BMJ Open

Association of Neighborhood Socioeconomic Status and Diabetes Burden using Electronic Health Records in Madrid (Spain): The Heart Healthy Hoods Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-021143
Article Type:	Research
Date Submitted by the Author:	13-Dec-2017

Complete List of Authors:	Bilal, Usama; Drexel Dornsife School of Public Health, Urban Health Collaborative; Universidad de Alcala de Henares Facultad de Medicina y Ciencias de la Salud, Social and Cardiovascular Research Group Hill-Briggs, F; Johns Hopkins University School of Medicine, Department of Medicine; Johns Hopkins Medical Institutions, Welch Center for Prevention, Epidemiology, and Clinical Research Sanchez-Perruca, Luis; Madrid Health Service, Information Services Directorate; Instituto de Salud Carlos III, Health Services Research on Chronic Patients Network (REDISSEC) del Cura, Isabel; Madrid Health Service, Primary Care Research Unit; Instituto de Salud Carlos III, Health Services Research on Chronic Patients Network (REDISSEC) Franco, M; Universidad de Alcala de Henares Facultad de Medicina y Ciencias de la Salud, Social and Cardiovascular Research Group

Keywords:	social epidemiology, social inequalities, Spain, record linkage, Diabetes, Neighborhoods

Association of Neighborhood Socioeconomic Status and Diabetes Burden using Electronic Health Records in Madrid (Spain): The Heart Healthy Hoods Study

Usama Bilal[1,2,3*], Felicia Hill-Briggs[1,4,5],
Luis Sánchez-Perruca[6,7], Isabel del Cura[7,8,9], and Manuel Franco[1,2]

1. Department of Epidemiology. Johns Hopkins Bloomberg School of Public Health. 615 N Wolfe St. Baltimore, MD, 21205 USA.
2. Social and Cardiovascular Epidemiology Research Group. Universidad de Alcalá. Ctra Madrid Barcelona km 33.6, Alcala de Henares, Madrid, 28871 Spain.
3. Urban Health Collaborative, Drexel Dornsife School of Public Health, 3600 Market St, Philadelphia, PA, 19104, USA
4. Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins Medical Institutions, 2024 E Monument St, Baltimore, MD 21205 USA
5. The Johns Hopkins University School of Medicine, Baltimore, MD
6. Information Services Directorate. Primary Care Management. Madrid Health Service, Madrid, Spain.
7. Health Services Research on Chronic Patients Network (REDISSEC). ISCIII, Spain.
8. Primary Care Research Unit. Primary Care Management. Madrid Health Service, Spain.
9. Area of Preventive Medicine and Public Health. Rey Juan Carlos University, Madrid, Spain.

Words: 2873; Abstract: 246; Figures: 2; Tables: 2; References: 29

Corresponding Author:
Usama Bilal, MD MPH PhD
Urban Health Collaborative
Drexel Dornsife School of Public Health
3600 Market St, 7th Floor, Philadelphia, PA, 19104
Telephone: +1 410 419 6170
Email: ubilal@drexel.edu

Abstract

Objective: To study the association between neighborhood socioeconomic status and diabetes prevalence, incidence, and control in the entire population of Northeastern Madrid, Spain.

Setting: Primary-care system in four districts of Madrid (Spain)

Participants: 269,942 people aged 40 or above, followed from 2013 to 2014

Exposure: Neighborhood Socioeconomic Status, measured using a composite index of 7 indicators of education, wealth, occupation, and living conditions.

Primary Outcome Measures: Diagnosis of diabetes based on ICPC-2 codes and glycated hemoglobin (HbA1c %)

Results: In regression analyses adjusted by age and sex and compared to individuals living in low SES neighborhoods, men living in medium and high SES neighborhoods had 10% (95% CI: 6 to 15%) and 29% (25 to 32%) decreased odds of diabetes, while women had 27% (23 to 30%) and 50% (47 to 52%) decreased odds. Moreover, men in medium and high SES neighborhoods had 13% (1 to 23%) and 20% (9 to 29%) decreased hazard of diabetes incidence while women had a decrease of 17% (3 to 29%) and 31% (20 to 41%). Individuals living in medium and high SES neighborhoods had 8% (2 to 15%) and 15% (9 to 21%) decreased odds of lack of diabetes control, and a decrease in average HbA1c % of 0.05 (0.01 to 0.10) and 0.11 (0.06 to 0.15).

Conclusions: Diabetes prevalence, incidence, and control were socially patterned by contextual SES in a Southern European city. Future studies should provide mechanistic insights and targets for intervention to address this health inequity.

Keywords: social epidemiology; social inequalities; record linkage; diabetes; Spain; neighborhood/place

Strengths and Limitations

- We study the entire population of an area of a very large city (Madrid) where almost 600,000 people live, resulting in a very large sample size and decreased concerns for selection bias as compared to regular cohort studies or surveys.
- The diagnosis of diabetes in our EHR has been validated before and shown to have a very high validity with a kappa of 0.99, but we cannot achieve the level of standardization of measurements of cohort studies.
- We use HbA1c which is a robust measure of diabetes control and is the standard of care in clinical practice.
- We used an exposure constructed from publicly available indicators, increasing the replicability of our findings and the applicability to other health outcomes, but restricting our capacity to build a complex exposure that may capture socioeconomic status better.
- 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 (especially individual level socioeconomic status).

Introduction

The burden of diabetes has seen a large increase in Western countries in recent decades[1]. 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[2]. Population preventive strategies are needed to decrease this burden[3], taking into consideration mass-influences that differ across populations[3].

Among these mass influences are neighborhood characteristics. A large body of literature has explored contextual socioeconomic influences on health. In particular, the association between neighborhood socioeconomic status and several measures of diabetes (prevalence, incidence or control) is robust and has been replicated in the US[4], other Anglo-Saxon countries[5 6], and Northern Europe[7 8] including in experimental or quasi-experimental settings[8 9]. Nonetheless, these influences have received scant attention in Southern Europe[10]. Moreover, previous studies have shown a strong social gradient in diabetes mortality in Spain, which warrants further mechanistic insights into its causes[11].

Finally, many of the studies outlined above use data from research-driven cohort studies. While these types of 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[12]. 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[12]. With Electronic Health Records (EHR) in a health system with universal health coverage these drawbacks may be overcome by avoiding the necessity for sampling altogether.

Taking the above into consideration, we studied the association between neighborhood socioeconomic status and diabetes prevalence, incidence and control in an electronic health record-based cohort of the entire population of Northeastern Madrid that includes data on more than 640,000 people.

Methods

Study setting

This study was conducted within the HeartHealthyHoods project (www.hhhproject.eu) in the city of Madrid, Spain[13]. We took data for 2013 and 2014 from all Health Care centers in four districts of the city of Madrid. These four districts house around 20% of the population of the entire city and include areas in the entire spectrum of socioeconomic status. Our unit of analysis is the census section (n=427), which is the smallest area for which the census collects data and has around 1200 people (range: 583 to 3865). Individual-level data was obtained from Electronic Health Records (EHR) including 640,217 individuals registered in any Health Center of the area. These EHR contain data on patient age, sex, residential location, clinical diagnoses, and laboratory values (lipids and HbA1c).

Given the low prevalence of type 2 diabetes in young individuals[14], and the change in medical practice for individuals above 40 in Madrid (that include cardiovascular prevention measures with risk factor assessments), we restricted our analysis to people aged 40 or above. Our final study sample was composed of 270,660 individuals, of which 23,908 had a diagnosis of diabetes. Primary Care EHR include 99.5% of the individuals living in the area per the Census[13].

Neighborhood Socioeconomic Status

The main exposure of this study was a Neighborhood Socioeconomic Status (NSES). We measured NSES using a composite index with 7 indicators in the 4 domains of the Spanish Commission to Reduce Health Inequalities[15]: education, wealth, occupation and living conditions. The 7 indicators were: in education, (1) Low levels of education (% people above 25 years of age with primary studies or below), (2) High levels of education (% people above 25 years of age with university education or above); in wealth, (3) Average housing prices (per sq. m); in occupation, (4) Part-time employment (% workers in part-time jobs), (5) Temporary employment (% workers in temporary jobs), (6) Manual occupational class (% workers in manual or unqualified jobs); and in living conditions (7) Unemployment rate (% registered unemployed individuals / people aged 16 to 64). Indicator data were obtained from the Padrón (a continuous and universal census collected for administrative purposes), the social security and employment services registries and the IDEALISTA report (a report from a large real estate corporation in Spain). All data were available by January 2013. The Online Resource contains more details on the operationalization of indicators. To compute the index, we standardized the seven indicators by centering by the mean and scaling by the standard deviation. We then calculated the composite index by taking the weighted mean of the three standardized indicators, so that each domain was weighted equally. We used this index in two ways: first, as a categorical variable in tertiles; and second, as a continuous variable if tertile analysis showed linearity.

