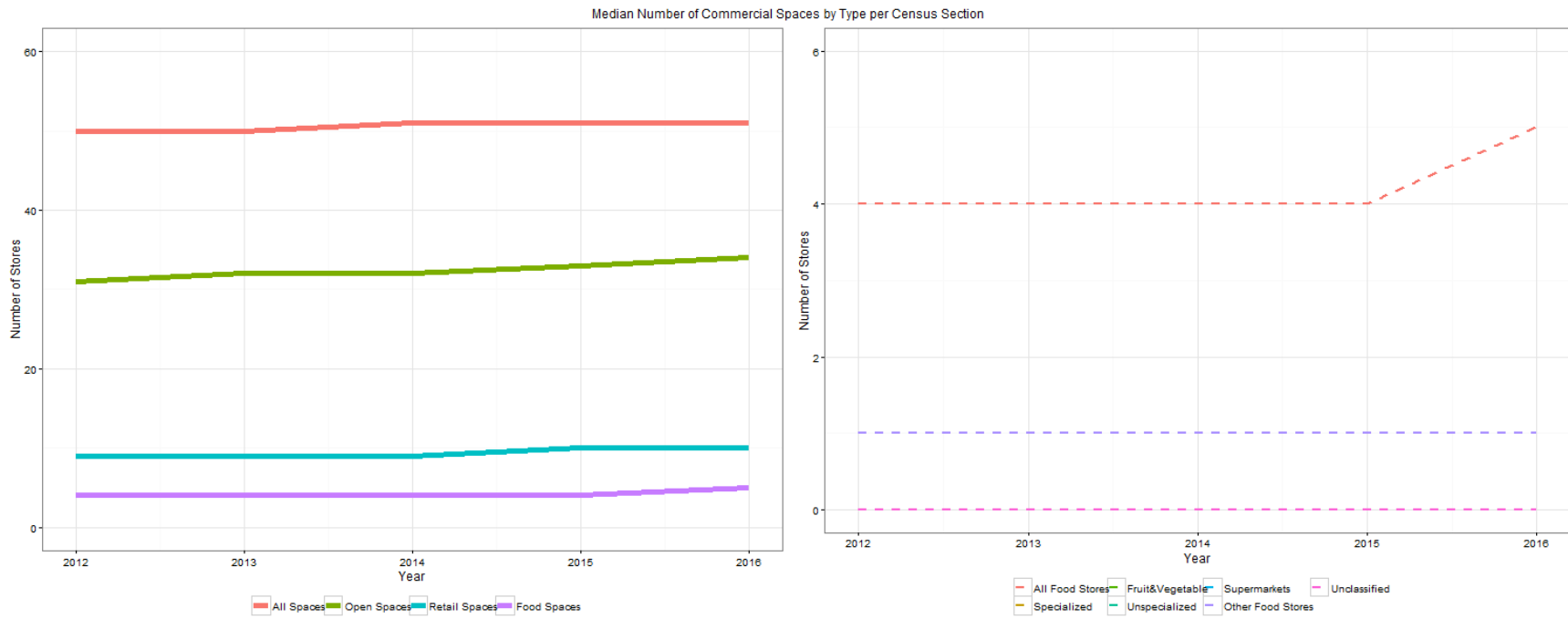


Figure 2.16. Exploratory Data Analysis for Aim 3: Map of the Study Area within Madrid and the Health Centers included in the Study

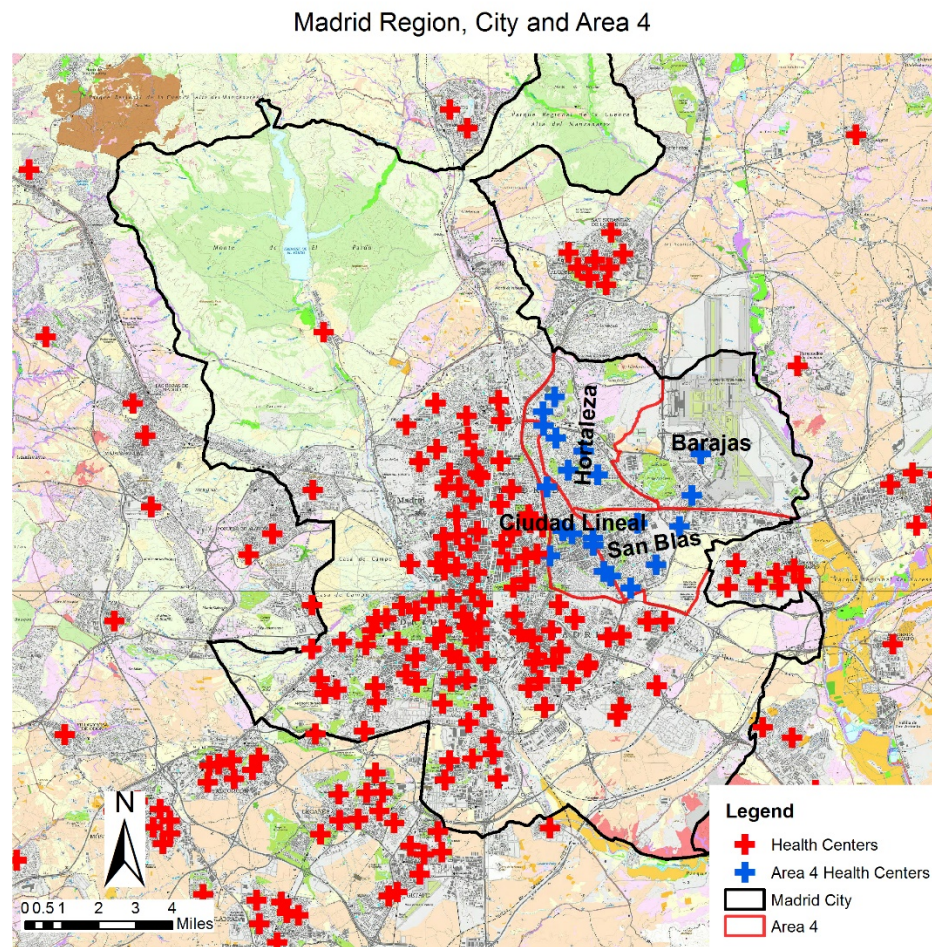


Figure 2.17. Exploratory Data Analysis for Aim 3: Distribution of key sociodemographic and socioeconomic variables in the four districts as compared to the entire city of Madrid

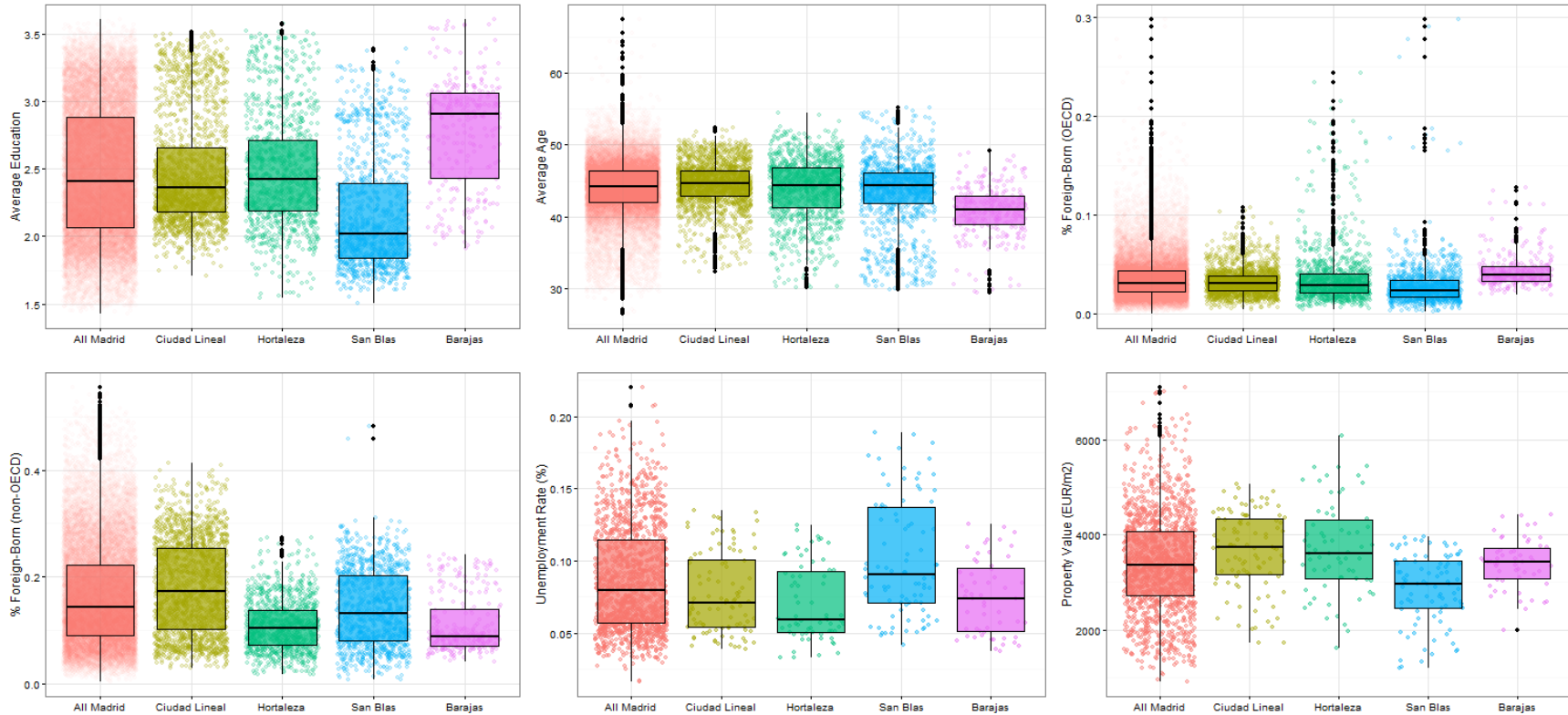
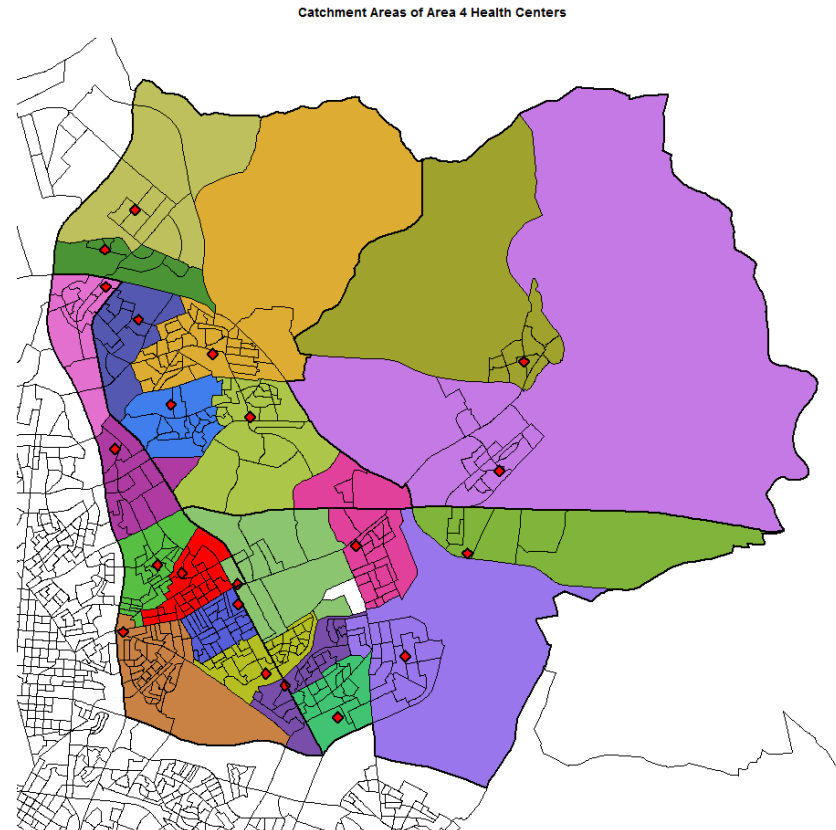


Figure 2.18. Exploratory Data Analysis for Aim 3: Map of the empirical catchment areas of the centers in 2013



Footnote: Empirical catchment areas are defined as: if at least half the patients from a census section come from a single health center, that census section is part of the catchment area of that center.

References

- Bilal, U., Diez, J., Alfayate, S., et al. (2016). Population cardiovascular health and urban environments: the Heart Healthy Hoods exploratory study in Madrid, Spain. *BMC medical research methodology*. 16, 104.
- Cuadra, C.B. (2011). Right of access to health care for undocumented migrants in EU: a comparative study of national policies. *European journal of public health*. 22 (2), 267-271.
- Díez, J., Bilal, U., Cebrecos, A., 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.
- Eagleson, S., Escobar, F., 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), 185-200.
- Eagleson, S., Escobar, F., Williamson, I. (2003). Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems. *International Journal of Geographical Information Science*. 17 (2), 99-118.
- Flavián, C., Haberberg, A., Polo, Y. (2002). Food retailing strategies in the European Union. A comparative analysis in the UK and Spain. *Journal of Retailing and Consumer Services*. 9 (3), 125-138.
- Fotheringham, A.S., Wong, D.W.S. (1991). The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A*. 23 (7), 1025-1044.
- Glanz, K., Sallis, J.F., Saelens, B.E. (2015). Advances in physical activity and nutrition environment assessment tools and applications: recommendations. *American journal of preventive medicine*. 48 (5), 615-619.
- Organization, W.H. (2011). *International classification of primary care, (ICPC-2)*. Geneva.
- Rodríguez, E., De Puelles, P.G., Jovell, A.J. (1999). The Spanish health care system: lessons for newly industrialized countries. *Health policy and planning*. 14 (2), 164-173.
- Shannon, C.E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*. 5 (1), 3-55.

CHAPTER 3: MEASURING NEIGHBORHOOD SOCIAL AND ECONOMIC CHANGE FOR HEALTH STUDIES

Abstract

Background: Neighborhood change is a complex phenomenon representing from processes of residential mobility and the actions of actors external to the neighborhood. To study its consequences on health behaviors and outcomes, we developed a measurement model of neighborhood social and economic change for the city of Madrid (Spain) from 2005 to 2015.

Methods: We use a finite mixture modeling approach, where membership in discrete types of change in indicators of any kind can be modeled. Using data from several administrative sources we constructed a set of 60 indicators. Through an iterative approach, we built a model starting with a single year (2009), testing for the best number of types in terms of fit, entropy and concordance with *a priori* theoretical principles. We iteratively increased model complexity by adding more indicators, changing the covariance structure and, subsequently, more years of data.

Results: The final model had 4 types of neighborhood change, included 15 indicators, a socioeconomic status index and an indicator variable for pre-post housing crash as predictors of type membership. The four types included: (Type 1) areas with increased migration and diversity and decreased SES; (Type 2) areas with high residential mobility, especially of young educated people along with new housing developments; (Type 3) areas with increased SES and decreased diversity, along with housing renovations; and (Type 4) areas with more aging native-born people, low residential mobility and no new

constructions. Transitions between types were common in types 2 (94% changed types the next year) and 4 (76% changed), while stability was more common in types 1 and 3 (62% remained in the same type next year). Transitions between types 1 and 4 were more common than other potential transitions. Types 1 and 3 showed high geospatial clustering, while type 4 was distributed throughout the city. The association with current area socioeconomic, sociodemographic and urban form characteristics was complex and non-linear.

Discussion: These measurement models can offer novel opportunities for the study of the consequences of residential environment on health behaviors and outcomes.

Introduction

Justifying the study of Change

The study of neighborhood effects on health has generally relied on static measurements of neighborhood characteristics (van Ham et al., 2011). A shift from static to dynamic neighborhood effects studies is needed to overcome several challenges in the field (van Ham et al., 2012). The current status of a neighborhood is the result of historical forces and specific processes (drivers of neighborhood change). A static conceptualization of neighborhoods, abundant in the public health literature, may be naïve to these neighborhood change drivers. In order to create a measurement model of neighborhood change, we have taken into consideration three aspects of neighborhood change: the two processes, neighborhood selection and external shocks, and the specific temporal nature of neighborhood change. In the following section, we discuss these aspects and how they informed our measurement model.

Neighborhood Selection Processes

Most of the neighborhood change literature relies on studying the processes by which individuals (or households) get “selected” or distributed among neighborhoods, otherwise known as neighborhood selection processes. Bailey et al. (2013) divide selection forces into two broad categories: residential mobility (overall flows, or lack thereof, of people in and out of neighborhoods) and selective migration (focusing more on differential patterns of mobility by demographic or socioeconomic groups). According to Bailey et al., residential

mobility decisions are related to neighborhood perceptions (both satisfaction and subjective characteristics of the area) and neighborhood change (which they argue are better predictors than current neighborhood composition) (Bailey et al., 2013). A measurement model that is sensitive to residential mobility must, therefore, include indicators of neighborhood composition change (such as education or ethnicity) and indicators of perceptions (such as property value). It must also include actual measures of residential mobility, such as the number of people moving into or out of the area.

Selective migration, the second force behind neighborhood selection, is mostly driven by demographic characteristics of households, creating a demographic conveyor where age distributes individuals across the city (i.e.: young adults to poor neighborhoods and middle-aged adults to richer ones, while the elderly tends to be more immobile) (Bailey et al., 2013). Bailey et al. (2013) also highlight the importance of stayers in determining neighborhood change, through natural growth (differential fertility) and socioeconomic change. Indicators of selective migration must, therefore, include demographic variables (such as changes in age composition or age-specific mobility) and changes that relate to non-movers (such as employment status changes or changes in fertility or mortality).

External Shock Processes

The second set of processes behind neighborhood change are those initiated by external actors. Aabers (2013) proposes a theory of how these

processes work, charging against what he calls “natural neighborhood change theories” which assume that neighborhood change occurs organically exclusively through flows of people. He argues that there are a series of “abstract space makers” whose impact on neighborhood change must be examined. These include real estate developers and banks who determine future neighborhood change through their practices (i.e., predatory lending and redlining). Indicators of the actions of external actors include housing stock changes (amount of housing units available or construction of new housing), renovation of the housing stock, lending practices (mortgage numbers or characteristics), large changes in the labor market (plant closures or openings), natural disasters (fires) or man-made ones (riots), and large urbanistic projects (new parks or developments).

The Temporal Nature of Neighborhood Change

A third aspect of neighborhood change is its temporal nature, which can vary from long-term slow drifts (or lack thereof) to very quick changes. Meen et al. (2013) conducted a study of neighborhood change showing that both long-term changes (century-long) and rapid changes are possible simultaneously. We focused on short-term changes by seeking indicators available annually using measurement model that could capture discrete types of change instead of smooth trajectories.

The difference between turnover and trajectories

Most research on neighborhood change in the health literature has employed the paradigm of trajectories. For example, a report from the CARDIA study created discrete types of trajectories (Richardson et al., 2014) under the assumption that neighborhood change follows a linear trend (e.g., “increasing”, “decreasing”, “stable”). In opposition to the study of trajectories, recent research has shown the utility of studying discrete change. First, Meen (2013) described neighborhoods in London for over a century, finding two types of phenomena: a long-term trajectory and short-term discrete changes. Second, Lekkas et al. (2017) have proposed a framework based on the lifecourse of places to suggest that types (and transitions between them) offer important information in health studies. Based on these empirical findings and theoretical propositions, we suggest considering two dimensions in the study of change: the first dimension is the magnitude of change, including population turnover, the building of new housing or the razing of old housing; the second is the direction of such change, especially the socioeconomic trend (upwards, downwards, or stable). Upwards and downwards trends are, by definition, areas where change is occurring. Stable trajectory areas may be the result of two different phenomena: first, a stable area, with low residential mobility or housing turnover; and second, an unstable area that stays in equilibrium, with high residential mobility that creates residential turnover. These two dimensions could be respected by using a continuous latent variable (e.g., including two components in a principal

component analysis or two factors in a confirmatory factor analysis) but this would not honor the discrete nature of change.

Steps in building a measurement model

Drawing from these three aspects of neighborhood change, we built a measurement model to provide a quantification of neighborhood social and economic change for studies on the health consequences of residential environments. Our analysis was done in five steps. First, we decided on a latent variable model that could capture the type of neighborhood change we were interested in (discrete short-term change). Second, we conducted an extensive review of potential indicators that could be accommodated to our latent variable model. Third, we built and diagnosed our model. Fourth, we represented the temporal and spatial nature of neighborhood change. Fifth, we validated our model comparing the results with other known indicators of neighborhood change.

Methods

Study setting

Our study was conducted in the city of Madrid, Spain. The city is divided into 21 Census Districts, 128 Neighborhoods and around 2400 census sections. See Table 2.1 (Chapter 2) for details. The Census Section has an average population of around 1500 people and is the smallest area for which census and other data is available. Some census sections may have populations as low as 700 or as high as 3500. Unless indicated otherwise, all analyses were conducted at the census section level. Their boundaries are updated every year for election purposes and may result in a split or merging of census sections. Chapter 2 describes in detail how we dealt with changes in census section boundaries. In summary, we constructed a “common” set of census sections ($n=2272$) that were consistent during the entire study period. Due to data availability, our analysis was restricted to the years 2005 through 2015. Large economic changes and urban transformations happened in Spain in these years. The unit of analysis was the “common” census section – year observation, creating a total of $n=22,720$ observations (2272 “common” census sections * 10 years). While the data covers 11 years (2005 to 2015), there are only 10 transitions between time periods (e.g., 2005 to 2006, etc.).

Data Sources

We obtained indicators from the following data sources (Chapter 2 contains more detailed information on each source). First, the *Padron* is a

continuous census of the entire Spanish population used to organize social services. Available from 2004 through 2015, it is updated every month and contains data on age, sex, education and country of origin on all residents. The *Padron* has two main components: cross-sectional data on the entire population on January 1st; residential mobility data from every census section to each other census section, by the variables above. The cross-sectional component allows for the detection of changes total population and in the proportion of residents by age, education or country of origin, along with diversity in these proportions. The residential mobility component allows for the study of mobility balance (difference between inflow and outflow of people), mobility throughput (sum of flows), and specific mobility flows by age, education and country of origin.

Second, the Cadaster (*Catastro*) is a tax registry for all properties in the entire Spanish territory. This registry helps local governments collect property taxes based on area and land use, and contains data on surface area of each property, year of construction and renovation status. Retrospective data is available from 2002 onwards. The cadaster allows for the detection of changes in the amount of housing surface area available every year.

Third, unemployment data is collected by the *Servicio de Empleo Publico Estatal* (National Employment Service), and includes data on people registered as job seekers. Data is available from 2003 through 2015 and updated monthly; we used data for the month of July of each year. We calculated the unemployment rate by dividing the total number of registered job seekers /

population 16 to 64 years of age. The numerator (registered job seekers) may be underestimated as some people may not be registered as such. Nonetheless this registration is mandatory in order to receive unemployment benefits. Ideally the denominator would be the number people 16 to 64 seeking jobs (not everyone in that age span is currently seeking a job), but information on job seekers is unavailable at this small geographical level, and hence we underestimate unemployment rates.

Last, the Idealista Report is a yearly report published from 2002 onwards by Idealista (www.idealista.com), the largest online real estate company in Spain. It includes the average sale price of each property sold through their website, yearly, for each neighborhood of Madrid. We used the average sale price of each property per m² as a marker of property value.

Latent variable Model

Our goal was to construct a measurement model of the construct of neighborhood social and economic change. We elected to use finite mixture modeling to capture discrete types of neighborhood change (McLachlan and Peel, 2004). In summary, finite mixtures are generalizations of latent class analysis that consider that the distribution of any variable is a mixture of distributions originated from several underlying sub-populations. In our case, this means that the distribution of changes in a given set of variables is actually originated from K different types of neighborhoods that are changing in different

ways. A general finite mixture model mathematical model is shown in equation 3.1:

$$(3.1) \quad Y_{ij} = \sum_{k=1}^K \pi_k(w) F(y|\beta)$$

Here Y is a vector of indicators for every year (i) and census section (j) of observation. The response is a function of the sum, over all k types, of two components. First, a model for the characteristics of each type, following a function F ; these functions can, for example, be a multivariate normal distribution with correlations between indicators, or a binomial distribution. These type characteristics are summarized in terms of an average response (mean for continuous indicators, or probabilities for categorical indicators) and a dispersion in the case of continuous indicators (variance). The second component is a model for type membership, as a function π of covariates (or an intercept-only model where the probabilities of each type are unconditional). In summary, these models estimate parameters for the means and variances or probabilities for each k type, $k-1$ probabilities of type membership and $k-1$ coefficients for the effect of each covariate on type membership.

The advantages of finite mixture modeling include the ability to handle indicators distributed in several ways (discrete data, normally-distributed continuous data, etc.); the flexibility in the inclusion of correlations between indicators; and the availability of software to fit these type of models (e.g., MPLUS and R). Potential disadvantages are the lack of an absolute measure of

model fit and the challenges derived from selecting a number of types to model (McLachlan and Peel, 2004). Model quality in finite mixtures can be assessed through measures of fit (such as the BIC, lower is better) and measures of classification (Entropy, higher is better). Entropy ranges from 0 (all units have a $1/K$ probability of type membership to each type) to 1 (all units have a 0 probability of belonging to each type except for a single type with a probability of 1). McMahan and Peel (2004), recommend the use of both types of measure, independently) and in combinations through the Integrated Completed Likelihood-Bayesian Information Criterion (ICL-BIC) (McLachlan and Peel, 2004). Our goal was to minimize ICL-BIC and BIC itself, and to maximize Entropy.

Candidate Indicators

In order to develop the measurement model, we specified a list of candidate indicators that represented either the consequence of neighborhood selection or external shocks on the area (as detailed in the introduction) and were available at least at the neighborhood or census section level.

All indicators were expressed in a metric of change. They either represented a delta measure (value at time i – value at time $i-1$) or an inherent measure of change (e.g.: absolute number of people moving into the area in the previous year). In a sensitivity analysis, we examined the effect of using a measure of change averaged over two years (average of the change in the last two years or average of the inherent measure of change over the last two years).

The objective of the list of candidate indicators was to assess the feasibility of measuring neighborhood change at the census section level from 2005 to 2015 with sufficient time granularity. Moreover, it allows for an assessment of the methods needed to fill potential temporal or spatial gaps in the data. From a total of 60 potential indicators we selected 32 that were available from 2005 to 2015 and were non-redundant (see Table 2.2 in Chapter 2 for the complete list). From the list of 32 potential indicators, we selected 10 indicators that we considered most important (see bolded indicators in Table 3.1), and started the model building process.

Model building and diagnosis

The first step in this process was to standardize all indicators by centering by the mean and scaling by the standard deviation each year of the study. We did this in order to avoid types clustering around long term trends. All interpretations of coefficients of continuous variables is relative to each other unit of observation each year. All indicators that were either discrete or that had a very skewed distribution were categorized as described below (see Table 3.1).

We started building the model by fitting a finite mixture with the 10 fundamental indicators for a single year (2009). We then used initial model results to select the number of types. Baseline models had different mean and variances per type and allowed no covariances between indicators. Two sets of modifications were conducted iteratively. First, we allowed the estimation of some covariances between indicators based on the empirical correlations

between indicators weighted by type membership. Second, we constrained the variance of some indicators to be equal across types if the results from the fitted model showed similarities in the variance of each indicator in all types. After improving model fit and/or classification, we re-assessed the number of types.

Second, we added indicators from the list of final indicators (32 minus the 10 we started with), ordered based on their *a priori* importance. After the addition of indicators stopped increasing model fit/classification, we re-assessed the number of types, improved the model regarding the estimated covariances and variances and re-assessed the number of types.

Third, we added data (sequentially) for other years beyond 2009 and performing diagnostics of model classification (based on entropy). After the final model (with data from 2006 to 2015) was fitted, we included predictors of type membership (\mathbf{w} in equation 3.1). These predictors improve the classification of units of observation (increasing Entropy) and are useful for the use of this model in further regression analysis (Bray et al., 2015). We included two predictors in our measurement model.

First, during the model building procedure we observed that the type structure may be different from 2006-2009 as compared to 2010-2015. Models fitted with data from 2006-2009 showed better entropy than models with data from 2010-2015. We interpret this to reflect a shift in the underlying neighborhood change model. This may reflect the impact of the economic recession of 2008 that led to changes in the housing market (Ortega and

Peñalosa, 2012)). To account for this, our final model (with pooled data from 2006 to 2015) includes an indicator variable for period (2010-2015 vs 2006-2009) as a predictor of type membership, allowing type membership in the second period to differ; we re-assessed the most appropriate number of types after adding this indicator again.

Second, we also added an index of current socioeconomic status as a predictor of type membership. This variable was a composite index of four indicators not included in the main model: % people with low education, % people with high education, current property value and current unemployment. The two education variables were weighted down 50%, to make education, wealth (property value) and unemployment weight equally. The four variables were standardized annually.

Representation of results

Finite mixture models produce two kinds of useful estimators originated from the two components of equation 3.1 above. First, the parameters β estimated in the function F , which describe the characteristics of each type. Second, a set of posterior probabilities of membership to each type for each area-year of observation, conditional on covariates w (equation 3.1 above).

To display the description of each type in Table 3.3 and Figure 3.1 we showed the mean of continuous normally-distributed variables and the probability of each category of discrete variables.

Model validation

There is no criterion for of neighborhood change so construct validity must be examined by addressing the relationships between types of change and other variables associated with change. To provide a sense of the correlates of neighborhood change (and hence to better understand what neighborhoods are in each type of type) we produced a series of non-parametric exploratory plots of the posterior probability of being in each type of change against current socioeconomic status index, current mean age, current proportion of Spaniards, population density (residents / km²) and distance to city center (measured as the distance between the centroid of each census section and the center of the Puerta del Sol square) (Figure 3.4).

Analytic Procedures

All data management and plotting was conducted in R v3.3.0, while all finite mixture modeling was conducted in Mplus v7.4 through the use of the MplusAutomation package from R. Maximum Likelihood Estimation with Robust Standard Errors were used for all analysis.

Results

Neighborhood Social and Economic Change Final Model

Table 3.1 shows the final list of 15 indicators. After the iterative process described in the methods section, we ended up with a 4-type model with 146 parameters. These included 76 type-specific means, 48 type-specific variances (one for each type-continuous indicator), 13 covariances, and 9 parameters for type membership (3 intercepts, 3 parameters for SES and 3 parameters for epoch). An entropy of 0.839 was achieved in the final 4 type model. In the following sections, we describe the results of the neighborhood social and economic change measurement model. We first describe the types and then follow with a description of the temporal and spatial patterns and the association of change types with known socioeconomic and sociodemographic indicators.

Neighborhood Social and Economic Change Types

Figure 3.1 displays means (converted to probabilities for the discrete variables) for all indicators in each of the four types. As a reminder, each indicator was standardized every year of the study. This means that the coefficients must be interpreted relative to all other areas every year. That is, in a year in which there was an overall decrease in property value, then a higher value (a relative increase in property value) would mean that the area did not lose as much property value as other areas. A lower value (a relative decrease in property value) would mean that the area lost more property value than other areas that year. Appendix 3.1 through 3.4 shows detailed information on the

model, including these means, probabilities, the variances, covariances and type membership predictors. What follows is a description of each of these four types.

Type 1 areas are changing by having the highest increase in the proportion of non-OECD migrants, with a small decrease in the average level of education but the strongest decrease in property value and strongest increase in unemployment. While these areas are losing population, the volume of people moving through these areas (mobility throughput) is high. The composition of migrants is mostly young people and people from non-OECD countries. The increases in diversity (both in education and country of origin) is the highest in all types of change. The amount of new housing and housing renovations is moderate, and along with the slight changes in total population this leads to stability in housing availability per person. In summary, Type 1 areas are changing through increased migrants from poor countries, a decrease in SES indicators and an increase in diversity. Type 1 areas are also the most prevalent, as 46% of the census section-year observations are in this type.

Type 2 areas are changing by showing the strongest decrease in average age and the strongest increase in average education level. This increase in education level is not followed by an increase in property value or decrease in unemployment, indicating that these areas are changing demographically, more than socioeconomically: there is an influx of younger people with college degrees that have not yet moved upwards socioeconomically. The increase in population of these areas is the highest, along with the number of people moving through

them (residential mobility throughput). These areas also display the highest probability of new housing and a high probability of housing renovations. This, along with the increase in total population, creates a heterogeneity within these areas, as some show a decrease in the amount of housing available per person while others show an increase. In summary, Type 2 areas are recent developments with new housing where young educated people are moving in. Type 2 areas are the least prevalent, as only 3% of the census section-year of observation are in this type.

Type 3 areas are changing through the strongest increases in property value and decreases in unemployment. This is not necessarily accompanied by the highest increase in education, showing that these areas are changing socioeconomically. These areas also display the strongest increase in the proportion of people from OECD countries, and the strongest decrease in the proportion of people from non-OECD countries, highlighting the migration of people from other developed countries. The degree of residential mobility is moderate and is mostly composed of adults or the elderly. These changes lead to the strongest decreases in diversity, both by education and by country of origin. In terms of housing, these areas are showing the highest probability of housing renovations. In summary, Type 3 areas are showing increases in socioeconomic markers through upwards mobility, and are displaying markers of high SES segregation; the increased probability of housing renovations is

marking potential gentrification. Type 3 areas are present in 27% of the census section-year observations.

Type 4 areas are changing through the highest increase in average age and decreases in the proportion of migrants from any area. The average education level shows the strongest decrease but is not followed by a decrease in property value or increase in unemployment, meaning these areas are changing demographically (older people with a lower likelihood of college degrees). Type 4 areas have the strongest decrease in population, with the lowest degrees in residential mobility and decrease in diversity by education. Moreover, these areas have the lowest probability of both housing renovations and new housing. In summary, type 4 areas are having a demographic shift towards older Spaniards, with low residential mobility and new/renovated housing. Type 4 areas are present in 24% of the census section-year observations.

In the sensitivity analysis using two years of change (instead of changes over one year), the type structure looked similar to the description above.

Spatial Distribution of the 4 Types of Neighborhood Change

Figure 3.2 shows the spatial distribution of neighborhood change types in two epochs (2006-2009 and 2010-2015). Three spatial patterns are evident from these maps. First, Type 4 is distributed more homogeneously across the city, with the exception of some clusters in the Northern part of the city in the first Epoch and in the West in the second Epoch. Both areas represent more

suburban and less dense developments. The second spatial pattern is the scattered and rare presence of type 2 through the city. As explained above Type 2 is the least common type and neighborhoods do not stay in it for long.

The last spatial pattern is the non-overlapping clustered distribution of Type 1 and Type 3. Type 1 is present in Epoch 1 in pockets around the city, especially in the inner-Northern part of the city, the Southern periphery and the East. Some of these pockets are persistent in the second Epoch, especially the ones in the North and East. Type 3 follows a similar pattern but in different areas. In particular, the first epoch shows a high concentration of type 3 neighborhoods in Northeast Madrid, West, and some parts of the South. Some areas of downtown Madrid have visible pockets of type 3 census sections. The second epoch (2010-2015) shares some of these areas, except for a lack of a clear presence in Southeastern Madrid and the new presence of type 3 areas in the Southwest.

Transitions between Types of Neighborhood Change

Table 3.4 shows the transition numbers and probabilities between each type of neighborhood social and economic change. Changes between types are common but differ across types. Type 1 and type 3 areas have a higher likelihood of remaining in the same area the next year (62%), while type 4 areas have a lower probability (31%) and type 2 areas do not remain in the same type the year after (6%). Transitions between Type 1 and Type 4 are more common, as 24% of the areas in Type 1 transition to Type 4 while 47% of the areas in

Type 4 transition to Type 1. From Type 3 transitions are similarly likely to both Type 1 and Type 4. From type 2, transitions to other types are common (94% change types the next year), and are more likely to type 1 (46%). In summary, areas in types 1 and 3 tend to stay in the same type, while it is not common for a neighborhood in type 2 or 4 to stay in the same type for more than one year.

Construct validity: Neighborhood Social and Economic Change and Socioeconomic Status

The construct validity of our measurement model can be examined by exploring how the typology we identify corresponds to other known features of places (such as SES, demographic profile and urban form). Figure 3.3 shows the association between the probability of belonging to each type and the current levels of the index of socioeconomic status, average age, proportion Spaniards and population density. Regarding the associations with socioeconomic status there's a gradient from Type 1 areas (lowest SES), Type 2 and 4 which are distributed thorough the spectrum, and Type 3 areas (highest SES). The main differences regarding current average age can be found between type 3 areas, who tend to be on the extremes (lowest and highest average age) and type 4 areas (around the middle of the distribution). The current proportion of Foreign-born people also differs by type of change, as Type 1 and Type 4 areas are on the higher end of the distribution while type 2 areas are on the lowest end. Population density is highest in type 1 and type 4 areas and lowest in type 3 areas. Regarding location within the city, type 1 areas tend to be further away

from the city center, especially at a distance of between 4 and 6km to the Puerta del Sol, while both type 3 and especially type 4 areas are in both extremes of the distribution, in areas close to the city center ($<2\text{km}$) and in areas further away ($>7.5\text{km}$).

Discussion

Our study has obtained data from several sources to build and validate a new measurement model of types of neighborhood social and economic change using a finite mixture modeling approach. We obtained four types that reflect: (Type 1) areas with increased migration and economic/demographic diversity and decreased SES; (Type 2) areas with high residential mobility, especially of young educated people, along with new housing developments; (Type 3) areas with increased SES and decreased diversity, along with housing renovations; and (Type 4) areas with more aging Spaniards, low residential mobility and no new construction. In essence, Type 1 and Type 3 areas show opposing trends in terms of SES and migrants while Type 2 and Type 4 areas show opposing trends in residential mobility and housing.

Our model has been able to tap sources of change that are independent (or complementary) to contextual socioeconomic status, the usual indicator of change. Previous analyses (Grigsby-Toussaint et al., 2010; Le-Scherban et al., 2014; Rummo et al., 2016; Wing et al., 2016) of neighborhood trajectories in the health literature that look at socioeconomic status would have conflated types 2 and 3, as neighborhoods with increasing socioeconomic status trajectories. As we have seen in our study, these two types represent neighborhoods in entirely different stages, as one represents areas with new housing developments and high residential mobility, while the other shows an increase in property value and housing renovations, decreases in unemployment and education diversity.

Moreover, in analysis using cross-sectional SES some areas in Type 1 and 4 would have been conflated together as low SES areas; as we have shown, these areas are in entirely different stages of change (Type 1 areas are diversifying, with more migrants and a longitudinal trend towards lower SES; Type 4 areas are losing population and aging, with very low residential mobility and no clear longitudinal SES trend).

The use of change in current markers of socioeconomic status is more common across the neighborhood change literature (Diez Roux et al., 2001; Gary-Webb et al., 2011; Singh, 2003). Two of our sets of indicators performed especially well in differentiating between neighborhoods and are not commonly used in the literature: mobility throughput (sum of residential mobility inflows and outflows) and change in diversity. To our knowledge, this is one of the first studies to use mobility throughput as an indicator of neighborhood change. Given the uniqueness of the Padron in Spain, that allows us to measure how many people enter and exit the neighborhood every year, we can measure whether a lack of change in population was due to a complete replacement (X people enter, X people exit) or due to a lack of mobility (0 people enter, 0 people exit). Areas with indicators of higher residential mobility had also a decreasing average age, potentially reflecting the demographic conveyors of Bailey et al. (2013), and points towards the utility of measuring change in the age composition in neighborhood studies. The second set of indicators, diversity, also provided a crisp differentiation between change types. By using Shannon's Entropy

(Shannon, 2001), a measure that is scale-independent, we were able to measure how diverse each area was in terms of age, country of origin and education. Only the last two were in the final list of indicators, but they differentiated well between areas gaining diversity (and generally losing socioeconomic status) and areas losing diversity (and generally gaining socioeconomic status). That areas with increased diversity lost SES and areas with decreased diversity gained SES potentially highlights an increase in the intensity of segregation phenomena, reported before for Madrid from 2001 to 2011 (Leal and Sorando, 2015).

Previous epidemiologic studies that measured neighborhood change have focused almost entirely on built environment characteristics, as has been the case with the MESA and the CARDIA neighborhood sub-studies (Hirsch et al., 2016). An exception to this is an analysis (Rummo et al., 2016) of the CARDIA study where finite mixture models were employed to classify neighborhoods according to age trajectory changes which were then regressed on several socioeconomic predictors.

Strengths and limitations

This study has several strengths. First, we used data from a multitude of data sources, leading to a very comprehensive list of indicators. Some of these data sources are universal in nature, removing any concern for sampling errors or other biases related to selection. Second, the spatial unit of analysis for our model and for most of the data was the census section, a very small area (n~1500 people) that allowed us to study change with precision. Third, we were

able to study year-to-year changes, removing some issues in the previous studies of neighborhood change that rely on decennial censuses. Fourth, finite mixture modeling is a robust but flexible approach to latent variable modelling, allowing for several types of variables to be included.

This study has some limitations. First, most of our data sources were collected for administrative purposes, which can lead to lower data quality than research-grade data collection. Nonetheless, the advantages outweigh potential errors in data collection. Second, while our unit of analysis is granular in space and time, some neighborhood change phenomena may occur at higher levels in both dimensions. Some changes may be city-wide (or even nation-wide, like recessions); nonetheless and given that our interest is in estimating neighborhood effects, we are interested in the spatial distribution of change, and changes that are affecting an entire city are therefore out of our scope. Time-wise, while some changes take decades (as studied by Meen et al. (2013)) some may either occur very rapidly (see Type 2 of change), and even long-term changes may be captured by the current trajectories (see our Type 1 and 3 of change). The concerns regarding differing levels of analysis for space and time have been well studied in geography (Cheng and Adepeju, 2014; Fotheringham and Wong, 1991), and future analysis should consider the effects of differing levels on inferences. Third, regarding the measurement model approach, the usual limitations of latent variable models apply (McLachlan and Peel, 2004). While we checked for the potential violations of the main assumptions of the

model (measurement invariance, conditional independence) and found them ignorable, there remains the possibility of undetected violations.

Conclusions

Neighborhood change is a complex exposure, as it originates from several sources, including neighborhood selection and external shocks. We propose a measurement model that encompasses indicators in all domains, using the city of Madrid as an example. Future research will study how these types of change relate to health behaviors and outcomes.

Tables

Table 3.1. Indicators of Neighborhood Social and Economic Change

Indicator	Operationalization*	Source
Δ Mean Age	% in each 5-year age bin in * Mid Point of age bin	Padron
Δ Proportion Foreign-Born in non-OECD	# Born in non-OECD Country / Total population	Padron
Δ Proportion Foreign-Born in OECD	# Born in OECD Country / Total population	Padron
Δ Mean Education Level	% in each education group (4 groups) * 1 to 4 (for no official studies, primary education, secondary education, university education)	Padron
Δ Property Value	Average price sale of all housing units sold in EUR/sqm.	Idealista Report
Δ Unemployment Rate	# Registered for Unemployment / Population 16 to 64	Unemployment Services
Any Renovation	Any renovation conducted in that year	Cadastre
New Housing	Any new housing built in that year	Cadastre
Δ Housing Space / Person	Total Housing Space / Total Population	Cadastre and Padron
Δ Total Population	Total Population	Padron
Mobility Throughput	# People Incoming to the area during time t-1 / Total People in the Area at time t-1 + # People Outgoing from the area during time t-1 / Total People in the Area at time t-1	Padron
Mobility Throughput of people aged 25 or below	% below 25 years of age in incoming people during time t-1 / % below 25 years of age at time t-1 + % below 25 years of age in outgoing people during time t-1 / % below 25 years of age at time t-1	Padron
Mobility Throughput of Foreign-Born (non-OECD)	% foreign-born in incoming people during time t-1 / % foreign born at time t-1 + % foreign-born in outgoing people during time t-1 / % foreign born at time t-1	Padron
Δ Education Diversity	- sum of [ln(proportion in each education group) * proportion in each education group]	Padron
Δ Country of Origin Diversity	- sum of [ln(proportion in each country of origin group) * proportion in each country of origin group]	Padron

Footnote:

*: All Δ (change) Indicator were operationalized as Indicator at time t – Indicator at time t-1. This column shows the operationalization of the Indicator itself.

Bolded indicators were included in the initial list of 10 indicators.

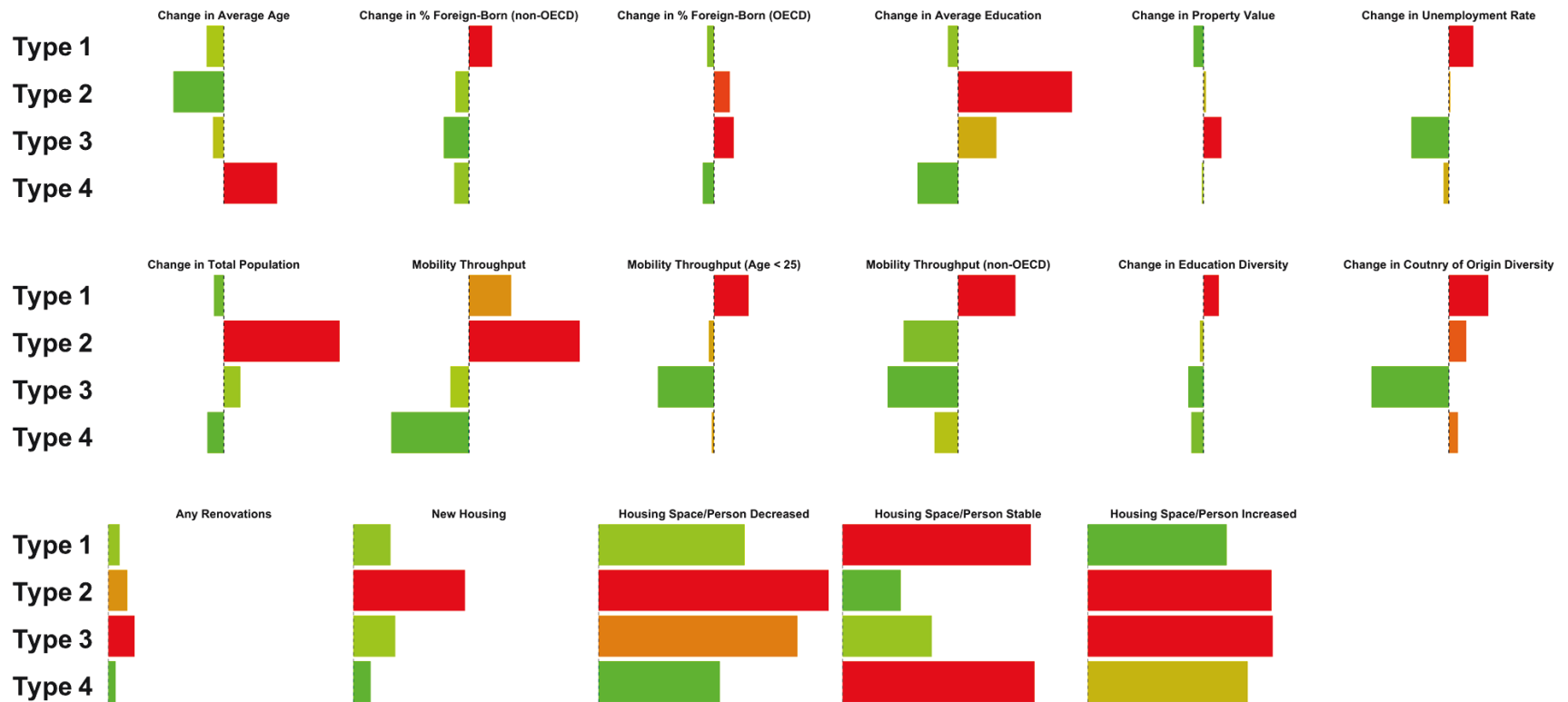
Table 3.2: Transitions Between Types of Neighborhood Change

		Year + 1						Year + 1			
(A)		Type 1	Type 2	Type 3	Type 4	(B)		Type 1	Type 2	Type 3	Type 4
Year	Type 1	5822	281	1057	2212	Year	Type 1	62%	3%	11%	24%
	Type 2	292	37	179	129		Type 2	46%	6%	28%	20%
	Type 3	974	190	3425	967		Type 3	18%	3%	62%	17%
	Type 4	2292	142	935	1514		Type 4	47%	3%	19%	31%

Footnote: Panel (A) are the sum of all census-section years of observations. The diagonal of the matrix are areas that stayed in the same type of neighborhood change the next year, while off-diagonals are transitions to other types. Panel (B) is the same data but converted to probabilities that sum to 1 in every row (i.e.: for all areas in Type 1 in year t, there's a 62% probability of staying in that type the next year, and a 3%, 11% and 24% probability of transitioning to type 2, type 3 or type 4).

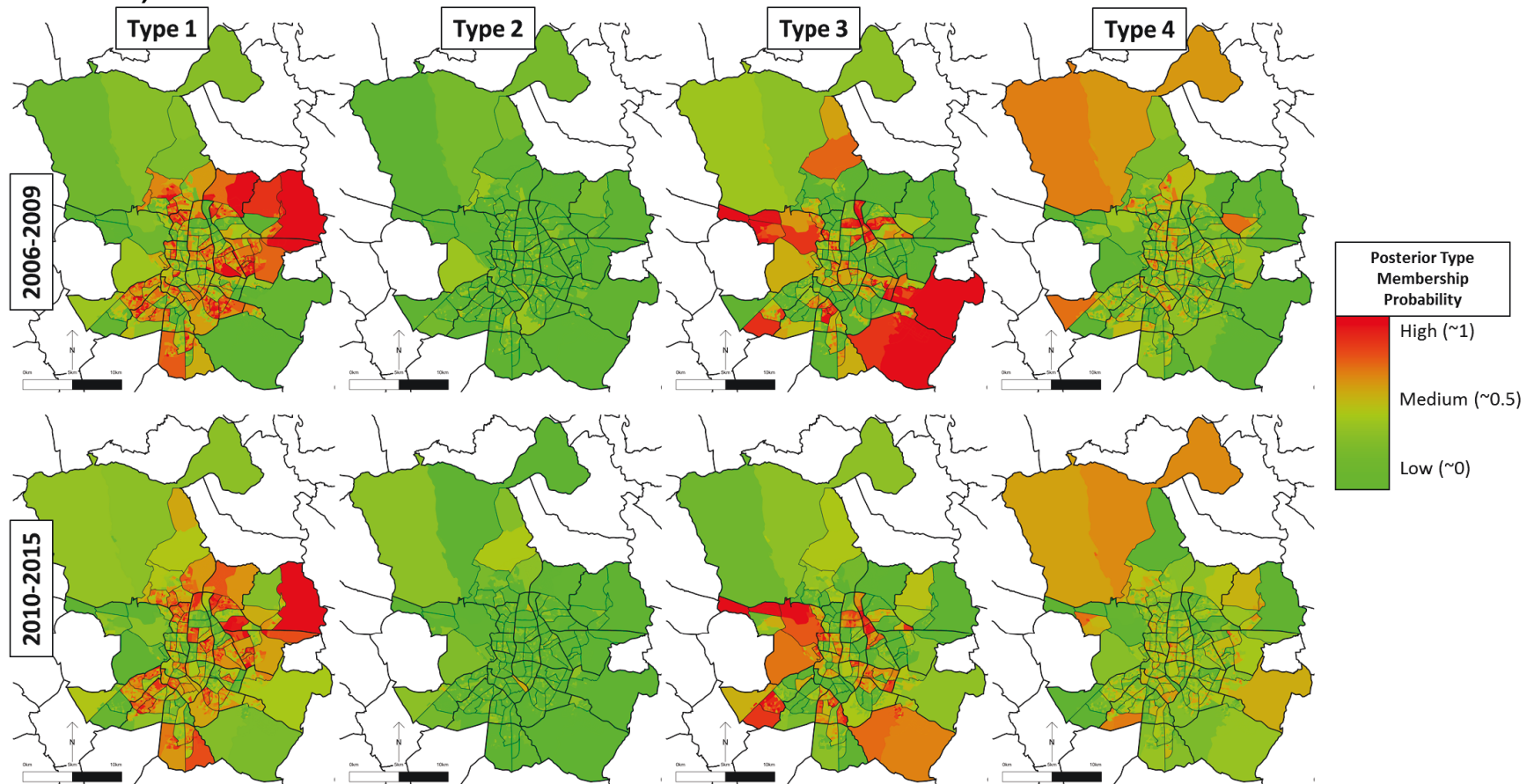
Figures

Figure 3.1: Characteristics of each type of neighborhood social and economic change.



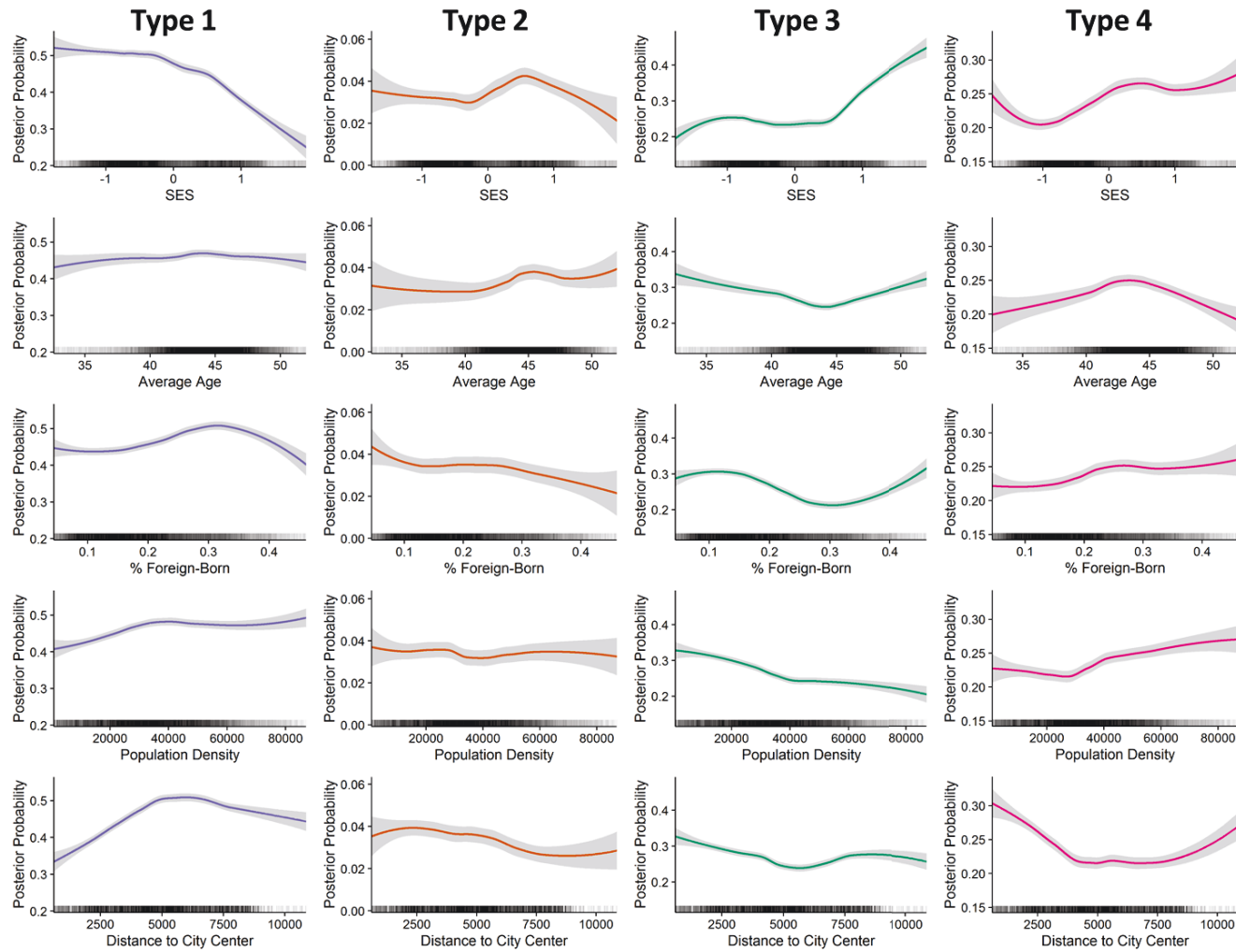
Footnote: graphical representation of Table 3.3. Darker bars are more positive values, while lighter bars are more negative values. In the case of categorical indicators (bottom row), all values are positive, so lighter values represent lower probabilities while darker values represent higher probabilities.