Diabetes Prevalence, Incidence and Control

A type-2 diabetes diagnosis 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%)[16]. Prevalent cases were defined as those diagnosed before January 1st 2013. Incident cases were those occurring from January 1st 2013 to December 31st 2014 in people free of diabetes by baseline. We operationalized lack of diabetes control as either a dichotomous variable (HbA1c >=7%) or a continuous variable. If more than one value of HbA1c was available, we used the last available measurement of the year.

Statistical Methods

The overall goal of this analysis is to study the association between neighborhood socioeconomic status and diabetes prevalence, incidence and control. We computed descriptive statistics by tertile of neighborhood socioeconomic status.

To study the association between neighborhood socioeconomic status and diabetes prevalence or lack of control (binary indicator) we used a logistic regression model with robust standard errors clustered at the census section level using a sandwich Huber-White estimator. These models were adjusted for age (in five categories; 40 to 50, 50 to 60, 60 to 70, 70 to 80 and 80 and above) and sex. Continuous HbA1c (for diabetes control) was examined using a linear regression with robust standard errors clustered at the census section level using a sandwich Huber-White estimator. Diabetes incidence was examined by the use of Kaplan-Meier survival estimates, where patients were censored on death, moving out of the study area, diabetes incidence, or administrative censoring (December 31st 2014), whichever happened first. Cox Proportional Hazards models were used to estimate the adjusted association, with clustered standard errors on the census section. To graphically display the association between the exposure and the outcome variables we also modeled the association using restricted cubic splines with 4 knots in the percentiles recommended by Harrell[17]. A previous report in the Spanish setting highlighted a significant interaction by gender of contextual socioeconomic status and diabetes[10], so we explored whether this interaction existed in our analysis and displayed stratified results if this was the case.

All analyses were conducted in R v3.3.0 (R Software Foundation). This study was approved by the Madrid Primary Care Research Committee.

Results

Study Population

Table 1 shows a description of the study population by textile of neighborhood socioeconomic status (NSES) and in the total population. The total sample size was 269,942 people, with around 25%, 30% and 45% of the population living in low, medium and high NSES areas. Overall, the average age was 56.5 (IQR=47.4 to 69.8) and 54.9% of the population were women. 8.8% of the population over 40 years of age had diabetes and the average HbA1c in diabetic people was 6.7 (IQR=6.2 to 7.5). Thirty-nine percent of all diabetic people had uncontrolled diabetes (HbA1c equal or above 7%). In this population over 40 years of age the prevalence of other risk factors was 33% for hypertension, 6% for any cardiovascular disease, 27% for dyslipidemia, 2% for chronic kidney disease and 0.4% for retinopathy. Stratifying the population by textile of NSES revealed that younger people lived in neighborhoods with higher SES. The prevalence of diabetes decreased sharply with neighborhood SES (11.9% in the lowest NSES, 9.6% in the medium NSES and 6.5% in the highest NSES). The median HbA1c did not change, but the proportion of uncontrolled diabetic people was lowest in neighborhoods with the highest SES (37.1%, as compared to 40.5% and 38.7% in low and medium NSES, respectively). The prevalence of other risk factors also varied by neighborhood SES following similar patterns: as NSES increases the prevalence of cardiovascular risk factors, cardiovascular disease and diabetes complications decreased.

Table 1: Study Population by January 1st 2013.

Variable	Tertile 1 (Lowest NSES)	Tertile 2 (Mid NSES)	Tertile 3 (High NSES)	Total
Sample Size (N)	68369	81072	120501	269942
Median Age [IQR]	58.6 [48.3;74.5]	58.1 [48.0;71.1]	54.7 [46.6;66.9]	56.5 [47.4;69.8]
% Men	44.60%	44.20%	45.90%	45.10%
% Women	55.40%	55.80%	54.10%	54.90%
% with Diabetes	11.90%	9.60%	6.50%	8.80%
Median HbA1c [IQR]	6.7 [6.2;7.5]	6.7 [6.2;7.5]	6.7 [6.2;7.4]	6.7 [6.2;7.5]
HbA1c >=7%	40.50%	38.70%	37.10%	38.80%
HbA1c < 5%	0.30%	0.40%	0.30%	0.30%
HbA1c 5-6.5%	40.00%	40.50%	42.70%	41.10%
HbA1c 6.5-7%	19.40%	20.60%	20.30%	20.10%
HbA1c 7-9%	34.00%	32.20%	30.90%	32.40%
HbA1c >9%	6.30%	6.30%	5.70%	6.10%
% with Hypertension	32.80%	29.00%	22.40%	27.00%
% with Any CVD	7.50%	6.70%	5.10%	6.20%
% with CKD	2.60%	2.20%	1.40%	2.00%
% with Retinopathy	0.60%	0.40%	0.30%	0.40%
Low Education, % [IQR]	34.6% [29.3;39.4]	24.1% [19.8;27.4]	10.4% [5.9;17.1]	22.7% [10.6;29.1]
High Education, % [IQR]	10.4% [7.9;13.6]	21.1% [17.2;26.5]	41.7% [32.7;55.8]	23.9% [15.3;40.7]
Unemployment Rate, % [IQR]	13.8% [12.9;16.4]	12.6% [12.3;12.9]	8.9% [7.8;10.6]	11.9% [10.1;13.8]
Part-Time Workers, % [IQR]	26.7% [25.9;26.8]	23.4% [23.4;25.9]	16.5% [12.5;19.4]	22.1% [16.5;25.9]
Temporary Workers, % [IQR]	20.9% [20.4;21.5]	20.4% [19.0;20.9]	15.0% [13.8;17.6]	18.8% [16.0;20.9]
Manual Class, % [IQR]	40.3% [40.0;43.1]	37.1% [37.1;40.0]	22.4% [17.4;30.2]	35.9% [27.1;40.0]
Property Value, EUR/m^2 [IQR]	1825.0 [1576.0;1976.0]	2244.0 [2129.0;2403.0]	2874.0 [2614.0;3360.0]	2407.0 [2031.0;2822.0]

*Average education is the weighted average of the four education levels (no studies, primary, secondary and university education; each has a score of 1, 2, 3 and 4).

NSES and Diabetes Prevalence

Table 2 shows the association between NSES and diabetes prevalence, control, and incidence. Diabetes prevalence was associated in a dose-response manner to neighborhood SES. This association was significantly stronger in women as compared to men (p for interaction <0.001). In particular, compared to men living in low NSES neighborhoods, those living in medium NSES neighborhoods had a 10% decrease in the odds of having diabetes (OR=0.90, 95% CI 0.85 to 0.94), while those living in the highest NSES neighborhoods had a 29% decrease in the odds of diabetes (OR=0.71, 95% CI 0.68 to 0.75). In the case of women, those living in medium and high NSES neighborhoods had 27% and 50% decreased odds of diabetes, respectively, as compared to those living low NSES neighborhoods (OR=0.73, 95% CI 0.70 to 0.77, and OR=0.50, 95% CI 0.48 to 0.53). These associations were consistent in models looking at continuous NSES: a 1 standard-deviation increase in NSES was associated with a 18% and 30% decrease in the odds of diabetes in men and women, respectively (OR=0.82, 95% CI 0.81 to 0.84, OR=0.70, 95% CI 0.69 to 0.72). Figure 1 (Left Panel) shows the association using continuous NSES with restricted cubic splines, where the steeper pattern for women is evident.

NSES and Diabetes Control

Table 2 also shows the association between NSES and diabetes control, operationalized as a dichotomous variable (lack of diabetes control, or HbA1c >=7%) or a continuous variable (HbA1c %). There was no significant interaction by sex in the NSES and diabetes control (p for interaction=0.219 and 0.358 in the dichotomous and continuous model). As compared to diabetic people living in the lowest NSES neighborhoods, those living in medium NSES areas had 8% lower odds of lack of controlled diabetes (OR=0.92, 95% CI 0.85 to 0.98), while those living in the highest NSES areas had 15% lower odds of lack of diabetes control (OR=0.85, 95% CI 0.79 to 0.91). Moreover, a 1 standard-deviation increase in NSES was associated with a 6% decrease in the odds of lack of diabetes control (OR=0.94, 95% CI 0.91 to 0.97). These associations were maintained when looking at continuous HbA1c: diabetic people living in medium and high SES neighborhoods had a lower average HbA1c % (see Table 2). Figure 1 shows the prevalence of lack of controlled diabetes and average HbA1c levels across levels of neighborhood SES using restricted cubic splines, showing a linear decrease both in lack of control and in average HbA1c % with increasing NSES.