Figure 3.2: Spatial Distribution of Neighborhood Social and Economic Change Types by Epoch (2005-2009 and 2010-2015)



*Footnote: posterior type membership probabilities were averaged for all years in each epoch.

Figure 3.3. Association of Neighborhood Social and Economic Change Types and Current Levels of Socioeconomic Status, Average Age, Proportion Foreign-Born and Population Density.



Footnote: The Y-axis is the posterior probability of type membership to each type (type 1 in the first column, type 2 in the second column, etc.). The X-axis is the current value of a given variable (e.g.: SES index in the first row, average age in the second row, etc.), and the thin lines adjacent to the axis are the distribution of each variable (one line per census section-year observation). The smoothed line represents a lowess non-parametric estimator of the association of type membership with each variable.

References

- Aalbers, M.B. (2013). How Do Mortgage Lenders Influence Neighbourhood Dynamics? Redlining and Predatory Lending. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research* pp. 63-85). Dordrecht: Springer Netherlands.
- Bailey, N., Barnes, H., Livingston, M., Mclennan, D. (2013). Understanding Neighbourhood Population Dynamics for Neighbourhood Effects Research: A Review of Recent Evidence and Data Source Developments. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research* pp. 23-41). Dordrecht: Springer Netherlands.
- Bray, B.C., Lanza, S.T., Tan, X. (2015). Eliminating Bias in Classify-Analyze Approaches for Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*. 22 (1), 1-11.
- Cheng, T., Adepeju, M. (2014). Modifiable Temporal Unit Problem (MTUP) and Its Effect on Space-Time Cluster Detection. *PLoS one*. 9 (6), e100465.
- Diez Roux, A.V., Merkin, S.S., Arnett, D., et al. (2001). Neighborhood of residence and incidence of coronary heart disease. *The New England journal of medicine*. 345 (2), 99-106.
- Fotheringham, A.S., Wong, D.W.S. (1991). The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A*. 23 (7), 1025-1044.
- Gary-Webb, T.L., Baptiste-Roberts, K., Pham, L., et al. (2011). Neighborhood socioeconomic status, depression, and health status in the Look AHEAD (Action for Health in Diabetes) study. *BMC public health*. 11, 349.
- Grigsby-Toussaint, D.S., Lipton, R., Chavez, N., Handler, A., Johnson, T.P., Kubo, J. (2010). Neighborhood Socioeconomic Change and Diabetes Risk: Findings from the Chicago Childhood Diabetes Registry. *Diabetes Care*. 33 (5), 1065-1068.
- Hirsch, J.A., Meyer, K.A., Peterson, M., et al. (2016). Obtaining Longitudinal Built Environment Data Retrospectively across 25 years in Four US Cities. *Frontiers in Public Health*. 4 (65),
- Le-Scherban, F., Albrecht, S.S., Osypuk, T.L., Sanchez, B.N., Diez Roux, A.V. (2014). Neighborhood Ethnic Composition, Spatial Assimilation, and Change in Body Mass Index Over Time Among Hispanic and Chinese Immigrants: Multi-Ethnic Study of Atherosclerosis. *American journal of public health*, e1-e9.
- Leal, J., Sorando, D. (2015). Economic crisis, social change and segregation processes in Madrid. In T. Tammaru, M. van Ham, S. Marcinićzak, & S. Musterd (Eds.), *Socio-Economic Segregation in European Capital Cities: East Meets West*. Taylor & Francis.

- Lekkas, P., Paquet, C., Howard, N.J., Daniel, M. (2017). Illuminating the lifecourse of place in the longitudinal study of neighbourhoods and health. *Social Science & Medicine*,
- McLachlan, G., Peel, D. (2004). *Finite mixture models*: John Wiley & Sons.
- Meen, G., Nygaard, C., Meen, J. (2013). The Causes of Long-Term Neighbourhood Change. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research* pp. 43-62). Dordrecht: Springer Netherlands.
- Ortega, E., Peñalosa, J. (2012). The Spanish economic crisis: key factors and growth challenges in the euro area. Banco de Espana Occasional Paper (1201),
- Richardson, A.S., Meyer, K.A., Howard, A.G., et al. (2014). Neighborhood socioeconomic status and food environment: A 20-year longitudinal latent class analysis among CARDIA participants. *Health & Place*. 30, 145-153.
- Rummo, P.E., Hirsch, J.A., Howard, A.G., Gordon-Larsen, P. (2016). In Which Neighborhoods Are Older Adult Populations Expanding? Sociodemographic and Built Environment Characteristics Across Neighborhood Trajectory Classes of Older Adult Populations in Four U.S. Cities Over 30 Years. *Gerontology and Geriatric Medicine*. 2,
- Shannon, C.E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*. 5 (1), 3-55.
- Singh, G.K. (2003). Area deprivation and widening inequalities in US mortality, 1969-1998. *American journal of public health*. 93 (7), 1137-1143.
- van Ham, M., Manley, D., Bailey, N., Simpson, L., Maclennan, D. (2011). *Neighbourhood Effects Research: New Perspectives*: Springer Netherlands.
- van Ham, M., Manley, D., Bailey, N., Simpson, L., Maclennan, D. (2012). *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research*: Springer Netherlands.
- Wing, J.J., August, E., Adar, S.D., et al. (2016). Change in Neighborhood Characteristics and Change in Coronary Artery Calcium. A Longitudinal Investigation in the MESA (Multi-Ethnic Study of Atherosclerosis) Cohort. 134 (7), 504-513.

**CHAPTER 4: NEIGHBORHOOD SOCIAL AND ECONOMIC CHANGE AND
RETAIL FOOD ENVIRONMENT CHANGES IN MADRID (SPAIN)**

Abstract

Background: The relationships between neighborhood characteristics, the retail food environment and health behaviors and outcomes are complex. To shed light on the first part of this association, we explored associations between neighborhood social and economic change and change in the retail food environment in Madrid (Spain) from 2012 to 2016.

Methods: We used a measurement model of neighborhood social and economic change using finite mixture models with 15 indicators and 4 types of neighborhood change. We classified areas by their most likely change type in the previous 5 years, or labeled them as neighborhoods in transition if no type was highly likely (posterior probability < 0.8). We classified and geocoded all food stores for a period of 5 years (2012 to 2016) using a universal retail spaces census from the City Government of Madrid. Stores were classified as 1) any food store, 2) specialized small stores, 3) supermarkets or 4) fruit and vegetable stores. We used a multinomial logistic regression model with robust clustered standard errors to examine the association between losses (of one or more stores) or gains (of one or more stores) compared to stability in the number of stores by type of neighborhood change.

Results: Madrid showed a dynamic food environment, where at least one third of the areas saw changes in the number of stores from year to year. Overall there was an increase in the number of all food stores, with stability or potential decreases in the number of small specialty stores, especially fruit and vegetable

stores. These changes differed by neighborhood change type: type 1 (new migrants from poor countries, decreasing SES, increased diversity), type 4 (very low residential mobility, aging, no new housing), and neighborhoods in transition saw significant increases in the odds of gaining a supermarket (OR=1.39, 1.80 and 1.31, respectively). The same areas saw increases in the odds of losing small specialty stores. Alternatively, type 3 areas (increased property value and decreased diversity) have an increasing presence of small specialty stores and decreasing supermarkets.

Discussion: These results highlight potential increases in disparities in the food environment, as areas with increased property value move towards small specialized stores where the availability of unhealthy foods is lower and the availability of healthy foods may be higher. Future research should study the health consequences of these changes in urban environments.

Introduction

The Local Food Environment is a Mass-Influence on Diet

Population dietary patterns are shaped by mass-influences that differ across populations or within the same population over time (Díez et al., 2016; Rose, 1985). As a contextual factor, the Local Food Environment (LFE) affects everyone living in an area and therefore qualifies as a potential mass-influence on diet (Rose, 1985). The LFE is defined as the set of contextual aspects of the local environment that have the potential to influence dietary behaviors (Franco et al., 2016). The components of the local food environment include the location and accessibility of food stores and the availability of healthy foods within them (Glanz et al., 2005, 2007). Changes in these factors have the potential to affect population dietary patterns so understanding what causes changes in food stores (and their content) may be a feasible way to improve diet (Story et al., 2008).

Neighborhood Change is Understudied

Studies of neighborhood social and economic change are mostly absent in the public health literature, and those that study changes in neighborhood characteristics generally use residential mobility of participants as the instrument to study these changes (Jokela, 2014, 2015; Ludwig et al., 2011; White et al., 2016). However, due to the “stickiness” of neighborhood characteristics, most people that move relocate to similar areas (Glass and Bilal, 2016); moreover, most of the population of an area does not move in a given year (Glass and Bilal, 2016). Therefore, the lack of studies looking at change in neighborhood

characteristics challenges our ability to study contextual characteristics, such as the local food environment, and their effect on health. Cross-sectional studies have shown evidence for a strong patterning of the food environment by socioeconomic and sociodemographic characteristics (Franco et al., 2008; Moore and Diez Roux, 2006; Morland et al., 2002). There is also strong evidence for a difference in how food environments change by levels of socioeconomic neighborhood characteristics (Cobb et al., 2015a; Rummo et al., 2016a; Rummo et al., 2016b). Moreover, research conducted in the CARDIA study has shown that some socioeconomic trajectories of neighborhoods are associated with differential patterns of change in the food environment (Richardson et al., 2014). However, research on longitudinal changes in neighborhood characteristics is usually restricted to socioeconomic factors and race/ethnicity and does not consider other factors such as housing changes or residential mobility.

A Measurement Model of Neighborhood Social and Economic Change

In Chapter 3 we described a finite mixture model that estimated four types of Neighborhood Social and Economic Change. These included: (Type 1) a type of neighborhoods with relative increased proportion of migrants from poor countries, decreased education level and property value along with increased unemployment, a moderate degree of residential mobility and an increase in diversity; (Type 2) a type of neighborhoods with relative decreased average age and increased education level, the highest degree of residential mobility along with new housing constructions; (Type 3) a type of neighborhoods with relative

increased education level and property value, decreased unemployment, a moderate degree of residential mobility and housing renovations; and (Type 4) a type of neighborhoods with relative increased average age, the lowest degree of residential mobility with population loss and no new housing or renovations. Figure 4.1 shows a description of the four types of neighborhood social and economic change and Appendix 4.1 shows the distribution of types over time.

Neighborhood Change and Food Environment Change: Mechanisms

Previous research has shown that more disadvantaged areas tend to have smaller stores as opposed to supermarkets (Dunkley et al., 2004). Gentrifying areas show more 'boutique' stores and large chain-stores (Zukin et al., 2009), as opposed to small locally owned stores. Large chain-stores have a higher capacity to earn profits (due to economies of scale) and can sustain increased property values in the area (Zukin et al., 2009). We therefore hypothesize that type 3 neighborhoods will show a decrease in the number of small stores and an increase in the number of supermarkets due to increased property value pressures.

Objective

Our **objectives** were: (1) to describe retail food environment changes over a 5-year period (2012 to 2016); and (2) to study the association between neighborhood social and economic change, from 2007 to 2011, and subsequent changes in the distribution of food stores in Madrid (Spain), from 2012 to 2016.

Methods

Study setting

Our study was conducted using data from the municipality of Madrid, Spain. The city is divided into 21 Census Districts, 128 Neighborhoods and, as the smallest census unit, around 2400 census sections (smallest census unit for which data is available, of around 1500 people). Table 2.1 of Chapter 2 describes the structure of these units. The Census Section has an average population of around 1500 people and is the smallest area for which census and other data is available. Some census sections may have populations as low as 700 or as high as 3500. Their boundaries are updated every year for election purposes and may result in a split or merging of census sections. Chapter 2 describes in detail how we dealt with changes in census section boundaries. In summary, we constructed a “common” set of census sections (n=2272) that were consistent during the entire study period. All analyses were conducted at the common census section level. Due to data availability, the analysis of retail food environment changes was restricted to the years 2012 through 2016.

Neighborhood Social and Economic Change

In a previous study (see Chapter 3), we defined and measured neighborhood social and economic change based on a theoretical framework drawn from Grigsby (Grigsby, 1987) and Van Ham (van Ham et al., 2012). This measurement model is based on a finite mixture modeling framework, which

generates types that share similar patterns over several indicators. Details on this model are available in Chapter 3.

Retail Food Environment Changes

To study the changes in the distribution of food stores, we created a tiered classification based on the economic activity of each retail space (see Figure 2.10 In Chapter 2). The first tier included all food stores, which consist of all unspecialized and specialized food stores. The second tier includes unspecialized stores (small grocery stores and supermarkets) and small specialized stores (fruit and vegetable, meat, seafood and bakeries). The third tier includes supermarkets (nested within unspecialized stores) and fruit and vegetable specialty stores (nested within specialty stores). Fruit and Vegetable Stores have lower scores on standard measures of healthy food availability (such as the HFAI (Bilal et al., 2016)) compared to supermarkets because they lack some healthy foods like whole grain breads or low-fat milk. However, they are a focus here because they are the primary source of fruits and vegetables and carry no unhealthy products (Bilal et al., 2016). While other specialty stores may carry less healthy products (meat or baked products), they usually lack processed and ultra-processed food (Bilal et al., 2016). We also studied changes in supermarkets to improve comparability with the existing retail food environment literature, where they are usually the food store of interest.

Data on the location of food stores was obtained from the *Censo de Locales* (commercial spaces census) from the Madrid City Government. This

census is collected for statistical purposes and to help with licensing and inspections. It includes, for every commercial space, its registered economic activities classified according to the National Classification of Economic Activities. Figure 2.10 in Chapter 2 shows the nested structure of this classification. To classify each store, we followed the algorithm shown in Figure 4.2. In summary, we classified unspecialized stores as supermarkets if their name was in a list of 60 supermarkets obtained from the yellow pages, and as small grocery stores if the name was not in the list. We classified stores with more than 1 specialized store code (e.g.: butcher [code 4722] and fruit store [code 4721]) as unspecialized stores, and applied the algorithm above. We classified stores as specialized stores if they had just one specialized store code (e.g.: Fruit and Vegetable Specialty stores were those with a single 4721 code). Table 4.1 lists the number of stores by year in all communities. This algorithm was trained using data from on-field audits in three census sections and validated in 42 census sections.

Data Analysis

The overall goal of this analysis is to study the association between neighborhood social and economic change and food environment changes over a 5-year period. The analysis had two parts: first, we categorized commercial spaces and explored retail food environment changes; second, we conducted an analysis studying the association of neighborhood social and economic change with food environment changes.

To categorize commercial spaces, we applied the algorithm described above. We then assigned each space to a census section by performing a spatial join with the layer of census sections. Once every space was assigned to a census section, we aggregated raw counts of spaces by food store category to the census section level. We used raw counts instead of densities per capita because changes in population density were already accounted for in the neighborhood change measurement model.

To study the association between neighborhood social and economic change and retail food environment changes we first needed to assign each census section a change type. We followed the method recommended by Bray et al. (Bray et al., 2015) that has been shown to reduce the amount of bias in latent type/finite mixture analysis with distal outcomes. To implement Bray's method we included the average number of stores in each category (total food stores, specialized stores, supermarkets and fruit and vegetable stores) from 2012-2016 as a predictor of type membership in the finite mixture model. Once we obtained a set of posterior probabilities of type membership for every census section-year of observation, we further reclassified each observation for the study according to their neighborhood change type in the previous five years. For this, we averaged the posterior probabilities of type membership for each census section from 2007 to 2011 and applied the following algorithm: if the average posterior probability in any type was above 0.8 (so that, on average, the census section was in that type for all years but one), the census section got assigned that type; if none of the

four types had an average posterior above 0.8 then the census section was assigned to a “transitional” fifth type. The number of census sections assigned to Type 2 was very low (8) and caused model convergence issues, so we dropped the eight Type 2 census sections from all analysis.

The second step in this association analysis was categorizing each census section by its changes in each food stores category from the previous year (i.e., gained at least one store, lost at least one store, stayed stable). We then used multinomial logistic regression with robust standard errors clustered at the census section level, where the dependent variable was either the odds of gaining or the odds of losing a store (reference outcome was stability; reference type given our hypothesis was type 3). We explored within-census section correlations in food store changes and found no correlation at any lag, so we opted not to use autoregressive errors. We ran a main model with adjustment for the baseline number of food stores and baseline socioeconomic status in the area. This variable was a composite index of four standardized indicators: % people with low education, % people with high education, current property value and current unemployment. The two education variables were weighted down 50%, to make education, wealth (property value) and unemployment weight equally. We also ran secondary analyses without these adjustment covariates.

All data management and statistical analyses were conducted in R version 3.3.0 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Retail Food Environment Changes

Table 4.1 shows a description of the number of food stores every year from 2012 to 2016 classified by type. The number of available spaces increased every year, from around 139,000 in 2012 to around 144,000 in 2016. The proportion of these that were open for business, according to the retail spaces census, significantly increased from 66% in 2012 to 68.3% in 2016 ($p=0.017$). Around 35% of the open spaces were classified as retail through the period, and the proportion of these that were in turn food stores significantly increased monotonically from 40.7% in 2012 to 46.9% in 2016 ($p=0.001$). Within food stores, there was an increase the proportion of unspecialized stores (21% in 2012 to 29% in 2016), with an increase in the share of supermarkets, and a decrease in specialized stores (42% in 2012 to 39% in 2016), with the number of fruit and vegetable stores remaining constant. The capacity of our algorithm to classify food stores remained constant over time, as only around 3.5% of all food stores remained unclassified after the application of the algorithm ($p=0.225$). The proportion of all open spaces without an economic activity code decreased over time, from 3% in 2012 to 1% in 2016.

Figure 4.3 shows the trends in the average number and proportion food stores by type by census section in each neighborhood change type. Overall, we see a similar increasing pattern in the number of all food stores, with areas in type 4 having a higher overall number of stores. The proportion of all retail stores

classified as food stores was higher in type 1 and 4 and increased in all areas over time. The number of supermarkets also followed an upward trend, but was lower overall in type 3 areas, where small specialized stores represented a higher proportion of all food stores. Last, the number of fruit and vegetable stores was stable or slightly trending upwards in all areas. Nonetheless, with respect to all food stores, the proportion of fruit and vegetable stores was trending downwards. Overall, the proportion of all food stores that were small specialized stores or fruit and vegetable stores was lower in Type 4 areas. Appendix 4.2 shows within area changes in the number of food stores (classified as losing, stable or gaining at least one store per type) for yearly changes. The proportion of areas losing a food store or not changing at all increased over time, while the proportion areas gaining one decreased from 29.6% in the 2012 to 2013 transition to 22% in the 2015 to 2016 transition. Supermarkets and specialized stores followed a similar trend, with a trend towards stability (no loss or gain).

Association of Neighborhood Change and Retail Food Environment

Changes

Table 4.2 shows the results of the main analysis of neighborhood change and food environment changes. After adjusting for the baseline number of stores and neighborhood socioeconomic status, Type 1 areas (more immigration and diversity), as compared to Type 3 areas (increase in property value), had decreased odds of losing supermarkets (OR=0.69, 95% CI 0.43 to 1.10) and a significant 39% increase in the odds of gaining supermarkets (OR=1.39, 95% CI

1.08 to 1.79). These areas also saw a significant 41% increase in the odds of losing small specialized stores (OR=1.41, 95% CI 1.04 to 1.93), and no change in the odds of gaining small specialized stores (OR=1.00, 95% CI 0.79 to 1.27). Overall, Type 1 areas saw an increase in the number of supermarkets and decrease in the number of small specialized stores.

Type 4 areas (population loss, aging), as compared to Type 3 areas (increase in property value) had a significant 80% increase in the odds of gaining a supermarket (OR=1.80, 95% CI 1.23 to 2.63). Although not significantly different from 1, there was a trend towards decreased stability in these areas, with an increase in the odds of either losing (OR=1.46, 95% CI 0.89 to 2.41) or gaining (OR=1.26, 95% CI 0.85 to 1.86) small specialized stores.

Areas with a large number of transitions between change types saw a significant 31% increase in the odds of gaining supermarkets (OR=1.31, 95% CI 1.04 to 1.65). No significant differences in the odds of gaining or losing FV stores was observed in any of the types of neighborhood change.

The role of Neighborhood Socioeconomic Status (NBSES) is shown in the second to last column of Table 4.2. For all food stores, supermarkets and small specialized stores, an increase in NBSES was associated with significantly increased odds of either losing or gaining stores. For FV stores, an increase in NBSES was only associated with a significant decrease in the odds of gaining these type of stores. An increase in the number of baseline stores in 2012 was associated with a higher odds of both gaining and losing stores during follow-up.

Appendix 4.3 shows the results of the secondary analysis not adjusting for baseline number of stores or NBSES. Appendix 4.4 shows the results of the secondary analysis only adjusting for baseline number of stores (and not NBSES). Both analysis show similar patterns as the main analysis with minor variations in the uncertainty around the estimates.

Discussion

In this study, we described the changes in the retail food environment of an entire city (Madrid) over 5 years (2012 to 2016). We found a dynamic environment, where at least one third of the census units (of around 1500 people) saw a change in the number of overall food stores. In general, we observed that the number of stores increased over time, with an increase in the number that were classified as supermarkets and a decrease in small specialized stores. These changes were especially evident in type 1 (gaining in poor migrants and diversity) and type 4 areas (losing population and aging), and were opposite in type 3 areas (increasing in property value.) Our main subject of study, fruit and vegetable stores, saw no changes overall, and only between 7 and 9% of census sections saw changes in their number, with no significant association with neighborhood change types.

This is the first study, to our knowledge, to describe detailed changes in the retail food environment in Spain. Previous research has looked at the cross-sectional picture, showing how food environments in Southern Europe differ widely from those in Anglo-Saxon countries. In particular, a study by Flavian showed that Spanish local food environments are dominated by small retailers, compared to Northern and Central European countries (Flavián et al., 2002). The number of outlets per resident is 3 times higher in Spain, Italy and Portugal compared to the UK, Finland, Denmark and Belgium (Flavián et al., 2002). Compared to Northern and Central Europe, market share of the top retailers is

reduced in Southern European Countries (Flavián et al., 2002) and so is the average number of supermarket or shopping malls per resident. In the case of Spain, this is related to two factors: (a) the availability of a transportation network that is especially dense in city centers and other dense areas (as opposed to suburbs where large food retailers may open) (Castillo-Manzano and López-Valpueda, 2009); and (b) the presence of small business owners lobbies that have guaranteed protective regulatory mechanisms related to the opening of large food retailers (Flavián et al., 2002). The presence of small food retailers in Madrid is abundant and appreciated by neighbors (Bilal et al., 2016). These food retailers (especially Fruit and Vegetable Stores) have a favorable ratio of healthy to unhealthy foods (Bilal et al., 2016; Díez et al., 2016).

In our study, we showed how areas gaining in diversity and migration (Type 1) had an increase in supermarkets and a decrease in specialized stores. This may be due to potential new markets opening with the incoming migrants. Our description of the baseline number of stores showed that these areas had a similar number of stores at baseline as compared to other areas, and our adjusted analysis by baseline number of stores was still showing the same inferences. Moreover, areas losing population and aging (type 4), along with areas without a clear change pattern (areas in transition), also saw significant increases in supermarkets. An alternative interpretation, considering that all other areas (other than the reference) saw an increase in the number of supermarkets is that our reference type (areas gaining in property value, reduced

unemployment, increase in education and in housing renovations – Type 3) actually lost supermarkets or gained less than other areas. Compared to this area, all other areas showed an increase (in some cases significant) in the odds of losing specialized stores, meaning that these areas were actually not losing these types of stores. In general, this means that areas with an increase in property value lost supermarkets and gained small specialized stores. While this runs contrary to our hypothesis that increases in property value would shift the food environment towards supermarkets, this is consistent with previous research in New York City (Zukin et al., 2009) showing that gentrifying areas may gain in “small boutique stores”. Our data did not allow us to differentiate between sub-types of small stores (“boutique” vs non-), so this remains a hypothesis to be confirmed in future research with more detailed data. We could not look at food affordability either, which would provide a marker for type of store. Actually, some previous research has described the phenomenon of “food mirages”, where accessibility and affordability are decoupled (Breyer and Voss-Andreae, 2013). If this is the case of type 3 areas, where availability is going up but affordability is decreasing, old residents with decreased purchasing power may be less able to obtain healthy foods in their own neighborhood.

Our previous research has showed that small stores (especially fruit and vegetable stores) have a high availability of certain healthy foods, while supermarkets display a high availability of both healthy and unhealthy foods (Bilal et al., 2016). This would mean that the overall healthiness of the food

environment is increasing in areas with increased property value and decreasing in all other areas. Given that these areas (as shown in Chapter 3) display the highest levels and increases in socioeconomic status, this has the potential to increase existing health disparities by neighborhood SES in Spain. Future research understanding the role that the dynamic nature of neighborhoods and food environment plays in health disparities has the potential to inform food-related policies in the future (Cobb et al., 2015b).

Limitations and Strengths

This study has some limitations. First, regarding the measurement of the exposure, the usual limitations of latent variable models apply. We checked for the potential violation of the main assumptions of the model (measurement invariance, conditional independence) and found them ignorable, but there is still a possibility for violations we could not detect. Importantly, we found that some areas transitioned quickly between neighborhood change types and did not have a clear change profile. We elected to classify these areas as “neighborhoods in transition”, but there is a potential for strong heterogeneity within them. Second, regarding the measurement of the outcome, we relied on a census of commercial spaces maintained by the local government for administrative purposes, which opens the possibility for measurement error. The number of unclassified stores was low (and decreased over time), but the potential for differential measurement error is of concern. Third, the change in the number of FV stores was low (<10% any given year), limiting our power to detect differences and potentially

contributing to the lack of significant results. Nonetheless the direction of the associations was consistent with that of small specialized stores.

Fourth, we lacked information on business hours of commercial spaces. As shown in previous research, time accessibility may also be an important determinant of population diets and is usually understudied (Widener et al., 2011a; Widener et al., 2011b; Widener and Shannon, 2014). This is of particular relevance in Madrid, where business hours were deregulated in 2012 (de Rada and González, 2015), leading to increased opening hours (and freedom to open on Sundays and Holidays).

This study has several strengths. First, we looked into neighborhood change in a decade of intense economic change that included a housing boom and subsequent recession, and into food environment changes for five of those years. Second, we looked into these changes for an entire city that included areas in all extremes of demographic, socioeconomic and urban characteristics. Third, our data sources are all universal in nature and are therefore free from sampling error. Fourth, we looked at within-area changes, which allows us to control for time-fixed confounding that may create spurious associations.

Conclusion

Our study is a first step towards shedding light on the dynamic composition of the food environment. Future research (see Chapter 5) should determine whether neighborhood change is a putative exposure on health, and whether the food environment plays a role in this association

Tables

Table 4.1: Number of food stores by type for the entire city of Madrid.

Space Type	7/2012	7/2013	7/2014	7/2015	7/2016	Abs. p-value*	% p-value ⁺
All Spaces	139262	140840	141774	143810	144811	0.001	N/A
-Open Spaces	91887 (66.0%)	93589 (66.5%)	94206 (66.4%)	96667 (67.2%)	98864 (68.3%)	0.003	0.017
--All Retail	32426 (35.3%)	33119 (35.4%)	33855 (35.9%)	34440 (35.6%)	34837 (35.2%)	0.001	0.905
--- All Food Stores	13182 (40.7%)	14163 (42.8%)	15037 (44.4%)	15813 (45.9%)	16325 (46.9%)	0.001	0.001
----General	2838 (21.5%)	3574 (25.2%)	3980 (26.5%)	4438 (28.1%)	4706 (28.8%)	0.002	0.010
----- Supermarkets	808 (28.5%)	1165 (32.6%)	1427 (35.9%)	1689 (38.1%)	1818 (38.6%)	0.002	0.008
----- Specialized	5576 (42.3%)	5773 (40.8%)	6059 (40.3%)	6198 (39.2%)	6330 (38.8%)	0.001	0.004
----- FV Stores	1702 (30.5%)	1731 (30.0%)	1829 (30.2%)	1870 (30.2%)	1920 (30.3%)	0.002	0.803
----Other	4292 (32.6%)	4322 (30.5%)	4485 (29.8%)	4639 (29.3%)	4721 (28.9%)	0.003	0.020
----Unclassified (Food)	476 (3.6%)	494 (3.5%)	513 (3.4%)	538 (3.4%)	568 (3.5%)	0.001	0.225
---Non-Food Retail	19244 (59.3%)	19008 (57.4%)	18900 (55.8%)	18803 (54.6%)	18713 (53.7%)	0.005	0.001
--Unclassified (Open)	4575 (3.3%)	4076 (2.9%)	2673 (1.9%)	1770 (1.2%)	1761 (1.2%)	0.009	0.008

Proportions are relative to the upper tier category (e.g.: FV stores proportions are over the total specialized stores count)

*: p-value for linear trend in absolute value from 2012 to 2016

+: p-value for linear trend in proportion (relative to upper tier category) from 2012 to 2016

Table 4.2: Neighborhood Change (averaged from 2007 to 2011) and gain/loss in food stores (2012 to 2016) adjusted for baseline number of stores and socioeconomic status at baseline (2006)

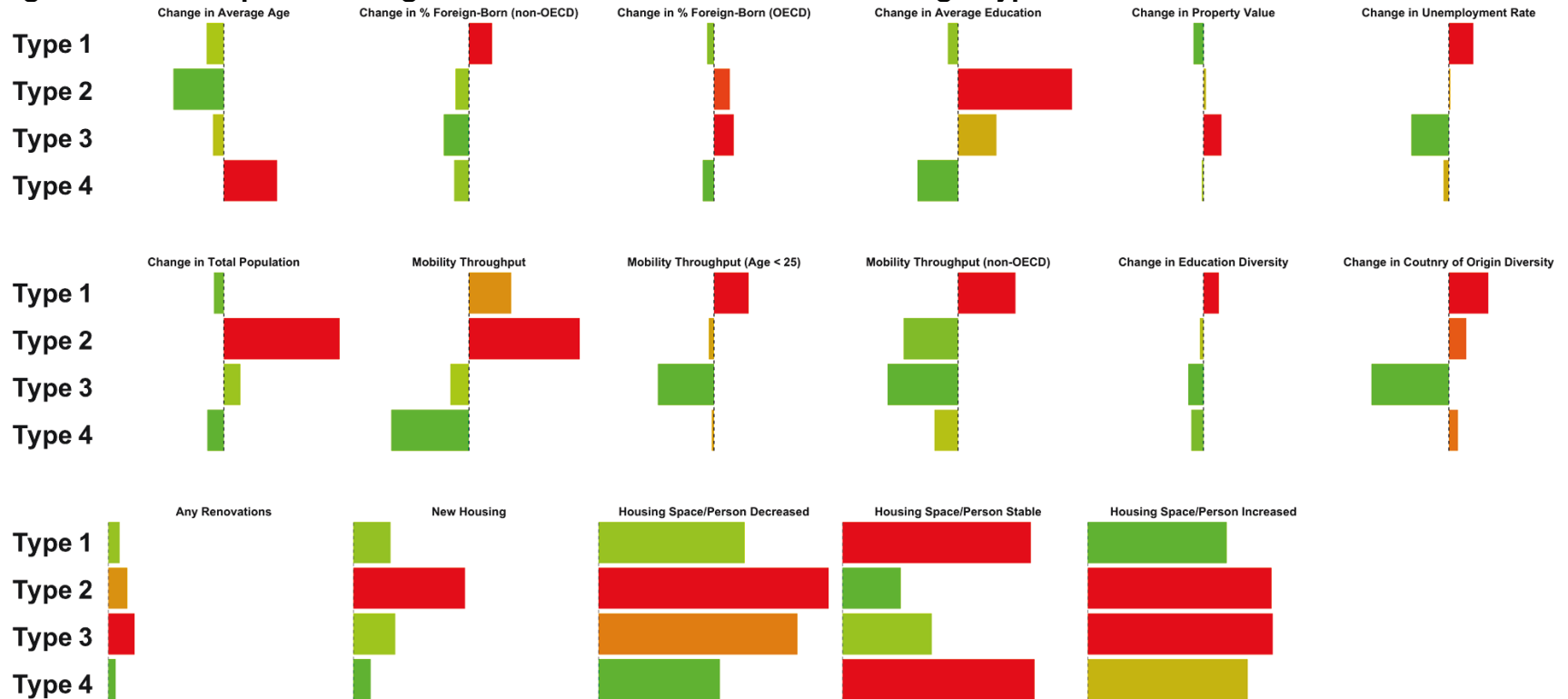
Store Loss	Type 1 (+Diversity)	Type 3 (+Prop. Value)	Type 4 (+Aging)	Areas in Transition	NBSES (+1 SD)	Baseline # of Stores (+1)
All Food Stores	1.02 (0.75;1.40)	1 (Ref.)	1.25 (0.77;2.02)	0.93 (0.71;1.23)	1.50 (1.35;1.67)	1.05 (1.03;1.06)
Supermarkets	0.69 (0.43;1.10)	1 (Ref.)	1.19 (0.57;2.47)	1.01 (0.69;1.50)	1.10 (0.94;1.29)	3.07 (2.70;3.50)
Specialized Stores	1.41 (1.04;1.93)	1 (Ref.)	1.46 (0.89;2.41)	1.19 (0.90;1.59)	1.29 (1.16;1.44)	1.07 (1.05;1.09)
FV Stores	1.23 (0.79;1.92)	1 (Ref.)	1.51 (0.75;3.06)	1.12 (0.75;1.69)	0.95 (0.81;1.11)	1.29 (1.20;1.40)
Store Gain	Type 1 (+Diversity)	Type 3 (+Prop. Value)	Type 4 (+Aging)	Areas in Transition	NBSES (+1 SD)	Baseline # of Stores (+1)
All Food Stores	1.08 (0.89;1.30)	1 (Ref.)	1.13 (0.84;1.51)	1.03 (0.87;1.21)	1.42 (1.33;1.52)	1.07 (1.06;1.09)
Supermarkets	1.39 (1.08;1.79)	1 (Ref.)	1.80 (1.23;2.63)	1.31 (1.04;1.65)	1.13 (1.04;1.22)	1.63 (1.49;1.78)
Specialized Stores	1.00 (0.79;1.27)	1 (Ref.)	1.26 (0.85;1.86)	0.91 (0.73;1.13)	1.16 (1.07;1.26)	1.08 (1.06;1.09)
FV Stores	0.97 (0.70;1.34)	1 (Ref.)	1.03 (0.59;1.78)	0.82 (0.61;1.12)	0.86 (0.76;0.96)	1.25 (1.18;1.33)

Footnote:

Model adjusted for the number of stores at baseline (2012), and neighborhood socioeconomic status (in 2006)

Figures

Figure 4.1: Description of Neighborhood Social and Economic Change Types



Footnote: graphical representation of Table 3.3. Darker bars are more positive values, while lighter bars are more negative values. In the case of categorical indicators (bottom row), all values are positive, so lighter values represent lower probabilities while darker values represent higher probabilities.

Figure 4.2. Algorithm to classify food stores based on declared CNAE activities and store name.

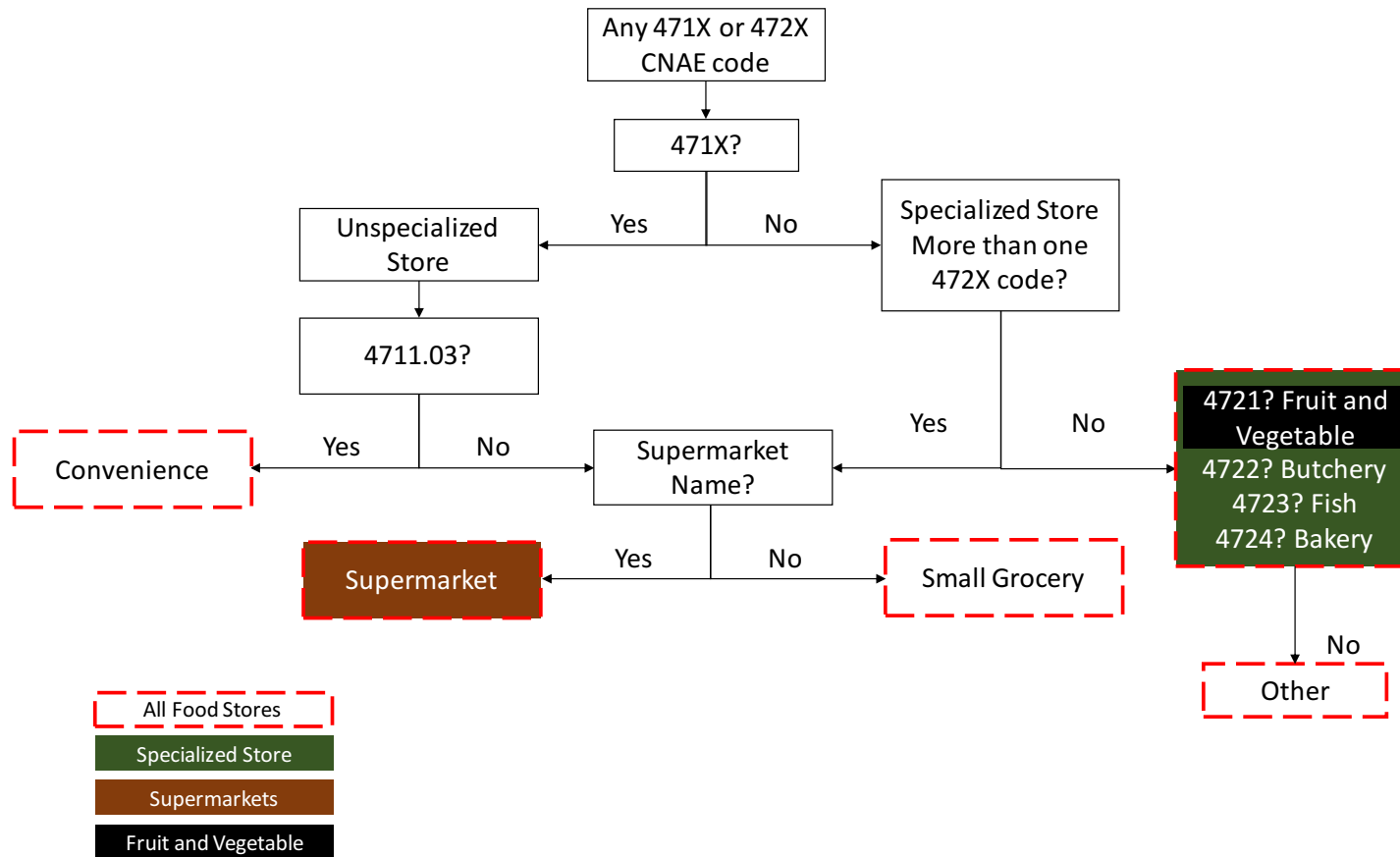
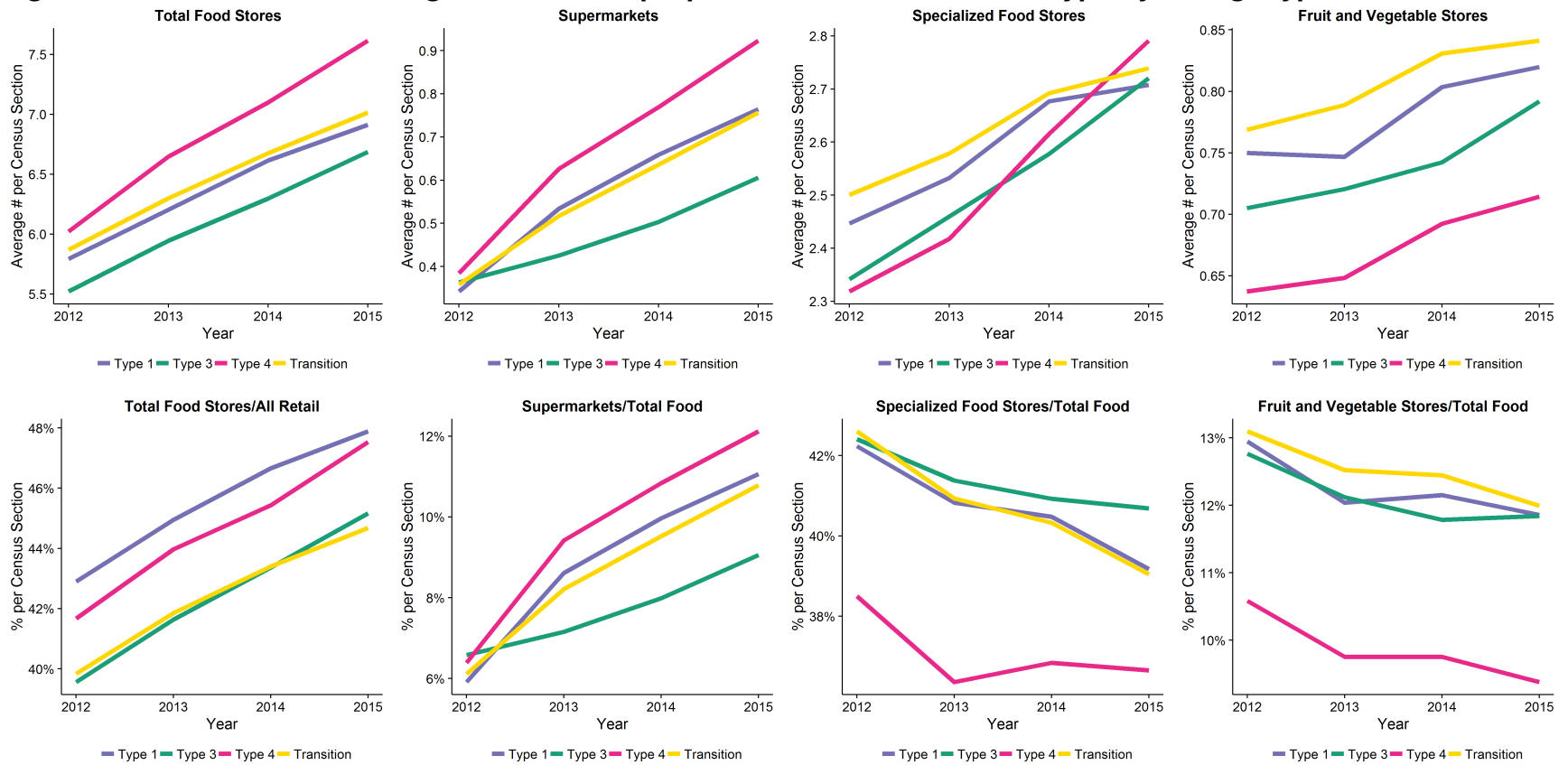


Figure 4.3: Trends in the Average number and proportion of each food store type by change type



References

- Bilal, U., Diez, J., Alfayate, S., et al. (2016). Population cardiovascular health and urban environments: the Heart Healthy Hoods exploratory study in Madrid, Spain. *BMC medical research methodology*. 16, 104.
- Bray, B.C., Lanza, S.T., Tan, X. (2015). Eliminating Bias in Classify-Analyze Approaches for Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*. 22 (1), 1-11.
- Breyer, B., Voss-Andreae, A. (2013). Food mirages: geographic and economic barriers to healthful food access in Portland, Oregon. *Health & place*. 24, 131-139.
- Castillo-Manzano, J.I., López-Valpuesta, L. (2009). Urban retail fabric and the metro: A complex relationship. Lessons from middle-sized Spanish cities. *Cities*. 26 (3), 141-147.
- Cobb, L.K., Anderson, C.A., Appel, L., et al. (2015a). Baltimore City Stores Increased The Availability Of Healthy Food After WIC Policy Change. *Health Affairs*. 34 (11), 1849-1857.
- Cobb, L.K., Appel, L.J., Franco, M., Jones-Smith, J.C., Nur, A., Anderson, C.A. (2015b). The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results. *Obesity*. 23 (7), 1331-1344.
- de Rada, V.D., González, R.L. (2015). La liberalización de los horarios comerciales. Un análisis de la opinión de la sociedad española. *Inguruak* (59),
- Díez, J., Bilal, U., Cebrecos, A., 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.
- Dunkley, B., Helling, A., Sawicki, D.S. (2004). Accessibility Versus Scale. *Journal of Planning Education and Research*. 23 (4), 387-401.
- Flavián, C., Haberberg, A., Polo, Y. (2002). Food retailing strategies in the European Union. A comparative analysis in the UK and Spain. *Journal of Retailing and Consumer Services*. 9 (3), 125-138.
- Franco, M., Bilal, U., Díez, J. (2016). Food Environment. In B. Caballero, P.M. Finglas, & F. Toldrá (Eds.), *Encyclopedia of Food and Health* pp. 22-26). Oxford: Academic Press.
- Franco, M., Diez Roux, A.V., Glass, T.A., Caballero, B., Brancati, F.L. (2008). Neighborhood characteristics and availability of healthy foods in Baltimore. *American journal of preventive medicine*. 35 (6), 561-567.
- Glanz, K., Sallis, J.F., Saelens, B.E., Frank, L.D. (2005). Healthy nutrition environments: concepts and measures. *Am J Health Promot*. 19,
- Glanz, K., Sallis, J.F., Saelens, B.E., Frank, L.D. (2007). Nutrition Environment Measures Survey in stores (NEMS-S): development and evaluation. *American journal of preventive medicine*. 32,

- Glass, T.A., Bilal, U. (2016). Are neighborhoods causal? Complications arising from the 'stickiness' of ZNA. *Soc Sci Med*,
- Grigsby, W.G. (1987). *The Dynamics of Neighborhood Change and Decline: Pergamon Journals*.
- Jokela, M. (2014). Are neighborhood health associations causal? A 10-year prospective cohort study with repeated measurements. *Am J Epidemiol*. 180 (8), 776-784.
- Jokela, M. (2015). Does neighbourhood deprivation cause poor health? Within-individual analysis of movers in a prospective cohort study. *J Epidemiol Community Health*. 69 (9), 899-904.
- Ludwig, J., Sanbonmatsu, L., Gennetian, L., et al. (2011). Neighborhoods, obesity, and diabetes--a randomized social experiment. *The New England journal of medicine*. 365 (16), 1509-1519.
- Moore, L.V., Diez Roux, A.V. (2006). Associations of neighborhood characteristics with the location and type of food stores. *American journal of public health*. 92,
- Morland, K., Wing, S., Diez Roux, A., Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American journal of preventive medicine*. 22,
- Richardson, A.S., Meyer, K.A., Howard, A.G., et al. (2014). Neighborhood socioeconomic status and food environment: A 20-year longitudinal latent class analysis among CARDIA participants. *Health & Place*. 30, 145-153.
- Rose, G. (1985). Sick individuals and sick populations. *International journal of epidemiology*. 14 (1), 32-38.
- Rummo, P.E., Guilkey, D.K., Shikany, J.M., Reis, J.P., Gordon-Larsen, P. (2016a). How do individual-level sociodemographics and neighbourhood-level characteristics influence residential location behaviour in the context of the food and built environment? Findings from 25 years of follow-up in the CARDIA Study. *Journal of Epidemiology and Community Health*,
- Rummo, P.E., Hirsch, J.A., Howard, A.G., Gordon-Larsen, P. (2016b). In Which Neighborhoods Are Older Adult Populations Expanding? Sociodemographic and Built Environment Characteristics Across Neighborhood Trajectory Classes of Older Adult Populations in Four U.S. Cities Over 30 Years. *Gerontology and Geriatric Medicine*. 2,
- Story, M., Kaphingst, K.M., Robinson -O'Brien, R., Glanz, K. (2008). Creating healthy food and eating environments: policy and environmental approaches. *Annual Reviews of Public Health*. 29,
- van Ham, M., Manley, D., Bailey, N., Simpson, L., Maclennan, D. (2012). *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research: Springer Netherlands*.
- White, J.S., Hamad, R., Li, X., et al. (2016). Long-term effects of neighbourhood deprivation on diabetes risk: quasi-experimental evidence from a refugee dispersal policy in Sweden. *The Lancet Diabetes & Endocrinology*. 4 (6), 517-524.

- Widener, M.J., Metcalf, S.S., Bar-Yam, Y. (2011a). Dynamic urban food environments a temporal analysis of access to healthy foods. *American journal of preventive medicine*. 41 (4), 439-441.
- Widener, M.J., Metcalf, S.S., Bar-Yam, Y. (2011b). Dynamic Urban Food Environments: A Temporal Analysis of Access to Healthy Foods. *American journal of preventive medicine*. 41 (4), 439-441.
- Widener, M.J., Shannon, J. (2014). When are food deserts? Integrating time into research on food accessibility. *Health & place*. 30, 1-3.
- Zukin, S., Trujillo, V., Frase, P., Jackson, D., Recuber, T., Walker, A. (2009). New Retail Capital and Neighborhood Change: Boutiques and Gentrification in New York City. *City & Community*. 8 (1), 47-64.

**CHAPTER 5: NEIGHBORHOOD SOCIAL AND ECONOMIC CHANGE AND
DIABETES INDICENCE OVER 6 YEARS IN MADRID, SPAIN**

Abstract

Background: We studied the association between neighborhood social and economic change and diabetes incidence in Madrid (Spain) from 2009 to 2014.

Methods: We designed a prospective cohort study from electronic health records of the entire population of 4 districts of Madrid (n=801,663). We included 200,670 individuals aged 40 or above and free of diabetes by January 1st 2009 and followed them for up to 6 years to ascertain diabetes incidence. We measured neighborhood social and economic change using a finite mixture model with 15 indicators that estimated membership in four types of change. We categorized areas by neighborhood change type from 2006 to 2009 and applied this to diabetes incidence measured from 2009-2014. We used Cox Proportional Hazards models to estimate the association between neighborhood change and diabetes adjusted by age, sex and baseline area socioeconomic status (SES).

Results: After adjusting for age, sex, and baseline SES, there was a significant association between neighborhood change and diabetes incidence. Compared to those living in neighborhoods characterized by increased SES and reduced diversity, people living in areas with decreasing SES, increased diversity, and increased non-OECD migrants, and areas with low residential mobility, aging, and no new housing had a 10% and 17% reduction in the hazard of diabetes (p<0.001).

Discussion: We found that neighborhood social and economic change was significantly associated with diabetes incidence. This evidence can help guide policies that create healthier environments for diabetes prevention.

Introduction

Diabetes Prevention Through the Study of Mass Influences

The burden of diabetes has seen a large increase in Western countries in recent decades (NCD-RisC, 2016). Diabetes-attributable costs in the European Union have been estimated to be over \$100 billion per year and are predicted to continue increasing in the following decades (Zhang et al., 2010). Population preventive strategies are needed to decrease this burden (Rose, 1985), taking into consideration mass influences that differ across populations (Rose, 1985). Among these mass influences are neighborhood characteristics and their dynamics.

Neighborhood Characteristics and Diabetes Burden

The association between current neighborhood socioeconomic status and several measures of diabetes (prevalence, incidence or control) is robust and has been replicated in the US (Geraghty et al.; Piccolo et al., 2015), other Anglo-Saxon countries (Booth et al., 2013; Connolly et al., 2000; Cox et al., 2007; Hippisley-Cox et al., 2004), Central (Müller et al., 2013) and Northern Europe (Mezuk et al., 2013). While these influences have received scant attention in Southern Europe, there is evidence of the presence of this association (Larrañaga et al., 2005). However, the policy implications of these types of analysis are hindered by insufficient attention paid to the dynamics of residential environments (van Ham et al., 2012). In particular, a lack of a longitudinal

approach to neighborhood effects hinders our ability to recommend policies for population-wide diabetes prevention.

Neighborhood Relocation and Diabetes

The most compelling evidence of the effects of neighborhood characteristics on diabetes comes from the Moving to Opportunity Study (MTO) that randomized people to receive vouchers for residential mobility and observed a reduction in extreme obesity and HbA1c (Ludwig et al., 2011) in people moving to neighborhoods with lower rates of poverty. A study from Sweden analyzed residential relocation as a natural experiment and found a decreased rate of diabetes in refugees relocated to wealthier areas (Unwin and Hambleton, 2016). Nonetheless, both studies focus on change through residential mobility, instead of change in areas themselves through policy or economic changes.

Neighborhood Change and Diabetes

The literature is, however, scarce on studies that examine the effect of longitudinal changes in the social and economic environment. A few studies have looked at changes in specific neighborhood built and social environments and intermediate behaviors related to diabetes, such as physical activity or weight gain (Zenk et al., 2016; Zhang et al., 2016). Zenk et al. (2016) found an increase in the density of small grocery stores was associated with decreased BMI, but no association was found for other types of stores. Zhang et al. (2016) found that the opening of new supermarkets (which often have unhealthy and processed foods) was not associated with a change in BMI.

Potential Mechanisms for the effect of Neighborhood Change on Diabetes

Diabetes is a sensitive marker of contextual economic conditions. The patterns of diabetes incidence in Cuba have followed economic downturns and recoveries over short time frames (Franco et al., 2013). While the changes in Cuba were strong in magnitude, there is also an evidence for a negative effect of short-term economic growth in Western economies (Catalano et al., 2011). In particular, mortality increases when the economy grows in the short term, an effect that is independent of individual-level changes in economic conditions (Tapia Granados et al., 2014). The association between economic growth and cardiovascular mortality is three times stronger in urban environments as compared to rural areas (Sameem and Sylwester, 2017). A potential mechanism for this association is that rapid changes in urban areas are associated with increase in cardiovascular risk factors, including diabetes. In a study looking into type 1 and type 2 diabetes in Chicago (Grigsby-Toussaint et al., 2010) found that “non-type 1 Diabetes” risk was increased in “Emerging High Income” neighborhoods. No study, to our knowledge, has studied changes in local social or economic conditions and diabetes risk in Europe, where the patterns of segregation and neighborhood selection and change differ widely from the US (Kazepov, 2005; Tammaru et al., 2015).

Measuring Neighborhood Change

To improve on the measurement of neighborhood change we constructed a finite mixture model that estimated four types of Neighborhood Social and

Economic Change (see Chapter 3 for more details). These included: (Type 1) a type of neighborhoods with an increased proportion of migrants from poor countries, decreased education level and property values along with increased unemployment, a moderate degree of residential mobility and an increase in diversity; (Type 2) a type of neighborhoods with decreased average age and increased education level, the highest degree of residential mobility along with new housing construction; (Type 3) a type of neighborhoods with increased education level and property value, decreased unemployment, a moderate degree of residential mobility and elevated housing renovations; and (Type 4) a type of neighborhoods with increased average age, the lowest degree of residential mobility with population loss and no new housing or renovations. Figure 5.1 shows a description of the four types of neighborhood social and economic change.

Objective

Taking the above into consideration, we studied the association between neighborhood social and economic change and diabetes incidence. As shown above, the current literature suggests a potential negative effect of rapid social and economic change on diabetes incidence shortly thereafter. We hypothesized that type 2 and type 3 areas (areas with higher degrees of residential mobility or increases in property value) would have a higher incidence of diabetes.

Methods

Study setting

We conducted this study as a part of the HeartHealthyHoods Retrospective Study. This study collected retrospective data on all health centers of 4 districts of Madrid (see Chapter 2, Table 2.1 for a description of the geographical areas). These centers all belonged to the same Health Area (when Health Areas were used in the administration of primary health care in Madrid) and were among the first to incorporate and standardize electronic health records. These four districts had around 600,000 residents in total (18% of the total population of Madrid) and are representative of the rest of the city of Madrid (Appendix 5.1). The study area is divided into 393 census sections (smallest census unit for which data is available, of around 1500 people). All analyses were conducted at the census section level. Data from January 1st 2005 to December 31st 2015 was used to classify census sections into 4 types using a measurement model of neighborhood social and economic change (see chapter 3). Data on diabetes incidence covers the period from January 1st 2009 to December 31st 2014. These years represent a wide variety of economic and urban conditions given large economic changes and urban transformations that happened in Spain during this period.

Neighborhood Social and Economic Change

In a previous study (see Chapter 3), we defined and measured neighborhood social and economic change based on a theoretical framework

drawn from Grigsby (1987) and Van Ham (2012). This measurement model is based on a finite mixture modeling framework, which generates types that share similar patterns over several indicators.