Table 2: Association of Neighborhood Socioeconomic Status and Diabetes Outcomes

	Diabetes Prevalence					
	Total		Men		Women	
Variable	**OR (95% CI)**	**p-val**	**OR (95% CI)**	**p-val**	**OR (95% CI)**	**p-val**
Tertile 1 of NSES (Low)	1 (Ref.)		1 (Ref.)		1 (Ref.)	
Tertile 2 of NSES (Middle)	0.81(0.78;0.83)	<0.001	0.90(0.85;0.94)	<0.001	0.73(0.70;0.77)	<0.001
Tertile 3 of NSES (High)	0.60(0.58;0.62)	<0.001	0.71(0.68;0.75)	<0.001	0.50(0.48;0.53)	<0.001
Continuous NSES	0.76(0.75;0.78)	<0.001	0.82(0.81;0.84)	<0.001	0.70(0.69;0.72)	<0.001
	Lack of Diabetes Control (HbA1c >= 7%)					
Variable	**OR (95% CI)**	**p-val**	**OR (95% CI)**	**p-val**	**OR (95% CI)**	**p-val**
Tertile 1 of NSES (Low)	1 (Ref.)		1 (Ref.)		1 (Ref.)	
Tertile 2 of NSES (Middle)	0.92(0.85;0.98)	0.017	0.90(0.82;1.00)	0.049	0.93(0.84;1.03)	0.14
Tertile 3 of NSES (High)	0.85(0.79;0.91)	<0.001	0.80(0.72;0.88)	<0.001	0.91(0.82;1.02)	0.095
Continuous NSES	0.94(0.91;0.97)	<0.001	0.92(0.88;0.96)	<0.001	0.96(0.91;1.00)	0.059
	Lack of Diabetes Control (Continuous HbA1c %)					
Variable	**Beta (95% CI)**	**p-val**	**Beta (95% CI)**	**p-val**	**Beta (95% CI)**	**p-val**
Tertile 1 of NSES (Low)	0 (Ref.)		0 (Ref.)		0 (Ref.)	
Tertile 2 of NSES (Middle)	-0.05(-0.10;-0.01)	0.021	-0.07(-0.13;-0.01)	0.021	-0.03(-0.09;0.03)	0.31
Tertile 3 of NSES (High)	-0.11(-0.15;-0.06)	<0.001	-0.13(-0.19;-0.07)	<0.001	-0.08(-0.14;-0.02)	0.014
Continuous NSES	-0.04(-0.06;-0.02)	<0.001	-0.05(-0.07;-0.02)	<0.001	-0.03(-0.06;-0.01)	0.011
	Diabetes Incidence					
Variable	**HR (95% CI)**	**p-val**	**HR (95% CI)**	**p-val**	**HR (95% CI)**	**p-val**
Tertile 1 of NSES (Low)	1 (Ref.)		1 (Ref.)		1 (Ref.)	
Tertile 2 of NSES (Middle)	0.85(0.77;0.95)	0.003	0.87(0.77;0.99)	0.041	0.83(0.71;0.97)	0.021
Tertile 3 of NSES (High)	0.75(0.68;0.83)	<0.001	0.80(0.71;0.91)	<0.001	0.69(0.59;0.80)	<0.001
Continuous NSES	0.86(0.83;0.90)	<0.001	0.90(0.85;0.94)	<0.001	0.82(0.77;0.87)	<0.001

* Note: models adjusted by age, sex and year and clustered on the census section.

NSES and Diabetes Incidence

Overall, at one and two years of follow-up the diabetes incidence was 5.7 per 1000 and 10.5 per 1000. Figure 2 shows the Kaplan-Meier estimate of diabetes incidence by tertile of NSES, showing a social gradient in diabetes incidence (lower SES corresponding to higher diabetes incidence). Table 2 also shows the results of the adjusted Cox Proportional Hazards models. We found a significant interaction by sex (p for interaction = 0.004). Men living in medium and high NSES neighborhoods had a 13% and 20% reduced hazard of diabetes incidence, as compared men living in low NSES neighborhoods (HR=0.87, 95% CI 0.77 to 0.99, and HR=0.80, 95% CI 0.71 to 0.91). Women saw a stronger association, with those living in medium and high NSES neighborhoods showing a 17% and 31% decrease in the hazard of diabetes compared to low NSES neighborhoods (HR=0.83, 95% CI 0.71 to 0.97, and HR=0.69, 95% CI 0.59 to 0.80). These associations were consistent in models looking at continuous NSES: a 1 standard-deviation increase in NSES was associated with a 10% and 18% decrease in the hazard of incident diabetes in men and women, respectively (HR=0.90, 95% CI 0.85 to 0.94, and HR=0.82, 95% CI 0.77 to 0.87).

Discussion

This study has shown a strong association between neighborhood socioeconomic status and diabetes burden. In particular, there is a dose-response association: as neighborhood socioeconomic status increases, diabetes prevalence, lack of control and incidence decrease in a linear fashion. This association is seen for both a categorical (tertiles) and a continuous operationalization of the exposure. There seems to be an interaction by sex in the association with diabetes prevalence and incidence, which is stronger in women as compared to men.

Previous studies have shown analogous results to ours. A report by Larranaga found an increase in the prevalence of diabetes in more deprived neighborhoods in the Basque Country (Northern Spain), using a sample of primary care practices[10], displaying a similar interaction by sex as our study. Other studies using electronic health records in other countries have found significant associations between area-level poverty, deprivation or socioeconomic status and diabetes prevalence[5], incidence[7] and control[6]. Other studies using data from cross-sectional surveys or cohort studies, but with similar spatial units as ours have also found significant associations in the US[4]. Our study is the first in Spain (and to our knowledge in Southern Europe) to show an association between neighborhood socioeconomic status and diabetes control.

Strengths and Limitations

Our study has several strengths. First, we study the entire population of an area of a very large city (Madrid) where almost 600,000 people live[13]. This results in a very large sample size and decreased concerns for selection bias as compared to regular cohort studies or surveys[12]. Second, the diagnosis of diabetes in our EHR has been validated before and shown to have a very high validity with a kappa of 0.99[16]. Third, HbA1c represents a robust measure of diabetes control and is the standard of care in clinical practice. Fourth and last, we used an exposure constructed from publicly available indicators, increasing the replicability of our findings and the applicability to other health outcomes. Our study also has some limitations. First and foremost, while the validity of our measures of diabetes prevalence, incidence and control is high[16], we cannot achieve the standardization of measurements that cohort studies do. Second, while our exposure is built from publicly available indicators, this also restricts our capacity to build a complex exposure that may capture socioeconomic status better. Third, 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 do not have data on individual level socioeconomic status. Unmeasured confounding by neighborhood selection may be an important source of bias in our study. However, whether adjusting for individual level socioeconomic status brings estimates closer to the truth or induces over-adjustment may depend on the level of social mobility of each country[18].

The implications of our study are several. As this is the first study, to our knowledge, to show strong contextual gradients in diabetes burden in Spain, we believe these findings should be incorporated in the National Health Equity Strategy. Research wise, this study opens the possibility to study the connection between contextual factors (the food, physical activity, tobacco and alcohol environment) and diabetes. For reference, our results regarding the 2-year incidence of diabetes in high SES as compared to low SES areas (HR=0.80 and 0.66 in men and women, respectively) have an association with reduced diabetes incidence similar to a 1.2 kg and 2.1 kg reduction in body weight in the DPP trial[19].

The World Health Organization (WHO) has identified social determinants as underlying many of the health inequities observed within countries[20], and resulting strategies to ameliorate social determinants through systems change are underway in countries including Spain[21]. For diabetes, an unhealthy diet, lack of physical activity, and subsequent obesity are some of the main modifiable risk factors that are adversely impacted by social determinants. Auchincloss and Christine have reported over several studies[22 23] increased prevalence and incidence of diabetes with lower availability of healthy foods or physical-activity promoting resources. Nonetheless, there's a lack of research on these mechanistic pathways in Spain. In particular, the association of contextual socioeconomic status and unhealthy food environments has not been thoroughly replicated in Europe and may actually follow a different gradient[24]. Understanding the contextual contributors to the social patterning of diabetes we have described in this study can offer opportunities for prevention through structural changes[25]. Nonetheless, these strategies need not be restricted to macro-level changes. Globally, intensive lifestyle diabetes prevention programs[26] present an evidence-based opportunity that is not reliant on environmental structural change. Diabetes prevention programs using this model have proven effective in reducing diabetes incidence in persons in lower income communities in the U.S[27]. There is also initial evidence that patient diabetes self-management programs focused on barriers to care and social determinants can improve diabetes self-management skills, health behaviors, and HbA1c in low income patients and communities[28 29]. These may be directions for future intervention research in lower socioeconomic status neighborhoods in Spain.

Conclusion

To conclude, our study is the first to show a contextual social gradient in diabetes burden by contextual measures of socioeconomic status in Southern Europe. The use of universal electronic health records of an entire improves representability and statistical power, and provides a clear representation of population health patterns. Future studies should provide targets for intervention to address this population health inequity.

Author Contributions: UB and MF conceptualized the study. UB conducted the statistical analysis and drafted the first version of the manuscript. UB, MF and FHB interpreted results and revised the first version of the manuscript. LSP and IdC organized and conducted health data collection. MF obtained funding for the study. All authors approved the final version of the manuscript.

Data Sharing: No additional data available.

Ethical Approval: This study was approved by the Madrid Primary Care Research Committee.

Conflicts of Interest: The authors declare that they have no conflict of interest

Funding: Manuel Franco was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007– 2013/ERC Starting Grant HeartHealthyHoods Agreement n. 336893). Felicia Hill-Briggs was supported by the National Institute of Diabetes and Digestive and Kidney Diseases Diabetes Research Center (P30DK079637). The funding sources had no role in the analysis, writing or decision to submit the manuscript.

Acknowledgements: Usama Bilal was supported by a Johns Hopkins Center for a Livable Future – Lerner Fellowship and a Postgraduate Fellowship from the *Obra Social La Caixa*.

Figure Captions

Figure 1 Estimated Diabetes Prevalence and Control by levels of Neighborhood Socioeconomic Status

Figure 2 Adjusted Kaplan-Meier Survival Curve of Diabetes Incidence by Neighborhood Socioeconomic Status

Figure 2 footnote: results predicted from models adjusted by age, sex and year and clustered on the census section. For prediction purposes age was set to the 3rd category (60 to 70 years of age)

References

1. NCD-RisC. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *The Lancet* 2016;387(10027):1513-30.

2. Zhang P, Zhang X, Brown J, et al. Global healthcare expenditure on diabetes for 2010 and 2030. *Diabetes Research and Clinical Practice* 2010;87(3):293-301.

3. Rose G. Sick individuals and sick populations. *International journal of epidemiology* 1985;14(1):32-8.

4. Piccolo RS, Duncan DT, Pearce N, et al. 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)* 2015;130:79-90.

5. Connolly V, Unwin N, Sherriff P, et al. 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* 2000;54(3):173-77.

6. Hippisley-Cox J, O'Hanlon S, Coupland C. 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)* 2004;329(7477):1267-9.

7. Mezuk B, Chaikiat Å, Li X, et al. Depression, neighborhood deprivation and risk of type 2 diabetes. *Health & place* 2013;23:63-9.

8. White JS, Hamad R, Li X, et al. Long-term effects of neighbourhood deprivation on diabetes risk: quasi-experimental evidence from a refugee dispersal policy in Sweden. *The Lancet Diabetes & Endocrinology* 2016;4(6):517-24.

9. Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes--a randomized social experiment. *The New England journal of medicine* 2011;365(16):1509-19.

10. Larrañaga I, Arteagoitia JM, Rodriguez JL, et al. 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* 2005;22(8):1047-53.

11. Reques L, Giráldez-García C, Miqueleiz E, et al. Educational differences in mortality and the relative importance of different causes of death: a 7-year follow-up study of Spanish adults. *Journal of epidemiology and community health* 2014;68(12):1151-60.

12. Weisskopf MG, Sparrow D, Hu H, et al. Biased Exposure-Health Effect Estimates from Selection in Cohort Studies: Are Environmental Studies at Particular Risk? *Environmental health perspectives* 2015;123(11):1113-22. [published Online First: 2015/05/09]

13. Bilal U, Diez J, Alfayate S, et al. Population cardiovascular health and urban environments: the Heart Healthy Hoods exploratory study in Madrid, Spain. *BMC medical research methodology* 2016;16:104.

14. Soriguer F, Goday A, Bosch-Comas A, et al. Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Di@bet.es Study. *Diabetologia* 2012;55(1):88-93.

15. Borrell C, Malmusi D, Muntaner C. Introduction to the “Evaluating the Impact of Structural Policies on Health Inequalities and Their Social Determinants and

Fostering Change"(SOPHIE) Project. *International Journal of Health Services* 2017;47(1):10-17.

16. de Burgos-Lunar C, Salinero-Fort MA, Cardenas-Valladolid J, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. *BMC medical research methodology* 2011;11:146. [published Online First: 2011/11/01]
17. Harrell F. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer 2015.
18. Glass TA, Bilal U. Are neighborhoods causal? Complications arising from the 'stickiness' of ZNA. *Soc Sci Med* 2016;166:244-53.
19. Hamman RF, Wing RR, Edelstein SL, et al. Effect of Weight Loss With Lifestyle Intervention on Risk of Diabetes. *Diabetes Care* 2006;29(9):2102-07.
20. Marmot M. Social determinants of health inequalities. *The Lancet* 2005;365(9464):1099-104.
21. Merino Merino B, Vega Morales J, Gil Luciano A, et al. Integration of social determinants of health and equity into health strategies, programmes and activities: health equity training process in Spain. 2013
22. Auchincloss AH, Roux A, Mujahid MS, et al. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: The multi-ethnic study of atherosclerosis. *Archives of internal medicine* 2009;169(18):1698-704.
23. Christine PJ, Auchincloss AH, Bertoni AG, et al. Longitudinal associations between neighborhood physical and social environments and incident type 2 diabetes

mellitus: The multi-ethnic study of atherosclerosis (mesa). *JAMA Internal Medicine* 2015;175(8):1311-20.

24. Díez J, Bilal U, Cebrecos A, et al. Understanding differences in the local food environment across countries: A case study in Madrid (Spain) and Baltimore (USA). *Preventive medicine* 2016;89:237-44.

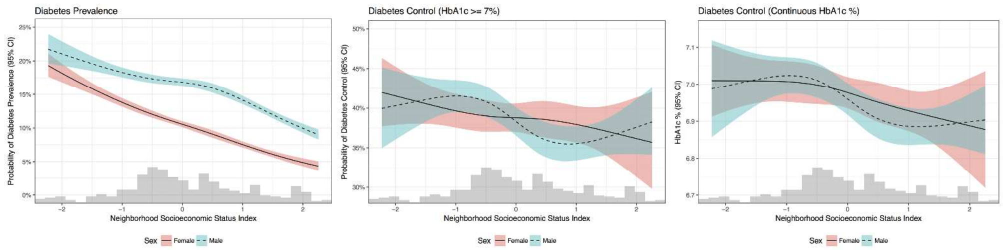
25. Franco M, Bilal U, Diez-Roux AV. Preventing non-communicable diseases through structural changes in urban environments. *J Epidemiol Community Health* 2015;69(6):509-11.

26. Diabetes Prevention Program Research G. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *The Lancet* 2009;374(9702):1677-86.

27. Ackermann RT, Finch EA, Brizendine E, et al. Translating the Diabetes Prevention Program into the Community: The DEPLOY Pilot Study. *American journal of preventive medicine* 2008;35(4):357-63.

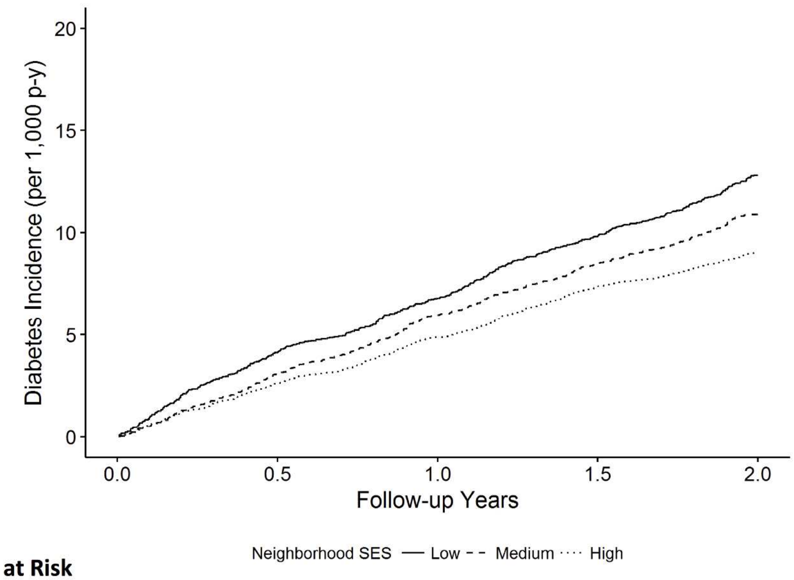
28. Fitzpatrick SL, Golden SH, Stewart K, et al. Effect of DECIDE (Decision-making Education for Choices In Diabetes Everyday) Program Delivery Modalities on Clinical and Behavioral Outcomes in Urban African Americans With Type 2 Diabetes: A Randomized Trial. *Diabetes Care* 2016

29. Hill-Briggs F, Lazo M, Peyrot M, et al. Effect of Problem-Solving-Based Diabetes Self-Management Training on Diabetes Control in a Low Income Patient Sample. *Journal of general internal medicine* 2011;26(9):972.