Electronic Health Records

Diabetes data was obtained from electronic health records of the entire population receiving care in the primary care health centers of the 4th Health Area of Madrid from the beginning of 2009 to the end of 2014 (6 years of potential follow-up time). In total, we had data for 3.8 million person-years of observation, including date of birth, sex, medical diagnoses and laboratory values (for 2013 and 2014). Due to the system in place to screen for cardiovascular risk factors in people aged 40 or above (Bilal et al., 2016b), we restricted our dataset to people born after January 1st 1969 (aged 40 or above by baseline). We also had information on the geocoded residential location of each individual by 2013 and 2014. We assigned the location in 2013 as the residence for the entire study period. If an individual died or moved out of the area before 2013, the geocoded residential location was missing and we excluded these individuals (19%). In total, we used data from 199,621 people aged 40 or above and free of diabetes by baseline with available geocoded residential location by 2013.

Diabetes Data

A diagnosis of Type-2 Diabetes was defined using the T90 diagnosis code of the ICPC-2 (“Diabetes non-insulin dependent”). A previous study has validated the diagnosis of diabetes in this dataset with a kappa of 0.99, with high sensitivity

(99.5%) and specificity (99.5%) (de Burgos-Lunar et al., 2011). We defined incident diabetes as a new diagnosis of type-2 diabetes in someone not otherwise classified as a prevalent diabetic by January 1st 2009. We recorded the date of diagnosis for every new case of diabetes (code T90).

Statistical Analysis

The overall goal of this analysis is to study the association between type of neighborhood social and economic change and diabetes risk over a period of six years. Below we describe: (1) the creation of the dataset and operationalization of variables; (2) exploratory and descriptive analysis of the data; (3) analysis of incidence; and (4) sensitivity analyses.

We first built a multilevel dataset in which each observation was an individual-year of follow-up with data on the individual's age (by January 1st 2009) and sex, diabetes diagnosis date (if any), along with its census section of residence. We operationalized neighborhood change types by averaging the posterior probabilities of neighborhood change type membership from 2006 to 2009 (the four years before the beginning of the follow-up period). We then classified each area as follows: if the averaged posterior probability of type membership was above 0.75 (that is, the neighborhood belonged to that type at least in all years except for one) then the area was assigned such type; if none of the 4 types had an averaged posterior of 0.75, then we assigned the area to a fifth type ("areas in transition"). Appendix 5.2 shows the make-up of each averaged change type, as compared to the original yearly types from 2006 to

2009. No area was classified as type 2 after averaging the change types from 2006 to 2009.

To conduct a basic description of the study sample data we explored sample characteristics at baseline (January 1st 2009) by type of neighborhood social and economic change at baseline (2006-2009), including age, sex, diabetes incidence, and prevalence of its complications (retinopathy and chronic kidney disease diagnoses) and other cardiovascular conditions (hypertension, dyslipidemia, cardiovascular disease).

To examine the association of neighborhood change type (2006-2009) with diabetes incidence (2009-2014) we excluded all individuals with prevalent diabetes by January 1st 2009. In each subsequent year of follow-up, each individual entered the sample on January 1st and exited on the diabetes diagnosis date (outcome), date of death or moving out of the area (censoring), or December 31st 2014 (administrative censoring). To explore the incidence of diabetes we computed both incidence rates and Kaplan-Meier estimates of diabetes-free survival by type of neighborhood social and economic change. To explore the association between neighborhood social and economic change and diabetes incidence we used a Cox Proportional Hazards model with Sandwich Robust Standard errors clustered on the census section. An unadjusted model was first estimated with dummy variables for type membership, followed by a model adjusted for age (in 5 categories) and sex, and a model further adjusted for Neighborhood Socioeconomic Status (NBSES). NBSES was a composite

index of four standardized indicators: % people with low education, % people with high education, current property value and current unemployment. The two education variables were weighted down 50%, to make education, wealth (property value) and unemployment weight equally.

We performed two sensitivity checks to assess the robustness of our inferences. First, we assessed the sensitivity of our assignment of individuals to census sections in 2009 based on 2013 data, by producing a more conservative estimate, where individuals that switched health centers at any point were excluded from the analysis (assuming they had switched residential locations when switching health centers). Second, we assessed whether our results are influenced by areas with a low count of diabetes cases by including only areas with 10 or more incident cases of diabetes over the 6 years of follow-up.

All analyses were conducted in R v3.3.0. Mplus v7.4 was used for the estimation of the finite mixture measurement model.

Results

Study Population

Table 5.1 shows a description of the study sample by type of neighborhood social and economic change at baseline (2006 to 2009). The mean age across types is similar ($p=0.492$), although the distribution in categories varies, with Type 1 and Type 4 areas having a higher proportion of younger individuals ($p=0.018$), and areas in transition having a higher proportion of older

individuals. The proportion of men is higher in type 1 and 4 (44.3 and 44.8% in each) as compared to type 3 and areas in transition (42.7 and 43.5%, $p=0.006$). The 6-year cumulative incidence of diabetes was 3.8% overall, and was similar across areas ($p=0.503$). The prevalence of other cardiovascular risk factors is similar across areas. In particular, the prevalence of hypertension, dyslipidemia, CVD, CKD and retinopathy was 24.5%, 19.6%, 4.3%, 1.2% and 0.3, respectively. The SES index distribution varied by neighborhood change type, with an increasing SES gradient going from Type 1 and Areas in Transition (lowest SES), Type 4, to Type 3 (highest SES) ($p<0.001$).

Diabetes Incidence

Figure 5.2 shows the unadjusted Kaplan-Meier curves of diabetes incidence by type of neighborhood social and economic change. There is a significant ($p<0.001$ for the log-rank test of curve equality) difference in the survival curves, as Type 1 areas have the highest incidence, followed by neighborhoods in transition, Type 3 and Type 4 areas (lowest incidence).

Association of Neighborhood Social and Economic Change with Diabetes Incidence

Table 5.2 shows the main results of this study. In Model 1 (unadjusted), there is no significant difference between types in diabetes incidence. The second column shows a model adjusted by age and sex. In this model, Type 1 areas have a significant increase in the hazard of diabetes (HR=1.13, 95% CI 1.02 to 1.25) as compared to Type 3 areas. The role of sex (higher incidence in

males, HR=1.76, 95% CI 1.68 to 1.85) and age is evident (higher incidence in older people, with hazard ratios ranging from 2.42 in ages 50-60 to 4.46 in ages 70-80, as compared to ages 40-50.).

Model 3 shows the results adjusted by age, sex and area-level socioeconomic status. Specifically, there is a statistically significant 10% decrease (HR=0.90, 95% CI 0.82 to 0.99) in the hazard of diabetes in people living in Type 1 areas and a 17% decrease in people living in Type 4 areas (HR=0.83, 95% CI 0.73 to 0.95), as compared to Type 3 areas. Moreover, there is also a 12% decrease in the hazard of diabetes in people living in Areas in Transition (HR=0.88, 95% CI 0.80 to 0.96) as compared to Type 3 areas. In all models, neighborhood socioeconomic status is significantly and negatively associated with diabetes incidence, with a 26% reduction in the hazard of diabetes per 1 SD increase in the index of neighborhood SES (HR=0.74, 95% CI 0.71 to 0.77) in Model 3.

Sensitivity analyses

Appendix 5.4 and 5.5 shows data on the sensitivity analysis excluding individuals that had moved between health centers and excluding areas with low counts (<10) of diabetes cases. In both cases the same patterns of association remained, although confidence intervals widened in the case of the sensitivity analysis excluding people that changed health centers.

Discussion

These analyses have shown that areas with increased diversity and foreign-born migrants (Type 1), areas with an aging population and low residential mobility (Type 4), and areas with a shifting change profile (Areas in Transition) have a decreased incidence of diabetes. Alternatively, areas characterized by increases in property values, decreases in diversity and unemployment and an increase in housing renovations (type 3) showed a significant increase in the hazard of diabetes. These results were robust to our sensitivity checks (health center changes and low event counts).

Latent variable models, as the one used in this study, allow epidemiologic studies to harness the relationships between exposures of interest. For example, some of the types in our study show similar patterns of change in the age composition (type 1 and 3), proportion of people from non-OECD countries (types 3 and 4), education (types 1 and 4) and unemployment (types 3 and 4). However, these areas differed from each other in how these changes clustered around some other indicators, such as residential mobility (higher in type 1, lower in type 3 and 4), origin diversity (lowest in type 3), housing (lowest in type 4), and renovations (highest in type 3). The inferences from this study can have two implications: (a) as a direct interpretation of an association between each type and diabetes; and (b) as a hypothesis generating process where combinations of variables can be tested in further studies (e.g., housing renovations and

increases in SES as a dominant characteristic of type 3 that may highlight gentrifying areas).

These results could emerge due to several mechanisms. First, Type 3 areas where property value is increasing (or not decreasing as much as in other areas) may see increased economic pressures in current residents, a marker of displacement before gentrification. Decreased housing affordability has been linked to poorer health outcomes (Pollack et al., 2010). A study in Philadelphia found that minority residents in gentrifying areas (who are in most cases the displaced population) have declines in self-reported health status, an effect absent in non-minority residents (Gibbons and Barton, 2016). This effect has also been found for minorities, for whom there is increased diabetes incidence in areas with increasing SES (Grigsby-Toussaint et al., 2010). However, with the data at our disposal we cannot determine whether our results are due to previous residents of the area remaining, people that are leaving or will leave the area, or new residents. Future research in our study with more detailed individual data will assess this hypothesis

Type 1 areas showed decreased diabetes incidence. These areas had an increase in migrants, declining SES (decrease in education and property value, increase in unemployment), and increased diversity. A potential benefit of decreasing SES seems unlikely. The increased proportion of foreign-born (non-OECD) migrants may drive diabetes rates down if the healthy migrant paradox for diabetes (Afable-Munsuz et al., 2013) is present in Spain, an hypothesis not

yet assessed. However, caution must be exercised to not fall into an atomistic fallacy (Diez Roux, 2002), as the individual effect of the process of migration may differ from a increase in the proportion in migrants. Last, a key difference between Types 1 and 3 is the increased diversity of Type 1 areas (especially in terms of foreign-born people) and the decreased diversity of Type 3 areas (especially in terms of education). Previous research has shown some positive effects of diversity on physical activity (Denton et al., 2014) and a negative effect of segregation (reduced diversity) on cardiovascular disease (Kershaw et al., 2015) and hypertension (Kershaw et al., 2011).

Type 4 areas also showed a decreased diabetes incidence. These areas were characterized by an increasing proportion of Spaniards (as compared to foreign-born people), aging of the population, population loss, lower levels of mobility and a decrease education levels. These decreases in education may be linked to an aging of the population. Similar findings have been described for smoking (Bilal et al., 2016a). In particular, either older people are moving in (less plausible) or younger people are moving out. If the latter is true, this may mean that those who stay in the area have a decreased risk of diabetes. Most of the neighborhood effects literature has been focused on movers (a minority of the population (Glass and Bilal, 2016). More research should focused on stayers, a segment of the population for which studies like the MTO (Moving to Opportunity) cannot make inferences (Sampson, 2008).

Our study has several strengths. First, we study the entire population registered in a universal and integrated health system in an area of a very large city (Madrid), where we included 200,000 people above 40 free of diabetes at baseline. Many studies looking at the contextual determinants of diabetes use data from research-driven cohort studies. While these studies have the advantage of standardized and high-quality data collection, they may suffer from a number of biases derived from a non-random sampling of the study participants (Chaix et al., 2011; Weisskopf et al., 2015). In particular, the role that context plays in determining selection into a study may be particularly relevant in studies on the effect of context on health (Weisskopf et al., 2015). The use of EHR from integrated universal health systems that share a common EHR for the entire population may be advantageous, providing higher rates of standardization in data collection, less selection bias and more complete coverage. In a pilot study conducted in Madrid using EHR (in a subset of 12 census sections of the study area in this manuscript), we found that 97.5% of the population listed in the census was registered in a health care center and could be geocoded to their residential address (Bilal et al., 2016b). A second strength is that our measure of diabetes prevalence has been validated with a kappa of 0.99 (de Burgos-Lunar et al., 2011). Third, our measurement model was constructed using publicly available indicators that increase the replicability of our findings and the applicability to other health outcomes.

We acknowledge several important limitations. The validity of our measures of diabetes prevalence and control is high, but we cannot achieve the levels of standardization and validity that cohort studies do. Second, the available data for individual level confounders was restricted to basic sociodemographics (age and sex), which opens the possibility for residual confounding in our inferences. In particular, we lacked the ability for adjust for the potential confounding influences of individual level socioeconomic status (SES). Given the strong patterns of segregation by SES (Tammaru et al., 2015) it is highly likely that people living in lower SES areas have a lower SES themselves. For now, our sensitivity analysis excluding people that had changed health centers at any point (a proxy for residential mobility) shows analogous inferences to our main analysis.

Third, we were not able to use our measurement model to its full extent. In particular, we had too few areas of Type 2 in our study area and no area was actually classified as Type 2 after averaging the change types from 2006 to 2009. This area, therefore, could not considered for this study, so we could not assess whether neighborhoods with a high intensity of residential mobility (type 2) had an association with diabetes. This was one of our main hypothesis (that these neighborhoods would have a higher incidence of diabetes), and remains untested until future studies can be conducted. We also were not able to completely characterize the areas classified as “in transition”, as they had an unstable change profile that disavowed characterization. However, as seen in

Appendix 5.2, these areas had an over-representation of Type 4 areas, and most (17 out of 19) area-year observations of Type 2 areas ended up in the “areas in transition” type.

This study has implications for policy development. First, if future studies confirm a potential putative effect of increased markers of SES, further research must tease out which markers can be modified through policy without harming other health outcomes. Property value would seem like a feasible target that is currently the focus of most of the research on gentrification and displacement (Gibbons and Barton, 2016; Whittle et al., 2015). Second, the potential protective effect of diversity (or negative effect of segregation) can be further evaluated and promoted through mixed-income developments (Joseph et al., 2007) or scattered public housing (Pollack et al., 2014). Last, and to put these results into perspective, the strength of the protective association of Type 4 areas (as compared to Type 3 areas) is similar in magnitude to the effect on diabetes incidence of a 1 kg. weight loss in the DPP lifestyle trial (Hamman et al., 2006).

Conclusion

This is, to our knowledge, one of the first studies to look at the association between neighborhood social and economic change and diabetes incidence at a population level. Future research should consider the potential social mechanisms behind these associations and the use of policy measures to mitigate the potential putative effects of area change.

Tables

Table 5.1. Description of study sample by neighborhood change type.

	Variable	Overall	Type 1 (+Diverse)	Type 3 (+Prop. Value)	Type 4 (Aging)	Areas in Transition	p-value
	Sample Size	199621	80578	29400	9544	80099	
Individual-level characteristics	Age (SD)	57.6 (12.8)	57.0 (12.6)	58.7 (12.6)	56.9 (13.0)	57.9 (12.9)	0.492
	Age: 40-50 (%)	36.30%	38.00%	31.80%	40.00%	35.80%	0.018
	Age: 50-60 (%)	24.00%	24.60%	23.50%	22.20%	23.90%	
	Age: 60-70 (%)	19.40%	18.50%	23.50%	18.00%	19.00%	
	Age: 70-80 (%)	14.60%	13.70%	15.60%	14.40%	15.20%	
	Age: 80+ (%)	5.60%	5.20%	5.60%	5.40%	6.10%	
	% Men	43.70%	44.30%	42.70%	44.80%	43.40%	0.027
	% Women	56.30%	55.70%	57.30%	55.20%	56.60%	
	Diabetes Incidence	3.80%	3.90%	3.70%	3.40%	3.70%	0.503
	% with Hypertension	21.20%	21.10%	20.80%	21%	21.60%	0.63
	% with Dyslipidemia	17.50%	17.40%	16.80%	18.30%	17.80%	0.433
% with Any CVD	3.60%	3.60%	3.60%	3.40%	3.60%	0.848	
% with CKD	1%	0.90%	1.10%	0.90%	1.10%	0.492	
% with Retinopathy	0.10%	0.10%	0.10%	0.10%	0.10%	0.071	
Contextual Characteristics	Ciudad Lineal District	33.80%	18.60%	67.50%	26.20%	37.60%	<0.001
	Hortaleza District	31.40%	40.80%	30.80%	23.20%	23.20%	
	San Blas District	26.70%	32.70%	1.70%	28.40%	29.70%	
	Barajas District	8.10%	7.90%	0%	22.20%	9.50%	
	SES Index [IQR]	-0.08 [-0.42; 0.39]	-0.2 [-0.5; 0.3]	0.50 [0.07; 1.06]	0.10 [-0.40; 0.40]	-0.18 [-0.45; 0.25]	<0.001

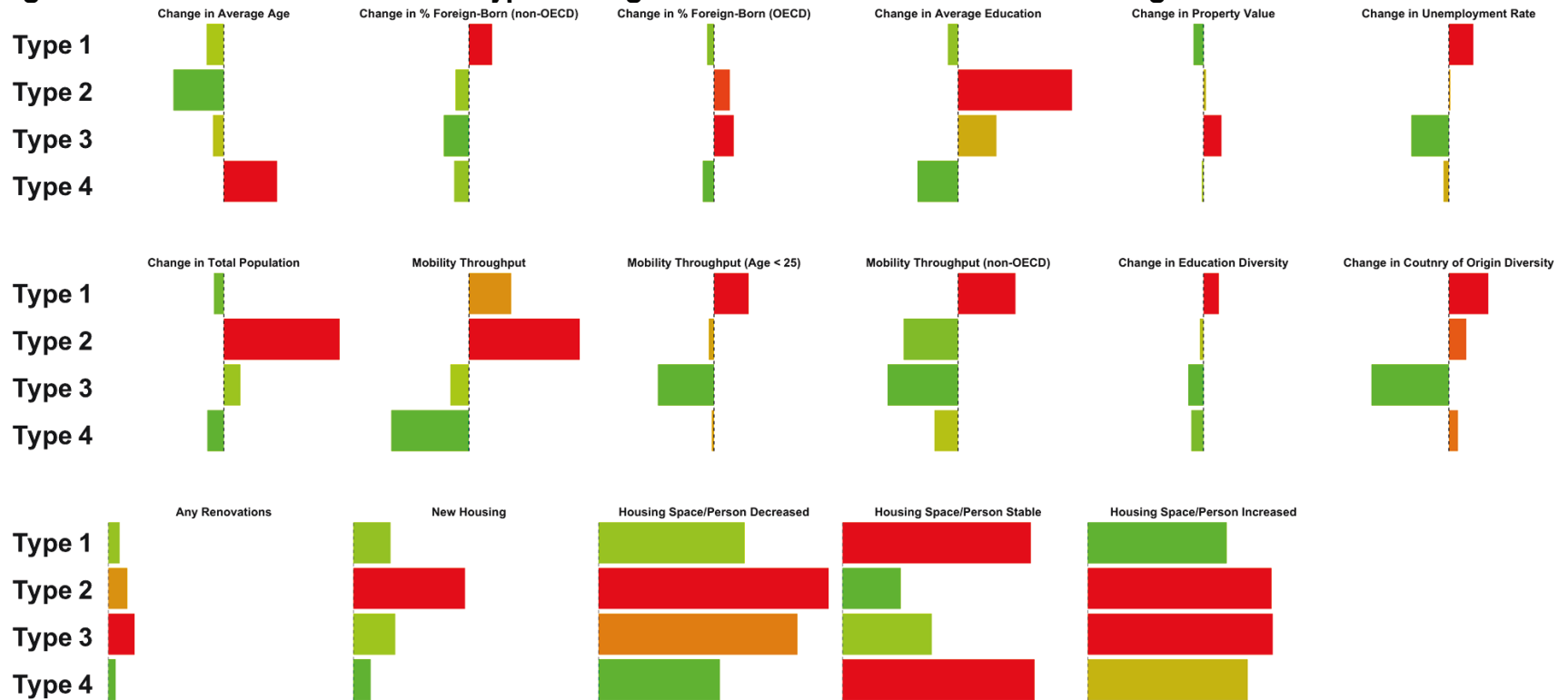
Footnote: all estimates (except diabetes incidence) are characteristics by January 1st 2009. p-values for continuous individual-level characteristics were computed using nested ANOVA; p-values for categorical individual-level characteristics were computed using Donner's χ^2 adjusted for clustered data (at the census section). P-values for contextual characteristics were conducted using χ^2 and ANOVA.

Table 5.2. Association (HR, 95% CI) of Neighborhood Social and Economic Change and Diabetes Incidence

Variable	Model 1 (Unadjusted)	Model 2 (M1+ Age and Sex)	Model 3 (M2 + NBSES)
Type 1 (+Diverse)	1.06 (0.95;1.20)	1.13 (1.02;1.25)	0.90 (0.82;0.99)
Type 3 (+Prop. Value)	1 (Ref.)	1 (Ref.)	1 (Ref.)
Type 4 (Aging)	0.93 (0.76;1.15)	1.00 (0.86;1.17)	0.83 (0.73;0.95)
Areas in Transition	1.03 (0.93;1.15)	1.07 (0.97;1.19)	0.88 (0.80;0.96)
Female		1 (Ref.)	1 (Ref.)
Male		1.76 (1.68;1.85)	1.78 (1.69;1.86)
Age: 40-50		1 (Ref.)	1 (Ref.)
Age: 50-60		2.42 (2.23;2.61)	2.41 (2.23;2.59)
Age: 60-70		3.60 (3.29;3.94)	3.58 (3.30;3.88)
Age: 70-80		4.46 (4.05;4.90)	4.21 (3.86;4.61)
Age: 80+		3.59 (3.23;3.99)	3.34 (3.03;3.70)
NB SES (+1 SD)			0.74 (0.71;0.77)

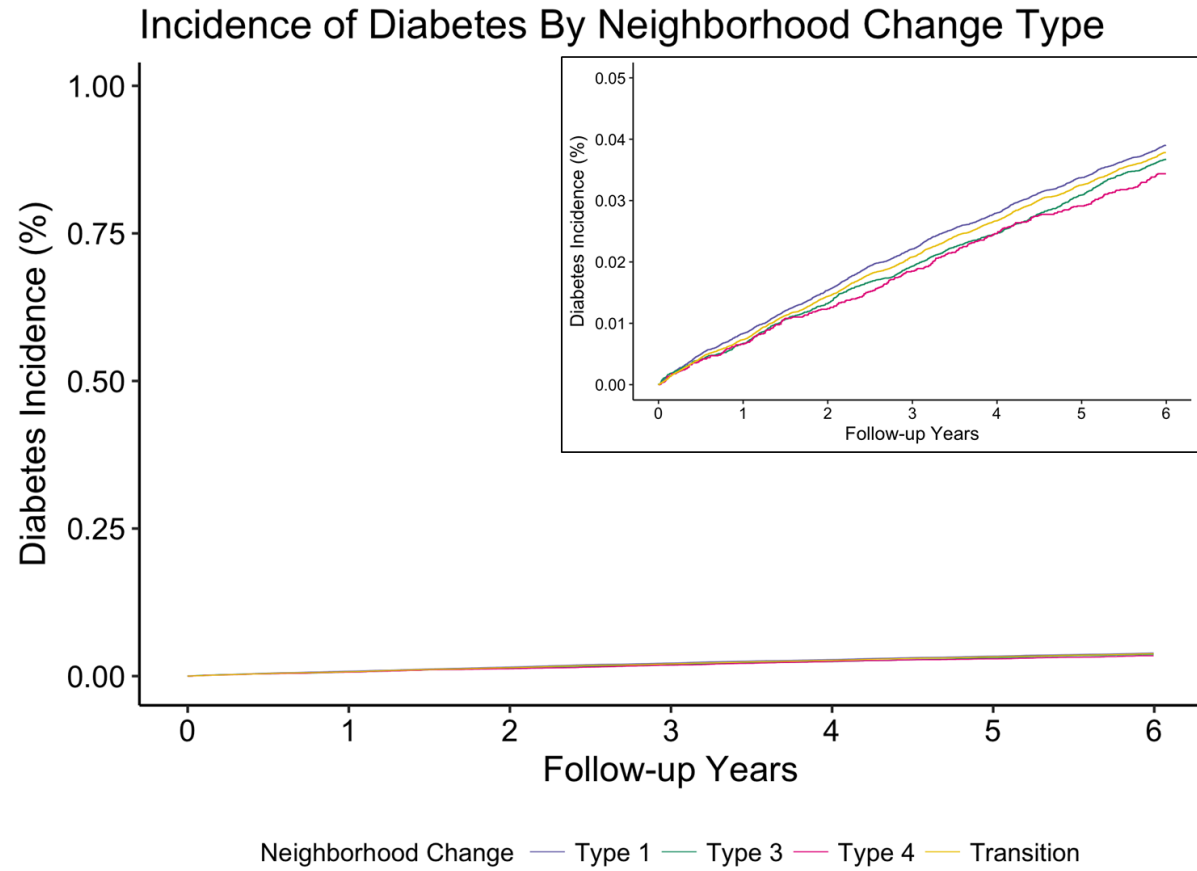
Figures

Figure 5.1: Characteristics of each type of neighborhood social and economic change.



Footnote: graphical representation of Table 3.3. Darker bars are more positive values, while lighter bars are more negative values. In the case of categorical indicators (bottom row), all values are positive, so lighter values represent lower probabilities while darker values represent higher probabilities.

Figure 5.2. Kaplan-Meier Survival Curve of Diabetes Incidence by Neighborhood Social and Economic Change Type.



References

- Afable-Munsuz, A., Mayeda, E.R., Pérez-Stable, E.J., Haan, M.N. (2013). Immigrant Generation and Diabetes Risk Among Mexican Americans: The Sacramento Area Latino Study on Aging. *American journal of public health*. 103 (5), e45-e52.
- Bilal, U., Beltrán, P., Fernández, E., Navas-Acien, A., Bolumar, F., Franco, M. (2016a). Gender equality and smoking: a theory-driven approach to smoking gender differences in Spain. *Tobacco control*. 25 (3), 295-300.
- Bilal, U., Diez, J., Alfayate, S., et al. (2016b). Population cardiovascular health and urban environments: the Heart Healthy Hoods exploratory study in Madrid, Spain. *BMC medical research methodology*. 16, 104.
- Booth, G.L., Creatore, M.I., Moinuddin, R., et al. (2013). Unwalkable neighborhoods, poverty, and the risk of diabetes among recent immigrants to Canada compared with long-term residents. *Diabetes care*. 36 (2), 302-308.
- Catalano, R., Goldman-Mellor, S., Saxton, K., et al. (2011). The health effects of economic decline. *Annu Rev Public Health*. 32, 431-450.
- Chaix, B., Billaudeau, N., Thomas, F., et al. (2011). Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology (Cambridge, Mass.)*. 22 (1), 18-26.
- Connolly, V., Unwin, N., Sherriff, P., Bilous, R., Kelly, W. (2000). Diabetes prevalence and socioeconomic status: a population based study showing increased prevalence of type 2 diabetes mellitus in deprived areas. *Journal of epidemiology and community health*. 54 (3), 173-177.
- Cox, M., Boyle, P.J., Davey, P.G., Feng, Z., Morris, A.D. (2007). Locality deprivation and Type 2 diabetes incidence: a local test of relative inequalities. *Social science & medicine (1982)*. 65 (9), 1953-1964.
- de Burgos-Lunar, C., Salinero-Fort, M.A., Cardenas-Valladolid, J., et al. (2011). Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. *BMC medical research methodology*. 11, 146.
- Denton, E.-g., Green, P., Newman, J., Siqin, Y., Davidson, K.W., Schwartz, J. (2014). The benefits of neighborhood racial diversity: Neighborhood factors and its association with increased physical activity in ACS patients. *International Journal of Cardiology*. 173 (3), 517-518.
- Diez Roux, A.V. (2002). A glossary for multilevel analysis. *Journal of epidemiology and community health*. 56 (8), 588-594.
- Franco, M., Bilal, U., Ordunez, P., et al. (2013). Population-wide weight loss and regain in relation to diabetes burden and cardiovascular mortality in Cuba 1980-2010: repeated cross sectional surveys and ecological comparison of secular trends. *Bmj*. 346, f1515.
- Geraghty, E.M., Balsbaugh, T., Nuovo, J., Tandon, S. Using Geographic Information Systems (GIS) to assess outcome disparities in patients with

- type 2 diabetes and hyperlipidemia. *Journal of the American Board of Family Medicine : JABFM.* 23 (1), 88-96.
- Gibbons, J., Barton, M.S. (2016). The Association of Minority Self-Rated Health with Black versus White Gentrification. *Journal of urban health : bulletin of the New York Academy of Medicine.* 93 (6), 909-922.
- Glass, T.A., Bilal, U. (2016). Are neighborhoods causal? Complications arising from the 'stickiness' of ZNA. *Soc Sci Med,*
- Grigsby-Toussaint, D.S., Lipton, R., Chavez, N., Handler, A., Johnson, T.P., Kubo, J. (2010). Neighborhood Socioeconomic Change and Diabetes Risk: Findings from the Chicago Childhood Diabetes Registry. *Diabetes Care.* 33 (5), 1065-1068.
- Grigsby, W.G. (1987). *The Dynamics of Neighborhood Change and Decline: Pergamon Journals.*
- Hamman, R.F., Wing, R.R., Edelstein, S.L., et al. (2006). Effect of Weight Loss With Lifestyle Intervention on Risk of Diabetes. *Diabetes Care.* 29 (9), 2102-2107.
- Hippisley-Cox, J., O'Hanlon, S., Coupland, C. (2004). Association of deprivation, ethnicity, and sex with quality indicators for diabetes: population based survey of 53,000 patients in primary care. *BMJ (Clinical research ed.).* 329 (7477), 1267-1269.
- Joseph, M.L., Chaskin, R.J., Webber, H.S. (2007). The Theoretical Basis for Addressing Poverty Through Mixed-Income Development. *Urban Affairs Review.* 42 (3), 369-409.
- Kazepov, Y. (2005). *Cities of Europe: Changing Contexts, Local Arrangement and the Challenge to Urban Cohesion: Wiley.*
- Kershaw, K.N., Diez Roux, A.V., Burgard, S.A., Lisabeth, L.D., Mujahid, M.S., Schulz, A.J. (2011). Metropolitan-level racial residential segregation and black-white disparities in hypertension. *Am J Epidemiol.* 174 (5), 537-545.
- Kershaw, K.N., Osypuk, T.L., Do, D.P., De Chavez, P.J., Diez Roux, A.V. (2015). Neighborhood-level racial/ethnic residential segregation and incident cardiovascular disease: the multi-ethnic study of atherosclerosis. *Circulation.* 131 (2), 141-148.
- Larrañaga, I., Arteagoitia, J.M., Rodriguez, J.L., Gonzalez, F., Esnaola, S., Piniés, J.A. (2005). Socio-economic inequalities in the prevalence of Type 2 diabetes, cardiovascular risk factors and chronic diabetic complications in the Basque Country, Spain. *Diabetic medicine : a journal of the British Diabetic Association.* 22 (8), 1047-1053.
- Ludwig, J., Sanbonmatsu, L., Gennetian, L., et al. (2011). Neighborhoods, obesity, and diabetes--a randomized social experiment. *The New England journal of medicine.* 365 (16), 1509-1519.
- Mezuk, B., Chaikiat, Å., Li, X., Sundquist, J., Kendler, K.S., Sundquist, K. (2013). Depression, neighborhood deprivation and risk of type 2 diabetes. *Health & place.* 23, 63-69.