Estimated Diabetes Prevalence and Control by levels of Neighborhood Socioeconomic Status

508x127mm (300 x 300 DPI)



People at Risk					
Low	60206	59957	59791	57506	56699
Medium	73293	73050	72834	70437	69551
High	112641	112325	112069	108687	107647

Adjusted Kaplan-Meier Survival Curve of Diabetes Incidence by Neighborhood Socioeconomic Status

254x190mm (300 x 300 DPI)

Neighborhood Socioeconomic Status: Operationalization of Indicators

To measure neighborhood socioeconomic status we explored all, to our knowledge, available data sources on social, economic and contextual factors in Madrid, Spain. We looked for readily available indicators (to ease replicability), that were measured at the neighborhood or census section level (to improve granularity) and that were available for several years (to allow for further studies looking at longitudinal changes). After a literature review and data exploration we used seven indicators in four domains.

Operationalization of indicators**:**

1. Education:
    a. Low level of education: people with primary studies or below / people aged 25 or above
    b. High level of education: people with university education or above / people aged 25 or above
2. Wealth:
    a. Property value: average sales price of housing properties in EUR per m^2
3. Occupation:
    a. Part-time employment: workers in part-time employment / all workers
    b. Temporary employment: workers in temporary employment / all workers
    c. Manual occupational class: workers in manual or unqualified jobs / all workers
4. Living conditions:
    a. Unemployment rate: individuals registered as unemployed / all people aged 16 to 64

Data Sources:

1. Education: The education indicators were obtained from the Padron, a continuous universal census of the entire population used for administrative purposes. It includes data on education level which we recategorized into the four typically used levels in Spain: no formal studies, primary education, secondary education, and university education. We also obtained proportion of people above age 25 to use as the denominator.
2. Wealth: Property value was obtained from the Idealista Report, a yearly study of neighborhood-level sale prices of all housing sold through the biggest real state corporation in Spain (Idealista). All data was

downloaded from the statistics website of the City Government of Madrid. Property value data from the IDEALISTA Report contains data for all houses listed for sale in their website on the first day of each year. The report contains data at the neighborhood level (n=128 each year). To translate this to the census section level, we obtained data from the IDEALISTA API (http://developers.idealista.com/access-request) on April 18th 2016. We collected all housing units for sale on that day, including their price, size and geocoded location. We overlayed a census section polygon file and assigned each housing unit to a census section. With this, we constructed a measure of average property value per census section for 2016. We then used a weighted linear mixed model with property value at the census section as the dependent variable, and property value at the neighborhood level (from the IDEALISTA Report 2016 data) as a fixed and random coefficient (at the neighborhood level ,with an unstructured covariance structure), and the following fixed effects for each census section: % low education, % high education, % immigration from non-oecd countries, % people below age 25, % people above age 25, and a quadratic fixed term for each indicator. Each observation was weighted by the number of housing units on sale on each census section. We then predicted the property value in each census section in 2013 by replacing the data above with the respective data from 2014. To diagnose this imputation we correlated the predicted values for 2016 with the observed values in 2016, finding a pearson correlation coefficient of 0.93.

3. Occupation: The total number of workers, and the number of workers in part-time and temporary employment along with the occupational class were obtained from the Social Security registries. These were downloaded from the statistics website of the City Government of Madrid.
4. Living conditions: Registered unemployment was obtained from the statistics of the Employment Service (SEPE), downloaded from the statistics website of the City Government of Madrid. The denominator was, given the lack of a better measure for the active population at this geographical level, the amount of people between 16 and 64 years of age in the neighborhood, obtained from the Padron.

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	**Title page (Abstract includes electronic health records)**
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Page 4
Objectives	3	State specific objectives, including any prespecified hypotheses			Page 5
Methods					
Study Design	4	Present key elements of study design early in the paper			Page 6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			Page 6
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the		RECORD 6.1: The methods of study population selection (such as codes or	Page 6, Page 8

		sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Page 7
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Page 6, 7, 8
Bias	9	Describe any efforts to address potential sources of bias			Page 8
Study size	10	Explain how the study size was			Page 6

		arrived at			
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Page 8
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			Page 8
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Page 6, 8
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-	Page 6, 8

				level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Page 6, Table 1
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Table 1
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			Table 1
Main results	16	(a) Give unadjusted estimates			Table 1, 2, Figures

		and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			1 and 2
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			Page 11, 12
Discussion					
Key results	18	Summarise key results with reference to study objectives			Page 14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Page 16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			Page 14,15, 16
Generalisability	21	Discuss the generalisability (external validity) of the study results			Page 14, 16

Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			Page 18
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code..	Page 18

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

Association of Neighborhood Socioeconomic Status and Diabetes Burden using Electronic Health Records in Madrid (Spain): The Heart Healthy Hoods Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-021143.R1
Article Type:	Research
Date Submitted by the Author:	13-Apr-2018
Complete List of Authors:	Bilal, Usama; Drexel Dornsife School of Public Health, Urban Health Collaborative; Universidad de Alcala de Henares Facultad de Medicina y Ciencias de la Salud, Social and Cardiovascular Research Group Hill-Briggs, F; Johns Hopkins University School of Medicine, Department of Medicine; Johns Hopkins Medical Institutions, Welch Center for Prevention, Epidemiology, and Clinical Research Sanchez-Perruca, Luis; Madrid Health Service, Information Services Directorate; Instituto de Salud Carlos III, Health Services Research on Chronic Patients Network (REDISSEC) del Cura, Isabel; Madrid Health Service, Primary Care Research Unit; Instituto de Salud Carlos III, Health Services Research on Chronic Patients Network (REDISSEC) Franco, M; Universidad de Alcala de Henares Facultad de Medicina y Ciencias de la Salud, Social and Cardiovascular Research Group
Primary Subject Heading:	Diabetes and endocrinology
Secondary Subject Heading:	Epidemiology, General practice / Family practice
Keywords:	social epidemiology, social inequalities, Spain, record linkage, Diabetes, Neighborhoods

SCHOLARONE™ Manuscripts

Association of Neighborhood Socioeconomic Status and Diabetes Burden using Electronic Health Records in Madrid (Spain): The Heart Healthy Hoods Study

Usama Bilal[1,2,3*], Felicia Hill-Briggs[1,4,5],
Luis Sánchez-Perruca[6,7], Isabel del Cura[7,8,9], and Manuel Franco[1,2]

1. Department of Epidemiology. Johns Hopkins Bloomberg School of Public Health. 615 N Wolfe St. Baltimore, MD, 21205 USA.
2. Social and Cardiovascular Epidemiology Research Group. Universidad de Alcalá. Ctra Madrid Barcelona km 33.6, Alcala de Henares, Madrid, 28871 Spain.
3. Urban Health Collaborative, Drexel Dornsife School of Public Health, 3600 Market St, Philadelphia, PA, 19104, USA
4. Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins Medical Institutions, 2024 E Monument St, Baltimore, MD 21205 USA
5. The Johns Hopkins University School of Medicine, Baltimore, MD
6. Information Services Directorate. Primary Care Management. Madrid Health Service, Madrid, Spain.
7. Health Services Research on Chronic Patients Network (REDISSEC). ISCIII, Spain.
8. Primary Care Research Unit. Primary Care Management. Madrid Health Service, Spain.
9. Area of Preventive Medicine and Public Health. Rey Juan Carlos University, Madrid, Spain.

Words: 3553; Abstract: 269; Figures: 3; Tables: 2; References: 53

Corresponding Author:
Usama Bilal, MD MPH PhD
Urban Health Collaborative
Drexel Dornsife School of Public Health
3600 Market St, 7th Floor, Philadelphia, PA, 19104
Telephone: +1 410 419 6170
Email: ubilal@drexel.edu

Abstract

Objective: To study the association between neighborhood socioeconomic status and diabetes prevalence, incidence, and control in the entire population of Northeastern Madrid, Spain.

Setting: Electronic health records of the primary-care system in four districts of Madrid (Spain)

Participants: 269,942 people aged 40 or above, followed from 2013 to 2014

Exposure: Neighborhood Socioeconomic Status (NSES), measured using a composite index of 7 indicators from 4 domains of education, wealth, occupation, and living conditions.

Primary Outcome Measures: Diagnosis of diabetes based on ICPC-2 codes and glycated hemoglobin (HbA1c %)

Results: In regression analyses adjusted by age and sex and compared to individuals living in low NSES neighborhoods, men living in medium and high NSES neighborhoods had 10% (95% CI: 6-15%) and 29% (95% CI: 25-32%) lower prevalence of diabetes, while women had 27% (95% CI: 23-30%) and 50% (95% CI: 47-52%) lower prevalence of diabetes. Moreover, the hazard of diabetes in men living in medium and high NSES neighborhoods was 13% (95% CI: 1-23%) and 20% (95% CI: 9-29%) lower, while the hazard of diabetes in women living in medium and high NSES neighborhoods was 17% (95% CI: 3-29%) and 31% (95% CI: 20-41%) lower. Individuals living in medium and high SES neighborhoods had 8% (95% CI: 2-15%) and 15% (95% CI: 9-21%) lower prevalence of lack of diabetes control, and a decrease in average HbA1c % of 0.05 (95% CI: 0.01-0.10) and 0.11 (95% CI: 0.06-0.15).

Conclusions: Diabetes prevalence, incidence, and lack of control increased with decreasing neighborhood socioeconomic status in a Southern European city. Future studies should provide mechanistic insights and targets for intervention to address this health inequity.

Keywords: social epidemiology; social inequalities; record linkage; diabetes; Spain; neighborhood/place

Strengths and Limitations

- We study the entire population of an area of a very large city (Madrid) where almost 600,000 people live, resulting in a very large sample size and decreased concerns for selection bias as compared to regular cohort studies or surveys.
- The diagnosis of diabetes in our EHR has been validated before and shown to have a very high validity with a kappa of 0.99, but we cannot achieve the level of standardization of measurements of cohort studies.
- We use HbA1c which is a robust measure of diabetes control and is the standard of care in clinical practice.
- We used an exposure constructed from publicly available indicators, increasing the replicability of our findings and the applicability to other health outcomes, but restricting our capacity to build a complex exposure that may capture socioeconomic status better.
- The available data for individual level confounders were restricted to basic sociodemographics (age and sex), which opens the possibility for residual confounding in our inferences (especially individual level socioeconomic status).

Introduction

The burden of diabetes has seen a large increase in Western countries in recent decades[1]. 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[2]. Population preventive strategies are needed to decrease this burden[3], taking into consideration mass-influences that differ across populations[3].

Among these mass influences are neighborhood characteristics. A large body of literature has explored contextual socioeconomic influences on health. In particular, the association between neighborhood socioeconomic status and several measures of diabetes (prevalence, incidence or control) is robust and has been replicated in the US[4-10], other Anglo-Saxon countries[11-19], and Northern and Central Europe[20-26] including in experimental or quasi-experimental settings[21 27]. Nonetheless, these influences have received scant attention in Southern Europe[28]. Moreover, previous studies have shown a strong social gradient in diabetes mortality in Spain, which warrants further mechanistic insights into its causes[29]. Recent studies have shown that segregation patterns and neighborhood selection phenomena is changing in Southern Europe[30], warranting an study of the health outcomes associated with these changes.

Finally, many of the studies outlined above use data from research-driven cohort studies. While these types of 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[31]. 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[31]. With Electronic Health Records in a health system with universal health coverage these drawbacks may be overcome by avoiding the necessity for sampling altogether.

Taking the above into consideration, we studied the association between neighborhood socioeconomic status and diabetes prevalence, incidence and control in an electronic health record-based cohort of the entire population of Northeastern Madrid that includes data on more than 640,000 people.

Methods

Study setting

This study was conducted within the HeartHealthyHoods project (www.hhhproject.eu) in the city of Madrid, Spain[32]. We took data for 2013 and 2014 from all Health Care centers in four districts of the city of Madrid, all belonging to the same health district. These four districts contain around 20% of the total population of Madrid and are representative of the rest of the city of Madrid (Appendix Figure 1). Our unit of analysis is the census section (n=427), which is the smallest area for which the census collects data and has around 1200 people (range: 583 to 3865). Individual-level data were obtained from Electronic Health Records (EHR) including 640,217 individuals registered in any Health Center of the area. These EHR contain data on patient age, sex, residential location, clinical diagnoses, and laboratory values (lipids and HbA1c).

Since this screening for cardiovascular risk factors is limited to people 40 years and older[32], we restricted our dataset to people born after January 1, 1973 (aged 40 or above by 2013). Our final study sample was composed of 270,660 individuals, of which 23,908 had a diagnosis of diabetes. Primary Care EHR include 99.5% of the individuals living in the area per the Census[32].

Neighborhood Socioeconomic Status

The main exposure of this study was Neighborhood Socioeconomic Status (NSES). To measure NSES, we considered the 4 domains of the Spanish Commission to Reduce Health Inequalities[33]: education, wealth, occupation and living conditions. To search for indicators to measure these 4 domains, we explored all, to our knowledge, available data sources on social, economic and contextual factors in Madrid, Spain. We looked for readily available indicators (to ease replicability), that were measured at the neighborhood or census section level (to improve granularity) and that were available for several years (to allow for further studies looking at longitudinal changes). After this process we selected 7 indicators that represent the 4 domains: education, (1) Primary education (% people above 25 years of age with primary studies or below), (2) University education (% people above 25 years of age with university education or above); wealth, (3) Average housing prices (per sq. m); occupation, (4) Part-time employment (% workers in part-time jobs), (5) Temporary employment (% workers in temporary jobs), (6) Manual occupational class (% workers in manual or unqualified jobs); and living conditions (7) Unemployment rate (% registered unemployed individuals / people aged 16 to 64). Indicator data were obtained from the Padrón (a continuous and universal census collected for administrative purposes), the social security and employment services registries and the IDEALISTA report (a report from a large real estate corporation in Spain). All data were available by January 2013. The Online Resource contains a detailed description of the operationalization of indicators.

We computed a weighted index of the seven indicators by: (1) making the directionality of the associations consistent, by reversing some of the indicators (primary education, part-time employment, temporary employment, manual occupational class, and unemployment rate) so that all indicators had a consistent association with the final index; (2) for each indicator, we centered by the mean and divided by the standard deviation in order to obtain a Z-score of each indicator; (3) in each domain, we averaged the Z-score of each indicator, resulting in a Z-score for each domain (education, wealth, occupation, and living conditions); (4) finally, we calculated the composite index of NSES by averaging the Z-score of each of the 4 domains. This composite NSES index was then operationalized in separate analyses as a categorical variable (NSES in tertiles) or as a continuous variable.

Diabetes Prevalence, Incidence and Control

Diabetes diagnoses were extracted from the EHR for all individuals, as recorded by primary care physicians during their usual clinical practice. A type-2 diabetes diagnosis 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%)[34]. Prevalent cases were defined as diabetes diagnoses dated before January 1st 2013. Incident cases were those occurring from January 1st 2013 to December 31st 2014 in people free of diabetes by baseline (January 1st 2013). We operationalized lack of diabetes control as either a dichotomous variable (HbA1c >=7%) or a continuous variable (HbA1c %). If more than one value of HbA1c was available, we used the last available measurement of the year.

Statistical Methods

The overall goal of this analysis is to study the association between neighborhood socioeconomic status and diabetes prevalence, incidence and control. We computed descriptive statistics by tertile of neighborhood socioeconomic status.

To study the association between neighborhood socioeconomic status and diabetes prevalence or lack of control (binary indicator) we used a log-binomial regression model with robust standard errors clustered at the census section level using a sandwich Huber-White estimator. These models were adjusted for age (in five categories; 40 to 49, 50 to 59, 60 to 69, 70 to 79 and 80 and above) and sex. Continuous HbA1c (for diabetes control) was examined using a linear regression with robust standard errors clustered at the census section level using a sandwich Huber-White estimator. Around 21% of the sample that had prevalent diabetes had no HbA1c % measured in 2013 or 2014. To assess whether this missing data phenomenon affected our inferences, we did a sensitivity analysis using a conditional mean imputation of HbA1c % in people with diabetes. In this model, we predicted the HbA1c % value using age, sex, health care center, NSES index, and diagnosis of other cardiovascular risk factors or conditions (hypertension, dyslipidemia, prevalent cardiovascular disease, chronic kidney disease and retinopathy). We then compared the point estimates of the association between prevalent lack of control and average HbA1c % obtained with and without conditional mean imputation.

In the analysis of diabetes incidence, each individual entered the sample on January 1st 2013 and exited on the date of diabetes diagnosis (outcome), date of death (censored), date of moving out of a health center in the area (censored), or study end by December 31st 2014 (administrative censoring). We used Kaplan-Meier survival estimates to explore differences in the hazard of diabetes incidence by NSES tertile. Cox Proportional Hazards models were used to estimate the adjusted association, with clustered standard errors on the census section. Since we censored individuals at death, a potential competing risk, our estimates from the model are analogous to cause-specific hazard ratios, and can therefore be interpreted as the increase in the hazard of diabetes if people that do not die. We checked the proportionality of hazards assumption by plotting Schoenfeld residuals and by checking their trend over time[35].