- Müller, G., Kluttig, A., Greiser, K.H., et al. (2013). Regional and neighborhood disparities in the odds of type 2 diabetes: results from 5 population-based studies in Germany (DIAB-CORE consortium). *American journal of epidemiology*. 178 (2), 221-230.
- NCD-RisC (2016). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *The Lancet*. 387 (10027), 1513-1530.
- Piccolo, R.S., Duncan, D.T., Pearce, N., McKinlay, J.B. (2015). The role of neighborhood characteristics in racial/ethnic disparities in type 2 diabetes: results from the Boston Area Community Health (BACH) Survey. *Social science & medicine* (1982). 130, 79-90.
- Pollack, C.E., Green, H.D., Jr., Kennedy, D.P., et al. (2014). The impact of public housing on social networks: a natural experiment. *American journal of public health*. 104 (9), 1642-1649.
- Pollack, C.E., Griffin, B.A., Lynch, J. (2010). Housing affordability and health among homeowners and renters. *American journal of preventive medicine*. 39 (6), 515-521.
- Rose, G. (1985). Sick individuals and sick populations. *Int. J. Epidemiol.* 14 (1), 32-38.
- Sameem, S., Sylwester, K. (2017). The business cycle and mortality: Urban versus rural counties. *Social Science & Medicine*. 175, 28-35.
- Sampson, R.J. (2008). Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure 1. *American Journal of Sociology*. 114 (1), 189-231.
- Tammaru, T., van Ham, M., Marcińczak, S., Musterd, S. (2015). *Socio-Economic Segregation in European Capital Cities: East Meets West*: Taylor & Francis.
- Tapia Granados, J.A., House, J.S., Ionides, E.L., Burgard, S., Schoeni, R.S. (2014). Individual joblessness, contextual unemployment, and mortality risk. *Am J Epidemiol*. 180 (3), 280-287.
- Unwin, N., Hambleton, I.R. (2016). Neighbourhood of residence and the risk of type 2 diabetes. *The Lancet Diabetes & Endocrinology*. 4 (6), 475-476.
- van Ham, M., Manley, D., Bailey, N., Simpson, L., Maclennan, D. (2012). *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research*: Springer Netherlands.
- Weisskopf, M.G., Sparrow, D., Hu, H., Power, M.C. (2015). Biased Exposure-Health Effect Estimates from Selection in Cohort Studies: Are Environmental Studies at Particular Risk? *Environmental health perspectives*. 123 (11), 1113-1122.
- Whittle, H.J., Palar, K., Hufstедler, L.L., Seligman, H.K., Frongillo, E.A., Weiser, S.D. (2015). Food insecurity, chronic illness, and gentrification in the San Francisco Bay Area: An example of structural violence in United States public policy. *Soc Sci Med*. 143, 154-161.

- Zenk, S.N., Mentz, G., Schulz, A.J., Johnson-Lawrence, V., Gaines, C.R. (2016). Longitudinal Associations Between Observed and Perceived Neighborhood Food Availability and Body Mass Index in a Multiethnic Urban Sample. *Health Education & Behavior*,
- Zhang, P., Zhang, X., Brown, J., et al. (2010). Global healthcare expenditure on diabetes for 2010 and 2030. *Diabetes Research and Clinical Practice*. 87 (3), 293-301.
- Zhang, Y.T., Laraia, B.A., Mujahid, M.S., et al. (2016). Is a reduction in distance to nearest supermarket associated with BMI change among type 2 diabetes patients? *Health & place*. 40, 15-20.

CHAPTER 6: CONCLUSION

Summary of Findings

Neighborhood Social and Economic Change

In Aim 1 (Chapter 3) we conducted a measurement model for neighborhood social and economic change in Madrid from 2005 to 2015, based on a latent variable structure that honored the discrete nature of neighborhood change. As predicted by our theory, we found two analytically distinct types of neighborhood change that represented areas with high residential mobility and housing construction and an increase in the proportion of young people (type 2 areas), and areas with less mobility, little new housing and population aging (type 4 areas). A second set of areas in the city were differentiated by the ways in which socioeconomic indicators changed: type 1 areas were at the lower end of education and property value increases and on the higher end of unemployment increases. Type 3 areas were in the opposite end of the spectrum in terms of SES, along with an increased probability of housing renovations.

One important feature of our results from Aim 1 was the finding that there were two sets of types that differed along two distinct dimensions, SES change and residential mobility/housing. Our measurement model focused on the spatial variation in the distribution of indicators at each point in time. However, we also found evidence of a second dimension related to temporal patterns of change and stability. Some areas belonged to the same type of neighborhood change throughout the study period, however the majority transitioned often between change types. Areas with stronger increases in property value (or weaker

decreases during the recession), housing renovations and migration of people from OECD countries tended to remain in their type and did not mix with other types of change. Areas with a high degree of mobility and new housing tended to stay in this type for only one year, and then settle into a different type of change subsequently.

Neighborhood Change and Food Environment Changes

In Aim 2 (Chapter 4) we further explored changes in the food environment and their association with neighborhood social and economic change. Madrid has a dynamic food environment, where at least one third of the areas saw changes in the number of food stores from year to year. Overall, there was an increase in the number of total food stores, with stability or potential decreases in the number of small specialty stores, especially fruit and vegetable stores. These changes differed by neighborhood change type: areas with increasing property values had a decrease in the number of supermarkets and an increase in the number of small specialty stores. This was contrary to our hypothesis; we expected areas with increased property values to see a decrease in small stores and an increase in supermarkets due to increased economic pressures on business.

Neighborhood Change and Diabetes Incidence

Next, in Aim 3 (Chapter 5) we explored the association between neighborhood change and diabetes incidence. Independent of neighborhood socioeconomic status, age and sex, we found a significant increase in diabetes in

areas where property values were increasing, diversity was decreasing, and markers of socioeconomic status were improving. An alternative statement is that diabetes incidence is lower in the comparison group, that is, in areas with a strong increase in average age, low levels of residential mobility, and no new housing. This was consistent with our hypothesis, as we expected areas characterized by increasing property value (or weaker decreases during the recession) to have a higher incidence of diabetes. Future efforts are warranted to determine if these associations are occurring in new or in previous residents.

Challenges

In the following section, we would like to discuss three challenges that emerged from the design, analysis and interpretation of the research conducted in this dissertation. These three challenges are: (1) the issue of scale in the study of types of neighborhood change; (2) priorities in the use of theory and/or data to make analytic decisions; (3) the consequences of data availability and quality in data-driven decisions.

Trajectories, Stages and the MAUP/MTUP

The results of our measurement model of neighborhood social and economic change highlight the discrete nature of change. Most studies in the literature study neighborhood change from the idea of the “trajectory”, although recent developments have begun to call for the study of neighborhood types (Lekkas et al., 2017). The idea of the trajectory assumes a linear trend in neighborhood change, where the entire trajectory of a neighborhood can be

described with a single descriptor for the slope and a single descriptor for the level (i.e., “upwards”, “stable high”, “stable low”, “downwards”). The trajectory model assumes that areas only follow a long-run trajectory of change. Some of the most recent neighborhood change literature (Meen et al., 2013) has shown that areas follow a sort of “stable stochasticity”. This means that while most areas do not change at all, some eventually experience a sudden change. The call by Lekkas et al. (2017) to look at stages of neighborhood development is certainly a step in the right direction and we have followed this new approach here.

Results from our measurement model

In our case, we found dynamic (changing) neighborhood environments. Only two thirds of each spatial unit of observation stayed in the same type in the subsequent year. In the case of areas with very high or very low mobility this number was even lower (6% and 30%). This is certainly not what might be expected from a trajectory perspective, meaning that characterizing an area with a single descriptor would be quite challenging and potentially inaccurate. For this reason, for Chapters 4 and 5 we recategorized areas into those that have a clear change pattern (in each of the four types) and a fifth type. This fifth type, named “areas in transition”, groups areas that have a high degree of transitions between types. In a sense, these areas have a lot of “change in the way they change”. These areas in transition share some characteristics with areas in other types, as they represent a heterogeneous group of areas.

The Signal vs. Noise problem

This finding is an example to the “signal vs. noise” problem: if the areas under study are too small, then we may be capturing noise (increases in age, education, etc.) that are the result of stochastic variations of limited meaning. We aimed to address this by averaging several years of observations and classifying areas using a higher-level descriptor beyond type of change that described areas by how much they changed between types over time. We found that this categorization associated with changes in the food environment and diabetes incidence (independent of age, sex and neighborhood socioeconomic status). However, future research on neighborhood change should address the issue of choosing the most appropriate scale in four dimensions. In particular, two aspects of scale should be addressed:

- Are yearly changes meaningful, or are long-run changes more important for health?
- Is change in a very small spatial scale meaningful or must one consider higher levels of aggregation?

The issue of granularity is a common one in complex systems (Walloth et al., 2016). In particular, there are coarse-graining methods that abstract away pieces of data that do not provide information (but add noise), leaving the researcher with more interpretable patterns to better understand the way the world works. Our attempt at converting census section-year type memberships to longer-term types is an example of this methodology. Future models of change

should wrestle with the delicate balance between longer-term trajectories that are too rigid to accommodate “stable stochasticity” and yearly discrete types of change that are too noisy to provide useful interpretations.

Modifiable Areal Unit Problem (MAUP)

An alternative strategy to deal with this issue is to evaluate change using several scales (temporal and spatial) and to determine ensuing variations in inferences. In geography, this is commonly known as the Modifiable Areal Unit Problem (MAUP) (Fotheringham and Wong, 1991). Its temporal version, the Modifiable Temporal Unit Problem (MTUP) has been recently described (Cheng and Adepeju, 2014). In essence, both issues stem from the differential effects that exposures may have at different scales. One way to check for this issue in this dissertation would have been to estimate the measurement model in Chapter 3 using either neighborhoods (instead of census sections) or longer-term changes (e.g., 5 year changes instead of 1 year changes). If the types of types that we obtain look substantively different, one can argue there is a MAUP/MTUP problem. The reason behind this is the arbitrary selection of a 1 year (or 5-year change, moving averages, smoothers, etc.) or census sections (or neighborhoods) as units of analysis. The MAUP/MTUP is present if inferences vary in such scenarios. In our case, we conducted a sensitivity analysis in Chapter 3 using 2-year moving averages of change. The resulting four types were similar to the ones obtained in the main model. The main difference was a higher degree of temporal stability in type assignment (areas tended to belong to

the same type over time), expected from the ensuing smoothing of change with moving averages.

Theory-driven vs. Data-driven decisions

The second major challenge I found in this dissertation was related to decisions confronted during the construction of the measurement model of Chapter 3. In particular, finite mixture models require the pre-specification of the number of types. As detailed in Chapter 3, there is no absolute objective measure to determine the best model by the number of types. Instead, the analyst should use a combination of data-driven and theory-driven hints. Data-driven markers include measures of goodness-of-fit (with the Bayesian Information Criterion as the main one for finite mixtures) and measures of classification (Entropy). Theory-driven decisions are made on *a-priori* considerations and the interpretability of types.

Data-driven approach

In our model, measures of model fit improved with the addition of new types. On the other hand, measures of classification (entropy) worsened as the number of types increased. In this problem of optimization, a usual approach is to construct a summary measure of fit and classification such as the ICL-BIC (McLachlan and Peel, 2004). Nonetheless, the decision of how to weight each component is essentially trivial, and no study has shown a clear advantage of the ICL-BIC over using BIC or entropy separately.

Theory-driven approach

A second way of optimizing is to fall back on previous theoretical considerations and observe how well the model fits the measurement theory informed by the literature. In our case, we theorized a 4-type model. The interpretability of each type of the 4-type model was detailed in Chapter 3 and used in chapters 4 and 5, and we believe it offers enough differentiation to be useful. An argument could be made for a reduction down to a 3-type model, given the very low prevalence (~3%) of type 2 areas. While this is a valid argument, the maximization of the likelihood of finite mixture models includes two components: the type membership and the type characteristic components. A type that is either very prevalent or very different as compared to other types will emerge in models sooner, as the number of types increases. That type 2 has a very low prevalence does not preclude that it is sufficiently different to warrant its inclusion. In particular, when building a 3-type model with the same indicators, type 2 persisted and types 1 and 4 merged into one group. Given the large differences in the interpretations of types 1 and 4, we believe in the validity of the pre-specified 4-type model.

Data Availability and Data-Driven Decisions

The third major challenge I encountered in this dissertation was related to decisions made around data availability. At first, this study proposed to look at three components of neighborhood change: residential mobility (measured through changes in occupational structure of new residents), residential

immobility (measured through changes in unemployment in old residents or stayers), and external actors (measured through changes in budgetary allocations to each area). These were to be correlated with food environment changes and health outcomes. Food environment changes were to be obtained from business registries, and health outcomes from electronic health records.

Two challenges then became apparent. First, health data was available at two levels. The less granular level was at the health center to which each patient was assigned from 2009 to 2014. The more detailed level was at the census section of residence from 2013 to 2014. As detailed in Chapter 2, the health center is not an ideal proxy for area of residence, however it closely follows “basic health areas”, which were the spatial division of health centers up until the late 2000s in Madrid. People are able to freely choose to use a different health center, and therefore while each health center has a spatially explicit catchment area, this may not be entirely accurate. Health data was also available at the census section level, that are then nested into the basic health areas.

The second challenging spatial data was food environment data. Business registries have a policy of not releasing data on areal units containing less than 5 stores. For data on total business by census section this was not entirely problematic, as only 2% of all census sections were missing data, and these could be imputed by subtracting from the total number of business in the neighborhood. For data on fruit and vegetable stores by census section this meant that 95% of all census sections had a missing value, no longer leaving

imputation as a viable option. The only solution here was to request data at the neighborhood level in which census sections are nested.

Neighborhoods and Basic Health Areas do not overlap with each other, and therefore it would be challenging to perform an analysis at this less granular level. The approach we used was to do all Neighborhood Change analysis at the census section unit. This way, all measures could then be aggregated up to both the Basic Health Area and the Neighborhood level for each aim. With this in mind, we set to look for indicators for Chapter 3 of this dissertation. The indicators that were originally proposed were only available at the neighborhood (or district) level, so a new measurement model had to be developed. This led to the discovery of several data sources that had not been used for public health research in the past, including the residential mobility data of the continuous census and the real estate tax registry.

However, with the advantages inherent in using smaller spatial (census section, $n \sim 2400$) and temporal units (yearly changes, $n \sim 10$ year-to-year changes from 2005 to 2015), there also came some disadvantages. By looking at yearly changes in census sections we were able to capture a very granular description of changes at the spatial and temporal level. However, changes at these granular levels may also have an unfavorable signal to noise profile. A change in inference when looking at different scales (to improve the signal to noise profile), may hint at the presence of the MAUP/MAUP (see previous challenge).

Implications and Next Steps

I would like to highlight a few key research implications derived from this dissertation and how they provide a pathway for future studies. The implications are divided in two main overarching themes: what I have learned from this process (the importance of theory development); and what surprised me during this process and how it informs future research (the importance and challenges of the study of dynamic relationships).

Lesson Learned: Theory Development (Popper goes to Santa Fe)

Epidemiology is the study of the distribution and causes of disease in populations (Porta, 2014). As with any other scientific discipline, there is a constant struggle between induction and deduction in the production of knowledge (Rothchild, 2006). Briefly, induction relies on gathering facts and pieces of knowledge and building a set of hypothesis (to be tested) or a theory that explains the findings (Rothchild, 2006). Deduction begins with theory development for hypothesis formulation, allowing for the potential rejection of these hypotheses to help refine the theory based on the results of subsequent testing (Rothchild, 2006).

Epidemiology, Deduction and Induction

According to Buck (1975), epidemiology would be better served by adopting a deductive approach that formulates hypothesis and then tries to refute them. According to Jacobsen (Jacobsen, 1976) this discussion is moot, since both approaches feed each other:

“The element of deduction in the formulation of the hypotheses does not make induction irrelevant, any more than the essential inductive inferences from observed data negate the importance of prior and subsequent deductions” (Jacobsen, 1976).

In fact, this iterative process was already suggested as the main driver of epidemiologic findings by Frost (in his introduction to John Snow (Snow et al., 1936)):

EPIDEMIOLOGY at any given time is something more than the total of its established facts. It includes their orderly arrangement into chains of inference which extend more or less beyond the bounds of direct observation. Such of these chains as are well and truly laid guide investigation to the facts of the future; those that are ill made fetter progress. But it is not easy, when divergent theories are presented, to distinguish immediately between those which are sound and those which are merely plausible. Therefore it is instructive to turn back to arguments which have been tested by the subsequent course of events; to cultivate discrimination by the study of those which the advance of definite knowledge has confirmed.

In any case, both approaches meet at the theory development stage, whether deriving from this stage or guiding and organizing data analysis. This is where I believe we, as academic Epidemiologists, could follow Frost’s prescription more closely, creating “orderly arrangement into chains of inference [of established facts] which extend more or less beyond the bounds of direct observation” (Snow et al., 1936). Epidemiology, over the last few decades, has created a strong preference for very specific and narrow hypotheses to be tested in experimental or observational settings (Schwartz et al., 2016). If these hypotheses are successfully verified, then a new piece of knowledge is created, giving strength to an argument for the design of an intervention to affect a specific risk factor or exposure. Nonetheless, the corpus of epidemiologic knowledge of any given disease (or exposure) is rarely examined for the creation

of a unifying theory that links those pieces of knowledge, creating an integrated narrative where the missing pieces can be inferred “beyond the bounds of direct observation”, as Frost would put it. It is even rarer to find a theory that precedes these studies, even after considering all the “calls for theory” in the literature over the past few decades (Diez Roux, 2007; Krieger, 2001; Roux, 2012), and even the more recent calls for a re-weighting of the importance of data and theory in epidemiology (Hernán, 2014; Marshall and Galea, 2014). Epidemiology calls for theories, but rarely tries to build them (either a priori or post-hoc). We are stuck in a world between deduction in induction, with no grand unifying theory (GUT) to guide us, but no ability (or training) for true data-driven induction.

The Theory Sandbox

In working on this dissertation, I have found creating a “theory sandbox” to be of great use. A “theory sandbox” is a strong narrative of causal connections where one can alter a factor and predict what the results may be (a sort of “penny sorter” where one drops an effect in one side and waits to see where it ends up). In dealing with the difficult choice of the number of types in a finite mixture model (a task with no absolute answer), reverting back to the proposed theory of neighborhood change provided answers otherwise unattainable by direct observation. The challenges derived from the selection of the number of types were described above in detail. All measurement models are representations of the way we think a phenomenon works. Measures of fit provide an idea of how well our data behaves, if our model was true. Combining

these measures with interpretations of how well our results fit our well-informed theory, may provide much better descriptions of the phenomenon of interest. In fact, the theoretical framework behind this dissertation predicted that there would be four main types of neighborhood social and economic change (two directions of change by two intensities of throughput). The results of Aim 1 (Chapter 3) were more nuanced than this two by two model, but it resembled the same structure.

An important lesson is that simplistic ideas about “directions of change” (“improvement” vs “worsening”) may not be sufficient ways of describing these complex phenomena. For example, two of the types of change (2 and 3) showed an “improvement” in education levels, but only one of them (type 3) had clear increases in property value and decreases in unemployment. The main difference is that a younger group of people migrated to one of the areas, bringing the education level up (because of cohort effects), but not necessarily increasing socioeconomic status. Regardless of these differences with our original theory (that now merits refinement), this process of theory development and use proved helpful in modeling change and interpreting the results.

Developing a Theory

Unanswered by the statement above is the question of how to generate such theory. For this process, I was inspired by two classical pieces of social epidemiologic literature: 1964 Sydney Kark’s paper “The Social Pathology of Syphilis in Africans” (Kark, 2003) and 1994 Nancy Krieger’s paper “Epidemiology

and the web of causation: has anyone seen the spider? (Krieger, 1994). These two papers posed two separate questions: “where’s the gold mine?” and “who is the spider?”.

The gold mine is the source of impetus for institutional actors to shape a social factor (Kark, 2003). In a capitalist society, the source of impetus is usually profit (Marx, 1976). If one looks at the sources of profit, one can find the decisions behind the shaping of many health-affecting factors. For example, if we assume that the development of ultra-processed foods is behind the obesity epidemic, a “where’s the gold mine?” approach would not look at the consequences of ultra-processed foods (UPF), or the reasons behind their consumptions by individuals, but would rather focus on: (a) discovering whether UPF are a gold mine (more profit to be extracted from them than from other calorie sources); and (b) what about their “gold mine”-ness makes them damaging. In the case of ultra-processed foods, it’s the removal of labor from the food production process that opens the door for increased profit margins and decreased food prices (relative to unprocessed foods).

Finding the spider involves identifying those profiting from the gold mine (Krieger, 1994). The food industry that is behind ultra-processed foods alters many policies and regulations, has a direct effect through marketing and product formulation, and even funds research with dubious intents (Bes-Rastrollo et al., 2014; Lesser et al., 2007). In finding the gold mine and the spider, one can start developing a theory that has causal forces emanating from them or through

them. In discovering these actors and their profit sources, theories become more complete and can point at specific policy levers that take into account the agency of the spider.

Figure 6.1 shows an example of this process. Starting with the undeveloped version of the theory framework (Panel A), I conducted an analysis of Goldmines/Spiders (Panel B). This revealed a preponderance of housing and real estate developers in the neighborhood change phenomenon. With these in mind, I reconstructed the framework in Panel A to clearly locate the "spiders" or drivers of neighborhood change (see Panel C). This proved to be helpful in the later search for indicators of neighborhood social and economic change, and provided exemplars of how areas that are changing may look. Last, the consequences of change were developed in Panel D, where I articulated how change is related to changes in the food environment and the development of chronic disease.

This process led to the development of neighborhood typologies, some of which shared common characteristics in their consequences on health. Neighborhood typologies are a useful device, since they already hint at the necessity for a method that accommodates discrete types instead of continuous latent variables of change. In our case, the typologies referred to the different ways in which neighborhoods can change and the resulting directions of change. These typologies highlight the differences between, for example, truly stable areas (no inflows or outflows) and areas in equilibrium (inflows equal to outflows).

The population output is different, but the inputs (and hence, the consequences) are different.

The nuances about the study of change, its inputs, outputs and the implications for epidemiologic research constitute the next research implication.

The Surprise: The Study of Dynamic Relationships and Throughput

Non-linear dynamics is the study of change in complexity theory. Among the key concepts in non-linear dynamics is **throughput**, or the amount of energy or matter flowing through a system. When throughput is increased, complex systems may enter chaotic states where future behavior is unpredictable (Mitchell, 2009). A key concept of this dissertation is the potential negative effects of an increase in throughput in an urban system.

Throughput and the City

In neighborhood change, throughput can be many things (e.g. traffic, people, houses, etc.), but one of them is the number of people involved in residential mobility (both in-migration and out-migration from an area). Focusing on throughput is a departure from focusing on differentials or deltas: instead of focusing on the balance of people going in versus people going out, throughput focuses on the actual sum of the two flows. This allows for the measurement of neighborhood change as a function of the total number of people involved in the process of mobility, instead of just the balance. These flows can be modified by the amount of new housing or the housing turnover through renovations that occur in an area. These factors, especially residential mobility throughput,

showed clear differences between types of change in our measurement model of neighborhood social and economic change. Residential mobility throughput identified areas with instability that were more prone to have new housing (or to renovate previous housing).

In this dissertation, we showed how some of the neighborhood change types representing increased throughput are associated with poorer health outcomes through increased diabetes incidence. However, we did not have the ability to explore differential patterns of diabetes incidence in Type 2 areas, those with the highest throughput and housing constructions. Therefore, future research should explore the health consequences of living in these types of areas and test whether interventions that reduce throughput may have the power to decrease diabetes incidence.

Throughput and Change

Controlling throughput is one of the key mechanisms to the control of systems (Hübler, 2005). The study of throughput should be linked to the study of change (dynamics). The amount of change in a system can be directly related to throughput, and therefore controlling change can be a way of controlling throughput and hence controlling systems. Change can occur in different ways, and can be temporally classified into short- and long-run changes, with their associated throughput levels (i.e. long-term change may involve a more sustained low-intensity energy flow while short-term change may involve spikes in energy that increase instability and lead to chaotic states). More importantly,

systems are adaptive, but these adaptation processes may be sensitive to the speed of change. When given time, systems may adapt to long-run changes and undergo throughput fluctuations without becoming chaotic. If, on the other hand, change occurs rapidly and with high intensity, systems may be more prone to instability. This dissertation focuses on short-term change, but also acknowledges that this phenomenon of short-term change can be the result of long-term changes with threshold effects.

Putative Social Mechanisms of Increased Throughput

Increased throughput in an urban system can affect population health through several mechanisms. These include: (1) decreased social cohesion due to increased population turnover; (2) increase in traffic-related noxious factors, such as noise and pollution; (3) changes in the food environment and other parts of the retail environment leading to decreased trust in retailers due to the breakages of social bonds; (4) increased local inequality due to the in-migration of new residents in different parts of the socioeconomic spectrum; and (5) increase in housing insecurity due to rising property value or pressure to increase housing turnover. These are just a few examples of mechanisms that are empirically testable with the right data. In this dissertation, we were only able to partially test the third mechanism, changes in the food environment. Future research should refine our measure of food environment changes and formally test how they change with neighborhood change.