To graphically display the association between the exposure and the outcome variables we also modeled the associations above using restricted cubic splines with 4 knots in the percentiles recommended by Harrell[36]. A previous report in the Spanish setting highlighted a significant interaction by sex of contextual socioeconomic status and diabetes[28], so we explored whether this interaction existed in our analysis and displayed stratified results if this was the case. All analyses were conducted in R v3.3.0 (R Software Foundation). This study was approved by the Madrid Primary Care Research Committee.

Results

Study Population

Table 1 shows a description of the study population by tertile of neighborhood socioeconomic status (NSES) and in the total population. The total sample size was 269,942 people, with around 25%, 30% and 45% of the population living in low, medium and high NSES areas. Overall, the median age was 56.5 (IQR=47.4 to 69.8) and 54.9% of the population were women. 8.8% of the population over 40 years of age had diabetes, 1.0% developed diabetes during follow up, and the average HbA1c in diabetic people was 6.7 (IQR=6.2 to 7.5). Thirty-nine percent of all diabetic people had uncontrolled diabetes (HbA1c equal or above 7%). Stratifying the population by tertile of NSES revealed that younger people lived in neighborhoods with higher SES. The prevalence of diabetes decreased sharply with NSES (11.9% in the lowest NSES, 9.6% in the medium NSES and 6.5% in the highest NSES), and the incidence of diabetes followed a similar gradient by NSES (1.3%, 1.1% and 0.9% in the lowest, medium and highest NSES areas).

Table 1: Study Population by January 1st 2013.

Variable	Total	Tertile 1 (Lowest NSES)	Tertile 2 (Mid NSES)	Tertile 3 (High NSES)	p-val*
Sample Size (N)	269942	68369	81072	120501	
Median Age [IQR]	56.5 [47.4;69.8]	58.6 [48.3;74.5]	58.1 [48.0;71.1]	54.7 [46.6;66.9]	<0.001
% Men	45.1%	44.6%	44.2%	45.9%	<0.001
% Women	54.9%	55.4%	55.8%	54.1%	
% Death during follow-up	1.2%	1.4%	1.3%	1.0%	<0.001
% Moved during follow-up	0.8%	0.8%	0.8%	0.8%	0.673
% with Prevalent Diabetes	8.8%	11.9%	9.6%	6.5%	<0.001
% with Incident Diabetes+	1.0%	1.3%	1.1%	0.9%	<0.001
Median HbA1c [IQR]	6.7 [6.2;7.5]	6.7 [6.2;7.5]	6.7 [6.2;7.5]	6.7 [6.2;7.4]	<0.001
HbA1c >=7%	38.8%	40.5%	38.7%	37.1%	0.237
HbA1c < 5%	0.3%	0.3%	0.4%	0.3%	0.285
HbA1c 5-6.5%	41.1%	40.0%	40.5%	42.7%	
HbA1c 6.5-7%	20.1%	19.4%	20.6%	20.3%	
HbA1c 7-9%	32.4%	34.0%	32.2%	30.9%	
HbA1c >9%	6.1%	6.3%	6.3%	5.7%	
Primary Education, % [IQR]	24.6% [15.1;32.2]	36.3% [30.7;40.3]	24.7% [20.8;27.9]	11.6% [7.1;19.5]	<0.001
University Education, % [IQR]	20.8% [13.0;33.7]	10.2% [7.4;13.0]	20.8% [16.8;24.7]	40.1% [29.9;52.5]	<0.001
Unemployment Rate, % [IQR]	12.6% [10.6;13.8]	13.8% [13.8;16.4]	12.6% [12.0;12.7]	8.9% [7.8;10.6]	<0.001
Part-Time Workers, % [IQR]	23.4% [18.7;25.9]	26.7% [24.8;26.8]	23.4% [22.4;25.9]	16.5% [12.7;19.4]	<0.001
Temporary Workers, % [IQR]	19.0% [17.3;20.9]	20.9% [20.4;21.5]	20.4% [18.9;20.9]	16.7% [13.8;18.2]	<0.001
Manual Class, % [IQR]	37.1% [27.4;40.0]	40.3% [40.0;43.1]	37.1% [36.2;40.0]	22.4% [17.4;30.2]	<0.001
Property Value, EUR/m^2 [IQR]	2286.0 [1975.0;2659.0]	1776.0 [1561.0;1971.0]	2243.0 [2128.0;2398.0]	2832.0 [2608.0;3382.0]	<0.001
SES Index [IQR]	0.0 [-0.6;0.6]	-0.8 [-1.2;-0.6]	-0.2 [-0.3;0.1]	1.0 [0.6;1.6]	<0.001

*p-value -values for continuous individual-level characteristics were computed using a clustered Somers' D comparison of medians; p-values for categorical individual-level characteristics were computed using Donner's Chi2 adjusted for clustered data. P-values for contextual characteristics were conducted at the neighborhood level using a Kruskal-Wallis test for the comparison of medians.+: Incident diabetes refers to new diagnoses of diabetes in 2013 or 2014 in people free of diabetes at baseline.

NSES and Diabetes Prevalence

Table 2 shows the association between NSES and diabetes prevalence, control, and incidence. Diabetes prevalence was associated in a dose-response manner to NSES. This association was significantly stronger in women as compared to men (p for interaction <0.001). In particular, compared to men living in low NSES neighborhoods, those living in medium NSES neighborhoods had 8% lower prevalence of having diabetes (PR=0.92, 95% CI 0.89 to 0.96), while those living in the highest NSES neighborhoods had 24% lower prevalence of diabetes (PR=0.76, 95% CI 0.74 to 0.80). In the case of women, those living in medium and high NSES neighborhoods had 24% and 46% lower prevalence of diabetes, respectively, as compared to those living low NSES neighborhoods (PR=0.76, 95% CI 0.73 to 0.79, and PR=0.54, 95% CI 0.52 to 0.57). These associations were consistent in models looking at continuous NSES: a 1 standard-deviation increase in NSES was associated with 14% and 26% lower prevalence of diabetes in men and women, respectively (PR=0.86, 95% CI 0.84 to 0.87, PR=0.74, 95% CI 0.72 to 0.75). Figure 1 shows the association using continuous NSES with restricted cubic splines, where the steeper pattern for women is evident.

NSES and Diabetes Control

Table 2 also shows the association between NSES and diabetes control, operationalized as a dichotomous variable (lack of diabetes control, or HbA1c >=7%) or a continuous variable (HbA1c %). There was no significant interaction by sex in the NSES and diabetes control (p for interaction=0.219 and 0.358 in the dichotomous and continuous model). As compared to diabetic people living in the lowest NSES neighborhoods, those living in medium NSES areas had 5% lower prevalence of lack of controlled diabetes (PR=0.95, 95% CI 0.91 to 0.99), while those living in the highest NSES areas had 9% lower prevalence of lack of diabetes control (PR=0.91, 95% CI 0.87 to 0.95). Moreover, a 1 standard-deviation increase in NSES was associated with 4% lower prevalence of lack of diabetes control (PR=0.96, 95% CI 0.94 to 0.98). These associations were maintained when looking at continuous HbA1c: diabetic people living in medium and high SES neighborhoods had a lower average HbA1c % (see Table 2). Figure 2 shows the prevalence of lack of controlled diabetes and average HbA1c levels across levels of NSES using restricted cubic splines, showing a linear decrease both in lack of control and in average HbA1c % with increasing NSES. In the sensitivity analysis using conditional mean imputation of HbA1c %, we found no change in our inferences after accounting for missing HbA1c % (see Appendix Figure 2).