Policy Implications

The policy implications of his dissertation are complex and require future studies with improved data on several factors. This section discusses three potential angles for policy development derived from the results of this dissertation. The first one, on controlling throughput, is directly derived from the section above. The following two are related to recent policy changes that occurred in Madrid (or in Spain in general) in the last two decades, and that could be studied with the methods discussed in this dissertation.

Controlling Throughput

Interventions to reduce throughput are warranted if we assume a putative causal connection between increased throughput and health. Even without such assumption, if we are able to predict what kind of policy levers reduce throughput we can evaluate existing interventions or policies as natural experiments and assess whether they have an effect on health.

As stated above, a key parameter to control throughput is to control change. In the case of Neighborhood Social and Economic Change, reducing population turnover is one mechanism for reducing mobility throughput. Assuming a balance in population so that neighborhoods do not become over or under-populated, a reduction in throughput necessarily goes through a reduction in the in- and out-migration to the area, which is strongly related to the amount of housing transactions occurring. Incentives for transactions include market-based housing transactions associated with housing turnover: if each housing

transaction (through new rentals or sales) generates profit, the market will find ways to increase this turnover and therefore generate putative health outcomes. Ideas to achieve this include: (1) rent control measures, that protect tenants against abusive profit extraction (Autor et al., 2014); (2) mixed-income developments, that reduce segregation and stop feedback cycles that increase or decrease property value excessively (Joseph et al., 2007) (although a number of authors have expressed concerns on this point (Lipman, 2008)); (3) measures against housing speculation, including penalties for empty housing (Beswick et al., 2016); and (4) regulation of predatory lending/subprime loans to reduce overinclusion in the lending market (Aalbers, 2013).

The Consequences of Removing Land-use Regulations

The first of the two policy changes that occurred in Madrid/Spain in the last decades relates to zoning regulations. Zoning laws are an understudied but potentially important determinant of health in cities. Previous work has pointed to zoning policies as key drivers of urban dynamics. For example, class and race segregation patterns have been observed to be influenced by zoning policies in 50 US metropolitan areas (Rothwell and Massey, 2010). More importantly, an article that studied the determinants of the distribution of the food environment in urban areas of the British Columbia found that zoning policies “are major processes determining how food outlets become distributed” (Black et al., 2011). The same study found little explanatory power of food environment disparities by

current neighborhood socioeconomic status while zoning policies and other urban form variables had major effects on these disparities (Black et al., 2011).

The authority regarding zoning regulations in Spain was transferred to regional governments in 1990, who in most cases let local governments design their own city-specific plans (González Pérez, 2007). Cities enact zoning policies every few decades, and these regulate land use in every city block. In 1997 the new Law to Liberalize Land Use removed all restrictions on urbanization (except for protected lands). These changes, along with a reduction in interest rates, created a surge in developments in many Spanish cities, including the creation of suburban residential areas (to the image of US or UK suburbs) where food availability may be reduced (Munoz, 2003). Some of the areas in our analysis, particularly those in type 2 of neighborhood change, were the most affected by these new regulations as they were newly built on previously undeveloped lands. Given that these regulatory changes occurred in 1997, understanding the effects of this policy is challenging as our data only spans the period 2005-2015. Nonetheless, future studies using this neighborhood change model should try to assess the consequences on health (or other indicators) of living in a type 2 area.

The Consequences of Removing Restrictions on Opening Hours

The second large policy change is the removal of restrictions regarding opening hours for retail businesses. The restrictions on business hours in Spain is aimed at protecting workers' rights and small businesses. The main motivation is that large corporations have a higher capacity to hire workers for longer hours

of operation. All businesses were allowed open for up to 72 hours every week and up to 12 Sundays (or Holidays) per year. The Madrid Regional Government (that has authority over the municipality of Madrid) changed this in 2012. This new regulation freed opening hours entirely, increased the number of Sundays (or Holidays) when a business can open and removed these restrictions entirely for some specific food stores (de Rada and González, 2015). As mentioned in Chapter 4, the kind of data that we had at our disposal did not allow us to evaluate the consequences of this policy on opening hours. Future research might use proxies for opening hours (such as store size or business load), or conduct field audits to gather information on opening hours. Recent research has highlighted the importance of considering temporal accessibility as a determinant of healthy food availability (Widener et al., 2011; Widener and Shannon, 2014).

Future research should exercise caution in not attributing increased accessibility to food stores an innate positive quality. In particular, accessibility to food stores selling both healthy and unhealthy foods (as is the case of supermarkets) can increase both healthy and unhealthy food availability. Under this scenario, variations in shopping behaviors by the time of the day should be given careful considerations. In particular, previous research on behavioral economics has shown that unhealthy food purchases are facilitated when both money (Thomas et al., 2010) and time (Park et al., 1989) constraints are levied. Gaining an understanding of these processes may not be feasible using

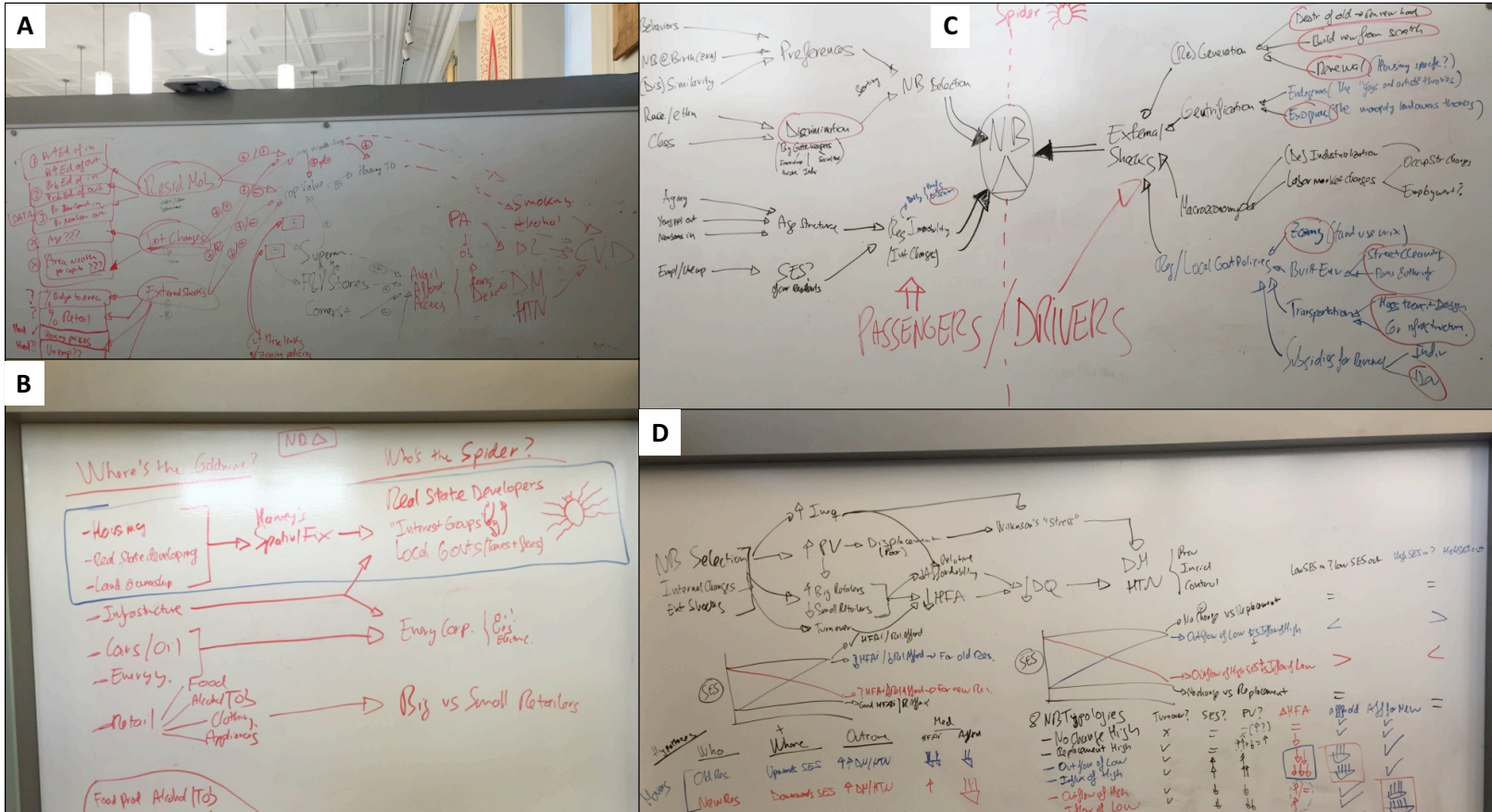
quantitative data and may require more intensive methods of research (Dunn, 2012), such as some qualitative techniques.

Conclusion

This dissertation explored the measurement of neighborhood social and economic change and its connection with food environment changes and diabetes incidence. We showed how a latent variable model that acknowledges the discrete nature of change is a feasible way to measure neighborhood social and economic change. We explored changes in the food environment, finding that areas with indicators of increased SES and housing renovations and reduced diversity had an increase in the number of small stores and a reduction in the number of supermarkets. This finding was contrary to our initial hypothesis. However, we found support for our hypothesis of increased diabetes incidence in these same areas. Future research should explore potential social mechanisms behind these associations and policy-levers to control neighborhood change and their consequences on health outcomes.

Figures

Figure 6.1: Theory Development



References

- Aalbers, M.B. (2013). How Do Mortgage Lenders Influence Neighbourhood Dynamics? Redlining and Predatory Lending. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research* pp. 63-85). Dordrecht: Springer Netherlands.
- Autor, D.H., Palmer, C.J., Pathak, P.A. (2014). Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts. *Journal of Political Economy*. 122 (3), 661-717.
- Bes-Rastrollo, M., Schulze, M.B., Ruiz-Canela, M., Martinez-Gonzalez, M.A. (2014). Financial Conflicts of Interest and Reporting Bias Regarding the Association between Sugar-Sweetened Beverages and Weight Gain: A Systematic Review of Systematic Reviews. *PLOS Medicine*. 10 (12), e1001578.
- Beswick, J., Alexandri, G., Byrne, M., et al. (2016). Speculating on London's housing future. *City*. 20 (2), 321-341.
- Black, J.L., Carpiano, R.M., Fleming, S., Lauster, N. (2011). Exploring the distribution of food stores in British Columbia: Associations with neighbourhood socio-demographic factors and urban form. *Health & place*. 17 (4), 961-970.
- Buck, C. (1975). Popper's philosophy for epidemiologists. *International journal of epidemiology*. 4 (3), 159-168.
- Cheng, T., Adepeju, M. (2014). Modifiable Temporal Unit Problem (MTUP) and Its Effect on Space-Time Cluster Detection. *PloS one*. 9 (6), e100465.
- de Rada, V.D., González, R.L. (2015). La liberalización de los horarios comerciales. Un análisis de la opinión de la sociedad española. *Inguruak* (59),
- Diez Roux, A.V. (2007). Neighborhoods and health: where are we and where do we go from here? *Revue d'Épidémiologie et de Santé Publique*. 55 (1), 13-21.
- Dunn, J.R. (2012). "Explanation," Philosophy and Theory in Health Inequalities Research: Towards a Critical Realist Approach. In P. O'Campo, & J.R. Dunn (Eds.), *Rethinking Social Epidemiology: Towards a Science of Change* pp. 23-42). Dordrecht: Springer Netherlands.
- Fotheringham, A.S., Wong, D.W.S. (1991). The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A*. 23 (7), 1025-1044.
- González Pérez, J.M. (2007). Urban Planning System in Contemporary Spain. *European Planning Studies*. 15 (1), 29-50.
- Hernán, M.A. (2014). Invited Commentary: Agent-Based Models for Causal Inference—Reweight Data and Theory in Epidemiology. *American Journal of Epidemiology*. 181 (2), 103-105.

- Hübler, A.W. (2005). Predicting complex systems with a holistic approach: The “throughput” criterion. *Complexity*. 10 (3), 11-16.
- Jacobsen, M. (1976). Against popperized epidemiology. *International journal of epidemiology*. 5 (1), 9-11.
- Joseph, M.L., Chaskin, R.J., Webber, H.S. (2007). The Theoretical Basis for Addressing Poverty Through Mixed-Income Development. *Urban Affairs Review*. 42 (3), 369-409.
- Kark, S.L. (2003). The social pathology of syphilis in Africans. *International journal of epidemiology*. 32 (2), 181-186.
- Krieger, N. (1994). Epidemiology and the web of causation: has anyone seen the spider? *Social science & medicine*. 39 (7), 887-903.
- Krieger, N. (2001). Theories for social epidemiology in the 21st century: an ecosocial perspective. *International journal of epidemiology*. 30 (4), 668-677.
- Lekkas, P., Paquet, C., Howard, N.J., Daniel, M. (2017). Illuminating the lifecourse of place in the longitudinal study of neighbourhoods and health. *Social Science & Medicine*,
- Lesser, L.I., Ebbeling, C.B., Goozner, M., Wypij, D., Ludwig, D.S. (2007). Relationship between Funding Source and Conclusion among Nutrition-Related Scientific Articles. *PLOS Medicine*. 4 (1), e5.
- Lipman, P. (2008). Mixed-income schools and housing: advancing the neoliberal urban agenda. *Journal of Education Policy*. 23 (2), 119-134.
- Marshall, B.D.L., Galea, S. (2014). Formalizing the Role of Agent-Based Modeling in Causal Inference and Epidemiology. *American Journal of Epidemiology*. 181 (2), 92-99.
- Marx, K. (1976). *Capital: A Critique of Political Economy*: Penguin Books Limited.
- McLachlan, G., Peel, D. (2004). *Finite mixture models*: John Wiley & Sons.
- Meen, G., Nygaard, C., Meen, J. (2013). The Causes of Long-Term Neighbourhood Change. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research* pp. 43-62). Dordrecht: Springer Netherlands.
- Mitchell, M. (2009). *Complexity: A guided tour*: Oxford University Press.
- Munoz, F. (2003). Lock living: Urban sprawl in Mediterranean cities. *Cities*. 20 (6), 381-385.
- Park, C.W., Iyer, E.S., Smith, D.C. (1989). The Effects of Situational Factors on In-Store Grocery Shopping Behavior: The Role of Store Environment and Time Available for Shopping. *Journal of Consumer Research*. 15 (4), 422-433.
- Porta, M. (2014). *A dictionary of epidemiology*: Oxford University Press.
- Rothchild, I. (2006). Induction, deduction, and the scientific method. *Soc. study Reprod*,
- Rothwell, J.T., Massey, D.S. (2010). Density zoning and class segregation in U.S. metropolitan areas. *Social science quarterly*. 91 (5), 1123-1143.

- Roux, A.V.D. (2012). Conceptual approaches to the study of health disparities. *Annual review of public health*. 33, 41.
- Schwartz, S., Prins, S.J., Campbell, U.B., Gatto, N.M. (2016). Is the “well-defined intervention assumption” politically conservative? *Social Science & Medicine*. 166, 254-257.
- Snow, J., Frost, W.H., Richardson, B.W. (1936). *Snow on cholera*: Commonwealth Fund.
- Thomas, M., Desai, K.K., Seenivasan, S. (2010). How Credit Card Payments Increase Unhealthy Food Purchases: Visceral Regulation of Vices. *Journal of Consumer Research*. 38 (1), 126-139.
- Walloth, C., Gebetsroither-Geringer, E., Atun, F. (2016). Introduction: Overcoming Limitations of Urban Systems Models and of Data Availability. In C. Walloth, E. Gebetsroither-Geringer, F. Atun, & L.C. Werner (Eds.), *Understanding Complex Urban Systems: Integrating Multidisciplinary Data in Urban Models* pp. 1-14). Cham: Springer International Publishing.
- Widener, M.J., Metcalf, S.S., Bar-Yam, Y. (2011). Dynamic urban food environments a temporal analysis of access to healthy foods. *American journal of preventive medicine*. 41 (4), 439-441.
- Widener, M.J., Shannon, J. (2014). When are food deserts? Integrating time into research on food accessibility. *Health & place*. 30, 1-3.

LIST OF APPENDICES

Appendix A: Appendices for Chapter 3 (Aim 1)	214
Appendix B: Appendices for Chapter 4 (Aim 2)	218
Appendix C: Appendices for Chapter 5 (Aim 3)	222

Appendix A: Appendices for Chapter 3 (Aim 1)

Appendix 3.1: Detailed results of the final measurement model of neighborhood social and economic change: Means and Probabilities

Indicator	Type 1	Type 2	Type 3	Type 4
Δ Mean Age	-0.19	-0.56	-0.12	0.60
Δ Proportion Foreign-Born in non-OECD	0.26	-0.15	-0.28	-0.16
Δ Proportion Foreign-Born in OECD	-0.08	0.18	0.22	-0.13
Δ Mean Education Level	-0.11	1.27	0.43	-0.45
Δ Property Value	-0.11	0.03	0.20	-0.02
Δ Unemployment Rate	0.27	0.02	-0.42	-0.06
Δ Total Population	-0.11	1.29	0.18	-0.18
Mobility Throughput	0.47	1.24	-0.21	-0.87
Δ Education Diversity	0.17	-0.04	-0.17	-0.14
Δ Country of Origin Diversity	0.44	0.19	-0.86	0.10
Mobility Throughput (age <25)	0.39	-0.06	-0.63	-0.03
Mobility Throughput (non-OECD foreign)	0.64	-0.61	-0.79	-0.26
<i>Any Renovation</i>	0.02	0.04	0.06	0.02
<i>New Housing</i>	0.08	0.24	0.09	0.04
Δ Housing Space / Person (Decrease)	0.31	0.49	0.42	0.26
Δ Housing Space / Person (Stable)	0.40	0.12	0.19	0.41
Δ Housing Space / Person (Increase)	0.29	0.39	0.39	0.34

Footnote: numbers for continuous indicators reflect mean value of the distribution of each type; numbers for categorical indicators (in *italics*) reflect probabilities for each type

**Appendix 3.2: Detailed results of the final measurement model of neighborhood social and economic change:
Variances**

Indicator	Type 1	Type 2	Type 3	Type 4
Δ Mean Age	0.87	2.79	0.80	0.72
Δ Proportion Foreign-Born in non-OECD	0.99	1.72	0.74	0.66
Δ Proportion Foreign-Born in OECD	0.83	3.47	1.04	0.72
Δ Mean Education Level	0.86	3.95	0.76	0.42
Δ Property Value	0.65	0.93	1.77	0.74
Δ Unemployment Rate	0.91	2.23	0.80	0.88
Δ Total Population	0.58	8.72	0.78	0.54
Mobility Throughput	0.91	2.01	0.51	0.21
Δ Education Diversity	1.00	1.61	0.91	1.04
Δ Country of Origin Diversity	0.62	6.40	0.61	0.17
Mobility Throughput (age <25)	0.79	1.57	0.45	1.22
Mobility Throughput (non-OECD foreign)	0.50	0.89	0.48	0.93
<i>Any Renovation</i>	N/A	N/A	N/A	N/A
<i>New Housing</i>	N/A	N/A	N/A	N/A
<i>Δ Housing Space / Person (Decrease)</i>	N/A	N/A	N/A	N/A
<i>Δ Housing Space / Person (Stable)</i>	N/A	N/A	N/A	N/A
<i>Δ Housing Space / Person (Increase)</i>	N/A	N/A	N/A	N/A

Footnote: N/A: variances are only available for continuous indicators.

**Appendix 3.3: Detailed results of the final measurement model of neighborhood social and economic change:
Covariance structure (expressed as correlations)**

Indicator	Δ Age	Δ Non-OECD	Δ OECD	Δ Educ.	Δ Prop. Value	Δ Unemp.	Δ Pop.	Mobil. Throughp.	Δ Ed. Divers	Δ Origin Divers.	Mobil (<25)	Mobil (Non-OECD)
Δ Age												
Δ Non-OECD	-0.315											
Δ OECD												
Δ Educ.	-0.159	0.009										
Δ Prop. Value												
Δ Unemp.												
Δ Population	-0.402	0.380	0.115*	0.070								
Mobility Throughp.		-0.011		0.010								
Δ Ed. Divers.	-0.298	0.693	0.410				0.353					
Δ Origin Divers.												
Mobility (<25)												
Mobility (non-OECD)								0.158				0.025

Footnote: all values are expressed as correlations (covariance (X1, X2) / (SD(X1) * SD(X2))). All correlations (except for *) are equal across types. *: this correlation is specific to Type 1 (and 0 for all other types). All missing cells are correlations of 0 (not estimated). Only the lower part of the correlation matrix is shown.

Appendix 3.4: Detailed results of the final measurement model of neighborhood social and economic change: Type Membership

	Type 1	Type 2	Type 3	Type 4
Type Prevalence (average posterior)	46%	3%	27%	24%
SES Index (+1 SD)	0.3***	0.71*	17.7***	1 (Ref.)
Epoch (2010-2015)	1.2*	1.1	2.1***	1 (Ref.)

Footnote: predictors of type membership (**bolded**) reflects odds of membership as compared to Type 4. Type prevalence reflects the averaged posterior over all census sections. *: p<0.05, **: p<0.01,, ***p<0.001.

Appendix B: Appendices for Chapter 4 (Aim 2)

Appendix 4.1: Distribution of Neighborhood Change Types over time

	Type 1	Type 2	Type 3	Type 4	Areas in Transition
Average 2006-2011	31.40%	2.90%	16.70%	5.20%	43.70%
Average 2012-2015	31.10%	6.10%	16.50%	6.90%	39.40%
2011	45.30%	3.10%	26.90%	24.70%	N/A
2012	46.50%	2.60%	26.60%	24.40%	N/A
2013	46.10%	3.00%	26.80%	24.10%	N/A
2014	46.30%	3.80%	26.10%	23.90%	N/A
2015	45.80%	3.60%	26.50%	24.20%	N/A

Footnote: areas in transition are defined as those with an average posterior probability of type membership < 0.8.

Appendix 4.2: proportion of census sections that lost, gained or were stable in the number of food stores in each category by year.

Store Type		7/2012 to 7/2013	7/2013 to 7/2014	7/2014 to 7/2015	7/2015 to 7/2016
All Food Stores	Lose	3.90%	4.90%	6.60%	7.70%
	Stable	66.50%	67.30%	67.20%	70.30%
	Gain	29.60%	27.80%	26.20%	22.00%
Supermarkets	Lose	2.20%	2.00%	2.80%	2.80%
	Stable	83.40%	86.80%	85.30%	89.70%
	Gain	14.40%	11.30%	12.00%	7.60%
Specialized Stores	Lose	6.90%	5.60%	7.60%	5.90%
	Stable	80.60%	81.10%	81.10%	84.10%
	Gain	12.50%	13.30%	11.30%	10.10%
FV Stores	Lose	3.90%	2.70%	3.40%	2.30%
	Stable	91.20%	91.40%	91.70%	93.30%
	Gain	4.90%	5.90%	4.80%	4.40%

Appendix 4.3: Neighborhood Change (averaged from 2007 to 2011) and gain/loss in food stores (2012 to 2016), unadjusted results

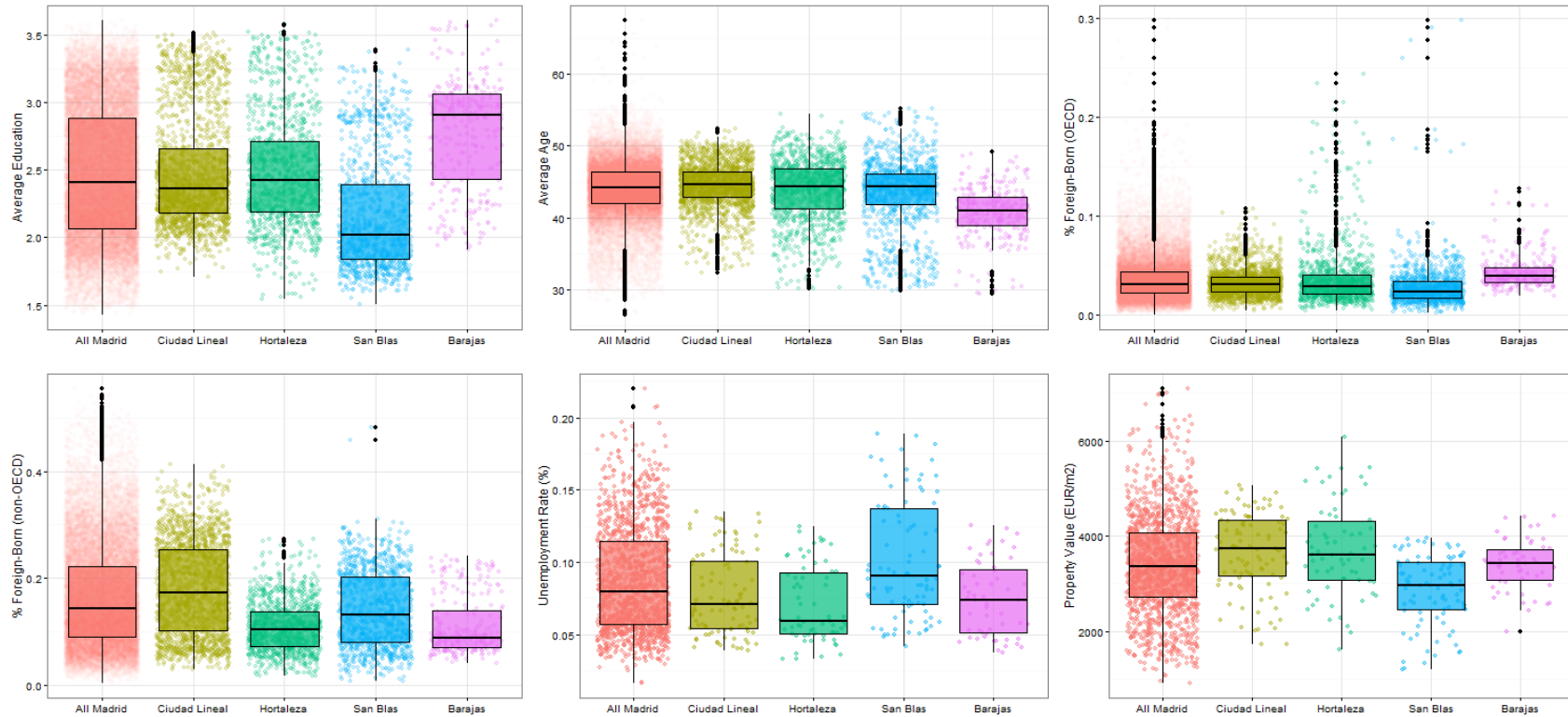
Store Loss	Type 1 (+Diversity)	Type 3 (+Prop. Value)	Type 4 (+Aging)	Areas in Transition
All Food Stores	0.93 (0.69;1.27)	1 (Ref.)	1.31 (0.81;2.11)	0.91 (0.69;1.20)
Supermarkets	0.71 (0.44;1.13)	1 (Ref.)	1.12 (0.55;2.29)	1.03 (0.70;1.53)
Specialized Stores	1.35 (0.99;1.85)	1 (Ref.)	1.46 (0.87;2.45)	1.19 (0.89;1.59)
FV Stores	1.33 (0.84;2.09)	1 (Ref.)	1.41 (0.68;2.92)	1.20 (0.79;1.83)
Store Gain	Type 1 (+Diversity)	Type 3 (+Prop. Value)	Type 4 (+Aging)	Areas in Transition
All Food Stores	0.97 (0.80;1.18)	1 (Ref.)	1.19 (0.87;1.62)	0.99 (0.83;1.18)
Supermarkets	1.33 (1.03;1.71)	1 (Ref.)	1.83 (1.24;2.69)	1.28 (1.01;1.61)
Specialized Stores	0.95 (0.74;1.22)	1 (Ref.)	1.24 (0.82;1.89)	0.90 (0.72;1.14)
FV Stores	1.03 (0.74;1.44)	1 (Ref.)	1.00 (0.55;1.82)	0.89 (0.65;1.21)

Appendix 4.4: Neighborhood Change (averaged from 2007 to 2011) and gain/loss in food stores (2012 to 2016), adjusted for baseline # of stores

Store Loss	Type 1 (+Diversity)	Type 3 (+Prop. Value)	Type 4 (+Aging)	Areas in Transition	Baseline # of Stores (+1)
All Food Stores	0.88 (0.65;1.2)	1 (Ref.)	1.26 (0.79;2.02)	0.87 (0.66;1.14)	1.05 (1.03;1.06)
Supermarkets	0.67 (0.42;1.06)	1 (Ref.)	1.19 (0.58;2.46)	1 (0.68;1.48)	3.08 (2.71;3.51)
Specialized Stores	1.29 (0.95;1.76)	1 (Ref.)	1.49 (0.91;2.43)	1.15 (0.86;1.53)	1.07 (1.05;1.09)
FV Stores	1.25 (0.8;1.95)	1 (Ref.)	1.51 (0.75;3.05)	1.13 (0.75;1.7)	1.3 (1.2;1.4)
Store Gain	Type 1 (+Diversity)	Type 3 (+Prop. Value)	Type 4 (+Aging)	Areas in Transition	Baseline # of Stores (+1)
All Food Stores	0.96 (0.8;1.15)	1 (Ref.)	1.16 (0.86;1.55)	0.97 (0.82;1.14)	1.07 (1.06;1.09)
Supermarkets	1.34 (1.04;1.72)	1 (Ref.)	1.81 (1.24;2.65)	1.28 (1.02;1.62)	1.64 (1.5;1.79)
Specialized Stores	0.95 (0.75;1.21)	1 (Ref.)	1.28 (0.87;1.88)	0.89 (0.71;1.1)	1.08 (1.06;1.09)
FV Stores	1.01 (0.73;1.4)	1 (Ref.)	1.02 (0.59;1.76)	0.84 (0.62;1.14)	1.25 (1.18;1.33)

Appendix C: Appendices for Chapter 5 (Aim 3)

Appendix 5.1: Comparison of the four study districts vs. the rest of Madrid in education, country of origin, age, unemployment and property value.



Appendix 5.2. Type Membership after the application of the classification algorithm (averaged from 2006-2009) vs actual modal type membership from 2006 to 2009

		Actual Modal Type Membership for every year 2006-2009				
		Type 1	Type 2	Type 3	Type 4	Total
Averaged Posteriors 2006-2009 + Algorithm	Type 1	513	0	1	22	536
	Type 2*	0	0	0	0	0
	Type 3	3	2	201	6	212
	Type 4	3	0	1	64	68
	Areas in Transition	221	17	102	236	576
	Total	740	19	305	328	1392

Footnote: * no areas were classified as Type 2 after applying the classification algorithm.

Appendix 5.3. Association (OR, 95% CI) of Neighborhood Social and Economic Change and Diabetes Prevalence

Variable	Model 1 (Unadjusted)	Model 2 (M1+ Age and Sex)	Model 3 (M2 + NBSES)
Type 1 (+Diverse)	1.18 (1.00;1.39)	1.31 (1.16;1.49)	0.94 (0.86;1.02)
Type 3 (+Prop. Value)	1 (Ref.)	1 (Ref.)	1 (Ref.)
Type 4 (Aging)	1.15 (0.89;1.50)	1.28 (1.08;1.53)	0.98 (0.87;1.11)
Areas in Transition	1.18 (1.02;1.36)	1.25 (1.11;1.41)	0.94 (0.86;1.03)
Female		1 (Ref.)	1 (Ref.)
Male		1.57 (1.52;1.63)	1.60 (1.54;1.66)
Age: 40-50		1 (Ref.)	1 (Ref.)
Age: 50-60		3.11 (2.89;3.34)	3.11 (2.91;3.33)
Age: 60-70		7.17 (6.68;7.69)	7.09 (6.67;7.53)
Age: 70-80		10.82 (10.05;11.66)	9.90 (9.31;10.53)
Age: 80+		8.90 (8.18;9.67)	8.16 (7.60;8.76)
NB SES (+1 SD)			0.65 (0.63;0.68)

Appendix 5.4. Association (HR, 95% CI) of Baseline Neighborhood Social and Economic Change and Diabetes Incidence (< 10 events excluded)

Variable	Model 1 (Unadjusted)	Model 2 (M1+ Age and Sex)	Model 3 (M2 + NBSES)
Type 1 (+Diverse)	1.08 (0.96;1.21)	1.15 (1.04;1.27)	0.92 (0.83;1.01)
Type 3 (+ Prop. Value)	1 (Ref.)	1 (Ref.)	1 (Ref.)
Type 4 (Aging)	0.93 (0.76;1.15)	1.00 (0.86;1.17)	0.84 (0.73;0.96)
Areas in Transition	1.05 (0.94;1.16)	1.09 (0.99;1.20)	0.90 (0.82;0.98)
Female		1 (Ref.)	1 (Ref.)
Male		1.76 (1.68;1.85)	1.77 (1.69;1.86)
Age: 40-50		1 (Ref.)	1 (Ref.)
Age: 50-60		2.43 (2.24;2.62)	2.41 (2.24;2.60)
Age: 60-70		3.64 (3.33;3.99)	3.61 (3.32;3.92)
Age: 70-80		4.48 (4.07;4.94)	4.24 (3.88;4.64)
Age: 80+		3.62 (3.26;4.03)	3.38 (3.06;3.74)
			0.75 (0.72;0.78)

Appendix 5.5. Association (HR, 95% CI) of Baseline Neighborhood Social and Economic Change and Diabetes Incidence excluding people that changed health centers

Variable	Model 1 (Unadjusted)	Model 2 (M1+ Age and Sex)	Model 3 (M2 + NBSES)
Type 1 (+Diverse)	0.95 (0.78;1.17)	1.05 (0.89;1.25)	0.79 (0.67;0.93)
Type 3 (+ Prop. Value)	1 (Ref.)	1 (Ref.)	1 (Ref.)
Type 4 (Aging)	0.89 (0.58;1.38)	0.99 (0.72;1.37)	0.80 (0.60;1.06)
Areas in Transition	0.99 (0.83;1.18)	1.06 (0.90;1.24)	0.82 (0.71;0.96)
Female		1 (Ref.)	1 (Ref.)
Male		1.72 (1.63;1.81)	1.73 (1.64;1.83)
Age: 40-50		1 (Ref.)	1 (Ref.)
Age: 50-60		2.52 (2.28;2.78)	2.49 (2.27;2.73)
Age: 60-70		3.72 (3.28;4.21)	3.65 (3.28;4.07)
Age: 70-80		4.34 (3.80;4.96)	4.04 (3.61;4.53)
Age: 80+		3.45 (2.98;4.00)	3.18 (2.80;3.61)
			0.73 (0.68;0.79)

March 13, 2017

CURRICULUM VITAE

USAMA BILAL, M.D., M.P.H.

PERSONAL DATA

Work Address

Department of Epidemiology, Rm W6604
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street, Baltimore, MD 21205

Email: ubilal@jhmi.edu

Personal Website: www.usamabilal.info

Birth Date and Location: January 31st, 1986; Gijon (Spain)

EDUCATION

- PhD/2017 Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
(Cardiovascular Epidemiology). Dissertation Title: “*Neighborhood Social and Economic Change, Food Environment Change and Diabetes Incidence in Madrid, Spain*”. Advisors: Dr. Thomas A. Glass (Dissertation) & David. D. Celentano (Academic)
- MPH/2012 University of Alcalá/National School of Public Health, Madrid, Spain
Thesis Title: “*Neighborhood Availability of Healthy Foods and Recreational Resources in Relation to Hypertension and Diabetes Control: the MESA study*”. Thesis Advisor: Dr. Manuel Franco
- MD/2011 University of Oviedo, Asturias, Spain. Thesis Title: “*Androgens block outward potassium currents and decrease spontaneous action potentials in GH3 cells*”. Thesis Advisor: Dr. Manuel Sanchez

OTHER TRAINING

- 2016 Santa Fe Institute Complex Systems Summer School (1 month).
- 2009-2011 Epidemiology Internship (6 months). National Center for Cardiovascular Research (CNIC), Madrid, Spain. *Supervisor:* Dr. Manuel Franco. *Topic:* Psychosocial Determinants of Cardiovascular Diseases.
- 2008-2011 Laboratory Internship (3 years). Medical School, University of Oviedo, Oviedo, Spain. *Supervisor:* Dr. Manuel Sanchez. *Topic:* Androgens and maxiK Channel Regulation.
- 2008 Laboratory Internship (1 month). Centre de Recherche de L’Hôpital Saint-Luc, Montreal, Canada. *Supervisor:* Dr. Suhayla Mukaddam-Daher. *Topic:* Imidazoline Receptors and Hypertension.

PROFESSIONAL EXPERIENCE

- 2014-2017 Research Assistant, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. *Supervisor:* Dr. Bryan Lau & Dr. Geetanjali Chander
- 2012-2013 Research Assistant, Social and Cardiovascular Epidemiology Research Group, University of Alcalá, Madrid, Spain. *Supervisor:* Dr. Manuel Franco
- 2011 Research Assistant, Department of Epidemiology, National Center for Cardiovascular Research, Madrid, Spain. *Supervisor:* Dr. Manuel Franco

PROFESSIONAL ACTIVITIES

Society Membership

- 2011-present Spanish Society of Epidemiology
- 2011-present Spanish Society of Public Health and Health Administration
- 2013-present American Heart Association, Epidemiology and Prevention Council
- 2013-present Society for Epidemiologic Research

EDITORIAL ACTIVITIES

Peer Review Activities. Journal Reviewer

- BMC Genomics
- PeerJ
- Preventive Medicine
- Tobacco Control
- Public Health Nutrition
- Epidemiology
- Gaceta Sanitaria
- International Journal of Health Geographics
- Circulation
- Journal of Epidemiology and Community Health

Review of Abstracts for Scientific Meetings

- 2013 Spanish Society of Epidemiology Annual Meeting
- 2014-2017 Society for Epidemiologic Research Annual Meeting

HONORS AND AWARDS

- 2010-2011 Ministry of Education Collaboration Scholarship, Spain
- 2009, 2010 *CICERONE* Undergraduate Research Scholarship, National Center for
2011 Cardiovascular Research, Madrid, Spain
- 2013-2015 *La Caixa* Postgraduate International Fellowship, Spain
- 2013-2015 *Enrique Najera* award for young epidemiologists, Spanish Society for
Epidemiology, Spain
- 2014 Society for Epidemiologic Research top 3 presentation at the 4th Annual
SERdigital Student Novel Methods Web Conference
- 2015 Charlotte Silverman Fund, Department of Epidemiology, Johns Hopkins
Bloomberg School of Public Health, Baltimore, MD
- 2015-2016, CLF-Lerner Fellow, Center for a Livable Future, Johns Hopkins Bloomberg
2016-2017 School of Public Health, Baltimore, MD
- 2016 Carol Buck Student Prize Paper Award (*Finalist*), Society for Epidemiologic
Research and 2016 Epidemiology Congress of the Americas, Miami, FL
- 2016 Dorothy and Arthur Samet Student Support Fund, Department of
Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore,
MD

PUBLICATIONS

Publications Under Review or in Preparation (available by request)

1. Gullon P, **Bilal U***, Cebrecos A, Badland HM, Galan I, Franco M. Socioeconomic Determinants of Small-Area Walkability in Madrid: The Heart Healthy Hoods Project. *Under Review*
2. **Bilal U**, Diez J, Diez-Roux AV, Franco M. Understanding Urban Health Determinants by Measuring Contextual Characteristics in Average Areas: the Median Neighborhood Index. *Under Review*
3. **Bilal U**, Knapp EA, Cooper RC. Swing Voting in the 2016 Presidential Election in Counties Where Midlife Mortality has been Rising in White Non-Hispanic Americans. *Under Review*
4. **Bilal U**, Lau B, Matthews CW, Mayer K, Geng E, Napravnik S, Eron JJ, Mugavero MJ, Hutton H, Cropsey K, McCaul ME, Crane H, Chander G. Measuring Alcohol Trajectories in

People Living with HIV: Modeling Finite Mixtures of Multinomial Distributions. *Under Review*

5. **Bilal U**, Franco M, Lau B, Celentano DD, Glass TA. Measuring neighborhood social and economic change for health studies. *In Preparation*.
6. **Bilal U**, Glass TA, Gittelsohn J, Jones-Smith J, Lawrence RS, Franco M. Neighborhood Social and Economic Change and Retail Food Environment Changes. *In Preparation*.
7. **Bilal U**, Coresh J, Jones-Smith J, del Cura Gonzalez I, Sanchez-Perruca L, Glass TA, Franco M. Neighborhood Social and Economic Change and Diabetes Incidence in Madrid, Spain. *In Preparation*
8. **Bilal U**, del Cura I, Sanchez-Perruca L, Hill-Briggs F, Franco M. Association of Neighborhood Socioeconomic Status and Diabetes Burden using Electronic Health Records in the Entire Population of Northeastern Madrid (Spain). *In Preparation*

Peer Reviewed Journal Articles (denotes co-first authorship)*

1. **Bilal U**, Cooper RS, Abreu FM, Nau C, Franco M, Glass TA. Economic growth and mortality: do social protection policies matter?. *International Journal of Epidemiology* 2017. *In Press*.
2. Fuentes S, **Bilal U**, Galan I, Villalbi JR, Espelt A, Bosque-Prous M, Franco M, Lazo M. Binge drinking and well-being in European Older Adults: Do gender and region matter? *European Journal of Public Health* 2017. *In Press*.
3. Cebrecos A, Diez J, Gullon P, **Bilal U**, Franco M, Escobar F. Characterizing physical activity and food urban environments: a GIS-based multicomponent proposal. *International Journal of Health Geographics*. 2016;15:35
4. **Bilal U**, Diez J, Alfayate S, Gullon P, del Cura I, Escobar F, Sandin M, Franco M. Population Cardiovascular Health and Urban Environments: The Heart Healthy Hoods exploratory study in Madrid, Spain. *BMC Medical Research Methodology*. 2016;16:104
5. Diez J, **Bilal U**, Cebrecos A, Buczynski A, Lawrence RS, Glass TA, Escobar F, Gittelsohn J, Franco M. Understanding differences in the local food environment across countries: A case study in Madrid (Spain) and Baltimore (USA). *Preventive Medicine*. 2016 Aug; 89: 237-244
6. Carrasco M, **Bilal U**. A Sign of the Times: To Have or To Be? Social Capital or Social Cohesion? *Social Science and Medicine*. 2016 Jun 159: 127-131
7. **Bilal U**, Beltran P, Navas-Acien A, Fernandez E, Bolumar F, Franco M. Gender Equality and Smoking: a theory-driven approach to smoking gender differences in Spain. *Tobacco Control* 2016 May;25(3):295-300

8. **Bilal U**, Lau BM, Lazo M, McCaul ME, Hutton HE, Sulkowski MS, Moore RD, Chander GD. Interaction Between Alcohol Consumption Patterns, Antiretroviral Therapy Type and Liver Fibrosis in Persons Living with HIV. *AIDS Patient Care and STDs* 2016 May;30(5):200-7.
9. Glass TA, **Bilal U**. Are neighborhoods causal? Complications arising from the ‘stickiness’ of ZNA. *Social Science and Medicine*. 2016 Oct;166:244-53.
10. Lazo M, **Bilal U***, Perez-Escamilla R. Epidemiology of NAFLD and Type 2 Diabetes: Health Disparities Among Persons of Hispanic Origin. *Current Diabetes Reports*. 2015 Dec;15(12):116
11. Cobb LK, Anderson CA, Appel LJ, Jones-Smith J, **Bilal U**, Gittelsohn J, Franco M. Changes in healthy food availability in Baltimore City food stores, 2006-2012 and the impact of the 2009 WIC policy change. *Health Affairs*. 2015 Nov 1;34(11):1849-5
12. Prasad V, **Bilal U**. The Role of Censoring on Progression Free Survival: Oncologist Discretion Advised. *European Journal of Cancer* 2015. 2015 Nov;51(16):2269-71
13. Gullon P, Badland HM, 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. *Journal of Urban Health*. 2015 Oct;92(5):923-39
14. Franco M, **Bilal U**, Diez-Roux AV. Preventing non-communicable diseases through structural changes in urban environments. *Journal of Epidemiology and Community Health*. 2015 Jun;69(6):509-11
15. Suarez L, **Bilal U***, Bordallo J, Cantabrana B, Sanchez M. Androgens block outward potassium currents and decrease spontaneous action potentials in GH3 cells. *Naunyn Schmiedebergs Arch Pharmacol*. 2015 Jan;388(1):67-78.
16. Redondo O, Villamor E, Valdes J, **Bilal U**, Caballero B, Roche D, Kroker F, Ramirez M, Franco M. Validation of a BMI cutoff point to predict an adverse cardiometabolic profile with DEXA adiposity measurements in Guatemalan children. *Public Health Nutrition* 2014 Jun 23:1-8.
17. **Bilal U**, Fernandez E, Beltran P, Navas-Acien A, Bolumar F, Franco M. Validation of a Method to Reconstruct Historical Smoking Prevalence Rates. *American Journal of Epidemiology* 2014;179 (1): 15-19.
18. Franco M, **Bilal U**, Orduñez P, Benet M, Morejon A, Caballero B, Kennelly JF, Cooper RS. Population-wide body weight loss and regain in relation to diabetes and cardiovascular disease mortality: Cuba 1980-2010. *British Medical Journal*. *BMJ* 2013;346:f1515.
19. Franco M, **Bilal U**, Guallar E, Sanz G, Gomez A, Fuster V, Cooper R. Systematic Review of Three Decades of Spanish Cardiovascular Epidemiology: Improving Translation for a Future of Prevention. *European Journal of Preventive Cardiology*. 2013 Ago;60(4):565-576.

20. Franco M, Cooper R, **Bilal U**, Fuster V. Challenges and opportunities for cardiovascular disease prevention. *American Journal of Medicine*. 2011 Feb;124(2):95-102.

Commentaries, Editorials and Letters

1. Carrasco M, **Bilal U**. Are subversion and conflict component parts of social cohesion? - A response to Lindstrom. *Social Science and Medicine* 2016 Sep 10;169:31-32
2. **Bilal U**, Padilla-Bernaldez J. [Cancer and Meat Consumption: Beyond Individual Lifestyles]. *Gaceta Sanitaria*. 2016 Jul-Aug;30(4):318.
3. **Bilal U**, Fernandez E, Navas-Acien A, Bolumar F, Franco M. Five Authors Reply. *American Journal of Epidemiology* 2014 Sep 15;180(6):659
4. Gullón P, **Bilal U**, Franco M. Physical activity environment measurement and same source bias. *Gaceta Sanitaria*. 2014 Jul-Aug;28(4):344-5
5. Padilla J, Carmona M, **Bilal U**. Trivializing conflict of interest, a crime against public health. *British Medical Journal (Rapid-Response)*, 12 November 2013)
6. **Bilal U**, Padilla J. Spanish government expels undocumented migrants from the health care system. *European Journal of Public Health (E-Letter)*, 8 June 2012)
7. Franco M, Cooper R, **Bilal U**, Fuster V. Control of Coronary Heart Disease Risk Factors and Evidence-Based Therapies: Joint Efforts for Coronary Heart Disease Prevention in Spain. *Revista Española Cardiología*. 2011 Nov;64(11):962-964

Book Chapters

1. Franco M, **Bilal U**, Diez J. Local Food Environment. In: Caballero, B., Finglas, P., and Toldrá, F. (eds.) *The Encyclopedia of Food and Health* vol. 4, pp. 139-144 (2016). Oxford: Academic Press.
2. Lazo M, **Bilal U**. Obesity: Epidemiology of. In: Caballero, B., Finglas, P., and Toldrá, F. (eds.) *The Encyclopedia of Food and Health* vol. 4, pp. 139-144 (2016). Oxford: Academic Press.
3. **Bilal U**. Urbanismo o Barbarie [Urbanism or Barbarism]. In: Padilla J, ed. *Salubrisimo o Barbarie [Public Health or Barbarism]*. Madrid: Atrapasueños Editorial; 2016. *Forthcoming*.
4. Franco M, **Bilal U**. [Epidemiology of Hypertension and Angina Pectoris in Spain]. *Formación Continuada en Atención Primaria*; 2012.
5. **Bilal U**, Belza MJ, Bolumar F. [Introduction to Epidemiology and its use in Health Administration]. In: [Master in Health Administration. Module 3: Public Health and Clinical Epidemiology]. Madrid: National University of Distance Education; 2012.

Op-eds and other pieces in the general media

1. Franco M, **Bilal U**, Cooper RC. [Baltimore, a showcase of inequality]. El PAIS. URL: http://elpais.com/elpais/2015/05/04/ciencia/1430735350_821550.html
2. Padilla J, **Bilal U**. Padilla J, Bilal U. [Should we restrict healthcare coverage to people with unhealthy behaviors?]. El Diario. URL: http://www.eldiario.es/tribunaabierta/Restringir-sanidad-publica-cuida_6_556154408.html

CURRICULUM VITAE

USAMA BILAL, M.D., M.P.H.

PART II

TEACHING

Classroom Instruction (Instructor or Course Director)

University of Alcala, Madrid, Spain

2013 Basic Epidemiology (1 semester), co-Instructor and lab instructor, 80 undergraduate Human Biology students.

National School of Public Health, Madrid, Spain

2013 Social Epidemiology (1 week), co-Instructor, 20 MPH Students.

2016 Social Epidemiology (1 week), co-Instructor, 25 MPH Students.

Other Significant Teaching (Guest Lecturer, Teaching Assistant, Laboratory Instructor)

University of Alcala, Madrid, Spain

2011, 2012 Basic Epidemiology, guest lecturer (2 lectures) and lab instructor, 75 undergraduate Human Biology students.

University of Oviedo, Asturias, Spain

2012, 2015 Introduction to Cardiovascular Epidemiology, guest lecturer (1 lecture), 25 MSc in Clinical Research students.

Johns Hopkins Bloomberg School of Public Health

2014-2016 Social Epidemiology (340.628), Teaching Assistant, 10-15 graduate students. Supervisors: Dr. Manuel Franco and Dr. Thomas A. Glass.

2015 Foundations of Social Epidemiology (340.666), Teaching Assistant, 40 graduate students. Supervisors: Dr. David Celentano and Dr. Amanda Latimore.

2015 Methodologic Challenges in Epidemiologic Research (340.754), Teaching Assistant, 80 graduate students. Supervisors: Dr. Bryan Lau and Dr. Allison Abraham.

2015 Methods for Clinical and Translational Research Workshop, Teaching Assistant, 20 graduate students. Supervisor: Dr. Jon Samet.

2014-2016 Tutor for 15 students in Principles of Epidemiology, Epidemiologic Methods I, II and III, Biostatistics 140.621, 622 and 623.

2015 Advanced Methods for the Design and Analysis of Cohort Studies (340.728), *Lab Instructor and Teaching Assistant*, 50 graduate students. Supervisors: Dr. Alvaro Muñoz and Dr. Christopher Cox.

2016 Advanced Seminar in Social Epidemiology (340.705), Teaching Assistant, 9 graduate students. Supervisor: Dr. Thomas A. Glass.

2016 Political Economy of Social Inequalities (308.610), Teaching Assistant, 25 graduate students. Supervisor: Dr. Vicente Navarro.

ACADEMIC SERVICE

School of Medicine, University of Oviedo, Asturias, Spain

2004-2007 Asturian Medical Students Association, webpage administrator
2007-2009 Spanish Medical Students Federation (IFMSA-Spain), webpage administrator
2008-2009 Asturian Medical Students Association, President
2009-2010 Spanish Medical Students Federation (IFMSA-Spain), President
2010 Head of Delegation at the 59th General Assembly of the International Federation of Medical Students Associations (IFMSA), Bangkok
2010 Head of Delegation at the 60th General Assembly of the International Federation of Medical Students Associations (IFMSA), Montreal
2011 Committee for the adaptation of the Doctor of Medicine Degree to the European Higher Education Area, student representative

School of Medicine, Universidad de Alcalá, Madrid, Spain

2012-2013 Social and Cardiovascular Epidemiology Research Group, Journal Club Coordinator

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

2013-2017 Social Epidemiology Student Organization, Director
2014-2015 Departmental Curriculum Committee, Student Representative
2015-2016 Departmental Faculty Committee, Student Representative

Welch Center for Prevention, Epidemiology and Clinical Research, JHMI

2014-2015 Coordinator for the Research Pearl at the Welch Center Grand Rounds of Clinical Research, Johns Hopkins Medical Institutions
2015-2016 Website Design Committee

Johns Hopkins Center for a Livable Future, Johns Hopkins Bloomberg School of Public Health

2015-2016 CLF-Lerner Fellows Journal Club, Co-Coordinator
2016-2017 CLF-Lerner Fellows Enrichment Activities, Co-Coordinator

RESEARCH GRANT PARTICIPATION

Active

Fondo de Investigaciones Sanitarias (Spain)

01/01/2016-12/31/2018

Sureda, P.I. (Bilal, Co-investigator)

Availability of Tobacco Products and Smoke free Policy Implementation in Madrid

The purpose of this study is to explore differential availability of tobacco products and differential implementation and compliance with smoke free policies across neighborhoods in Madrid (Spain).

PRESENTATIONS

Invited Seminars/Presentations

- 2012 Framework to study the Social Determinants of Health. Summer School of Public Health for Medical Students, Andalusian School of Public Health, Granada, Spain
- 2012 Sick Individuals and Sick Populations. School of Medicine, Complutense University, Madrid, Spain
- 2015 Macrosocial Changes and Chronic Diseases: Gender and Tobacco in Spain, 1960-2010. Seminar Series, National Center for Epidemiology (CNE), Madrid, Spain.
- 2016 The Health Effects of Economic Change: National Policies and Local Mechanisms. Public Health Forum. Institute of Psychology Health and Society, University of Liverpool, UK.

Presentations at Scientific Meetings (Oral Presentations)

- 2009 Estrogens and Cardiovascular Primary Prevention (**Oral Presentation**). 15th National Congress of Cardiology for Medical Students, Salamanca, Spain
- 2014 Fixed-effects versus Random-effects in the analysis of time-fixed confounding of ecological time series: a simulation study (**Live Presentation, award winning**). 4th Annual SERdigital Student Novel Methods Web Conference (Online)
- 2016 Population Cardiovascular Health and Urban Environments: The Heart Healthy Hoods exploratory study in Madrid, Spain (**Oral Presentation**); Healthy urban environment characterization focused on physical activity and food: A GIS-based method (**Oral Presentation**, on behalf of A. Cebrecos); Does walkability differ by area sociodemographic profile? A study of Madrid City (**Oral Presentation**, on behalf of P. Gullon). 13th International Conference on Urban Health, San Francisco, CA
- 2016 [Measuring Social and Economic Change in Small Areas of Madrid, Spain: Innovative Indicators and Alternative Data Sources] (**Oral Presentation**); Economic Growth and Mortality: Do Social Protection Policies Matter? (**Oral Presentation**). Annual Meeting of the Spanish Society of Epidemiology, Seville, Spain

Presentations at Scientific Meetings (Posters, presented on site)

- 2013 Gender Equality and female-to-male Smoking Ratio: Trends in Spain in the Last 50 Years (**Poster**). American Heart Association Epidemiology and Prevention 2013 Scientific Sessions, New Orleans, LA
- 2013 Validation of a Method to Reconstruct Historical Smoking Prevalence Rates (**Recorded Presentation**). 3rd Annual SERdigital Student Novel Methods Web Conference (Online).
- 2014 Food Stores, Food Markets and Healthy Food Availability in Comparable Urban Neighborhoods in Madrid and Baltimore (**Poster**), Measuring the Food, Tobacco, Alcohol and Physical Activity Urban Environments in Relation to Cardiovascular Health: The Heart Healthy Hoods Pilot Study in Madrid, Spain (**Poster**), Economic Crisis in Western Europe and Ischemic Heart Disease Mortality (**Poster**). American Heart Association Epidemiology and Prevention 2014 Scientific Sessions, San Francisco, CA
- 2014 Economic Crises and Ischemic Heart Disease Mortality in Europe: Effect Heterogeneity? (**Poster**), Selecting Comparable Neighborhoods across Cities: The Median Neighborhood Index (**Poster**). Society for Epidemiologic Research Annual Meeting, Seattle, WA
- 2015 Macroeconomic Growth is Associated with Increases in Cardiovascular Mortality in Countries with Lower Social Protection Spending (**Moderated Poster**). American Heart Association Epidemiology and Prevention 2015 Scientific Sessions, Baltimore, MD
- 2015 Fixed versus Random effects models for longitudinal data analysis of confounding in ecological time series: a simulation study (**Poster**). Society for Epidemiologic Research Annual Meeting, Denver, CO