Table 2: Association of Neighborhood Socioeconomic Status and Diabetes Outcomes

	Diabetes Prevalence					
	Total		Men		Women	
Variable	**PR (95% CI)**	**p-val**	**PR (95% CI)**	**p-val**	**PR (95% CI)**	**p-val**
Tertile 1 of NSES (Low)	1 (Ref.)		1 (Ref.)		1 (Ref.)	
Tertile 2 of NSES (Middle)	0.84(0.82;0.87)	<0.001	0.92(0.89;0.96)	<0.001	0.76(0.73;0.79)	<0.001
Tertile 3 of NSES (High)	0.66(0.64;0.68)	<0.001	0.76(0.74;0.80)	<0.001	0.54(0.52;0.57)	<0.001
Continuous NSES	0.80(0.79;0.81)	<0.001	0.86(0.84;0.87)	<0.001	0.74(0.72;0.75)	<0.001
	Lack of Diabetes Control (HbA1c >= 7%)					
Variable	**PR (95% CI)**	**p-val**	**PR (95% CI)**	**p-val**	**PR (95% CI)**	**p-val**
Tertile 1 of NSES (Low)	1 (Ref.)		1 (Ref.)		1 (Ref.)	
Tertile 2 of NSES (Middle)	0.95(0.91;0.99)	0.014	0.94(0.88;0.99)	0.033	0.96(0.90;1.02)	0.158
Tertile 3 of NSES (High)	0.91(0.87;0.95)	<0.001	0.88(0.83;0.93)	<0.001	0.95(0.89;1.01)	0.117
Continuous NSES	0.96(0.94;0.98)	<0.001	0.95(0.93;0.98)	<0.001	0.97(0.95;1.00)	0.07
	Lack of Diabetes Control (Continuous HbA1c %)					
Variable	**Beta (95% CI)**	**p-val**	**Beta (95% CI)**	**p-val**	**Beta (95% CI)**	**p-val**
Tertile 1 of NSES (Low)	0 (Ref.)		0 (Ref.)		0 (Ref.)	
Tertile 2 of NSES (Middle)	-0.05(-0.10;-0.01)	0.021	-0.07(-0.13;-0.01)	0.021	-0.03(-0.09;0.03)	0.31
Tertile 3 of NSES (High)	-0.11(-0.15;-0.06)	<0.001	-0.13(-0.19;-0.07)	<0.001	-0.08(-0.14;-0.02)	0.014
Continuous NSES	-0.04(-0.06;-0.02)	<0.001	-0.05(-0.07;-0.02)	<0.001	-0.03(-0.06;-0.01)	0.011
	Diabetes Incidence					
Variable	**HR (95% CI)**	**p-val**	**HR (95% CI)**	**p-val**	**HR (95% CI)**	**p-val**
Tertile 1 of NSES (Low)	1 (Ref.)		1 (Ref.)		1 (Ref.)	
Tertile 2 of NSES (Middle)	0.85(0.77;0.95)	0.003	0.87(0.77;0.99)	0.041	0.83(0.71;0.97)	0.021
Tertile 3 of NSES (High)	0.75(0.68;0.83)	<0.001	0.80(0.71;0.91)	<0.001	0.69(0.59;0.80)	<0.001
Continuous NSES	0.86(0.83;0.90)	<0.001	0.90(0.85;0.94)	<0.001	0.82(0.77;0.87)	<0.001

* Note: models adjusted by age, sex and year and clustered on the census section. Results for Diabetes Prevalence and Lack of Diabetes Control (binary) are shown in Prevalence Ratios (95% CI); results for Lack of Diabetes Control (continuous) are presented as changes in average HbA1c % (95% CI); results for Diabetes Incidence are presented as Hazard Ratios (95% CI).

NSES and Diabetes Incidence

Overall, at one and two years of follow-up the diabetes incidence was 5.7 per 1000 and 10.5 per 1000. Figure 3 shows the Kaplan-Meier estimate of diabetes incidence by tertile of NSES, showing a social gradient in diabetes incidence (lower NSES corresponding to higher diabetes incidence, $p<0.001$). Table 2 also shows the results of the adjusted Cox Proportional Hazards models. We found a significant interaction by sex (p for interaction = 0.004). The hazard of diabetes incidence in men living in medium and high NSES neighborhoods was 13% and 20% lower compared to men living in low NSES neighborhoods (HR=0.87, 95% CI 0.77 to 0.99, and HR=0.80, 95% CI 0.71 to 0.91). Women saw a stronger association, as the hazard of diabetes incidence in women living in medium and high NSES neighborhoods was 17% and 31% lower compared to women living in low NSES neighborhoods (HR=0.83, 95% CI 0.71 to 0.97, and HR=0.69, 95% CI 0.59 to 0.80). These associations were consistent in models looking at continuous NSES: a 1 standard-deviation increase in NSES was associated with a 10% and 18% decrease in the hazard of incident diabetes in men and women, respectively (HR=0.90, 95% CI 0.85 to 0.94, and HR=0.82, 95% CI 0.77 to 0.87). We tested the assumption of proportionality of hazards and found no evidence to reject the null hypothesis of proportionality (p for the global chi^2-test=0.604 for the unadjusted model, and 0.365 for the fully adjusted model).

Discussion

This study has shown a strong association between neighborhood socioeconomic status and diabetes burden. In particular, there is a dose-response association: as NSES increases, diabetes prevalence, lack of control and incidence decrease in a linear fashion. This association is seen for both a categorical (tertiles) and a continuous operationalization of the exposure. There seems to be an interaction by sex in the association with diabetes prevalence and incidence, which is stronger in women as compared to men.

Previous studies have shown analogous results to ours. A report by Larranaga found an increase in the prevalence of diabetes in more deprived neighborhoods in the Basque Country (Northern Spain), using a sample of primary care practices[28], displaying a similar interaction by sex as our study. Other studies using EHR in other countries have found significant associations between area-level poverty, deprivation or socioeconomic status and diabetes prevalence, incidence and control. A study by Cox[15] using EHR from a Scottish region found increased diabetes prevalence in more deprived areas, as measured using the Carstair Index of Deprivation. Studies by Mezuk[20] and Sundquist[26] showed a significant increase in diabetes incidence in the Swedish population living in medium and high deprivation neighborhoods, measured using four indicators of NSES. Several more studies in the UK[12 16 18 19], US[10], and Israel[37] have studied the association of NSES with diabetes control as measured by HbA1c % in EHR, finding a consistent gradient similar to ours (lower NSES associated with lower likelihood of control or higher HbA1c %). Other studies using data from cross-sectional surveys or cohort studies, but with similar spatial units as ours have also found significant associations in the US[4-6 9], France[22], and Sweden[23]. Our study is the first in Spain (and to our knowledge in Southern Europe) to show an association between NSES and diabetes control.

Strengths and Limitations

Our study has several strengths. First, we study the entire population of an area of a very large city (Madrid) where almost 600,000 people live[32]. This results in a very large sample size and decreased concerns for selection bias as compared to regular cohort studies or surveys[31]. Second, the diagnosis of diabetes in our EHR has been validated before and shown to have a very high validity with a kappa of 0.99[34]. Third, HbA1c represents a robust measure of diabetes control and is the standard of care in clinical practice. Fourth and last, we used an exposure constructed from publicly available indicators, increasing the replicability of our findings and the applicability to other health outcomes. Our study also has some limitations. First and foremost, while the validity of our measures of diabetes prevalence, incidence and control is high[34], we cannot achieve the standardization of measurements that cohort studies do. While there exists the possibility of differential measurement error, we have no reason to suspect that the accuracy of the measure of diabetes prevalence varies by socioeconomic status, given that Spain has a Universal Health Care system. Second, while our exposure is built from publicly available indicators, this also restricts our capacity to build a complex exposure that may capture socioeconomic status better. Third, the available data for individual level confounders were restricted to basic sociodemographic variables, age and sex, which opens the possibility for residual confounding in our inferences. In particular, we do not have data on individual level socioeconomic status. Unmeasured confounding by neighborhood selection may be an important source of bias in our study. However, whether adjusting for individual level socioeconomic status brings estimates closer to the truth or induces over-adjustment may depend on the level of social mobility of each country[38]. Last, the generalizability of these results to other Spanish or European cities may be limited for cities that do not have similar segregation patterns. Recent research has shown increased segregation in Madrid, with levels similar to London[30].

The implications of our study are several. As this is the first study, to our knowledge, to show strong contextual gradients in diabetes burden in Spain, we believe these findings should be incorporated in the National Health Equity Strategy. Research wise, this study opens the possibility to study the connection between contextual factors (the food, physical activity, tobacco and alcohol environment) and diabetes. Future studies may consider providing specific mechanistic insights into the contextual determinants of diabetes in Southern Europe. For example, Auchincloss and Christine have reported over several studies[39 40] increased prevalence and incidence of diabetes with lower availability of healthy foods or physical-activity promoting resources, but research on these mechanistic pathways is lacking in Spain and Southern Europe in general. In particular, the association of contextual socioeconomic status and unhealthy food environments has not been thoroughly replicated in Europe and may actually follow a different gradient[41]. We have previously shown that neighborhoods in Madrid with improving socioeconomic status indicators have an increased proportion of supermarkets and decreased proportion of fruit and vegetable stores[42], a contextual change undesired by neighbors and perceived as not conducive to better diets[43 44]. We have also previously shown that walkability may follow an inverse social gradient in

Madrid[45] (worse walkability in higher NSES areas), but that this association may not hold in gentrifying areas[45]. In summary, understanding the mechanisms (and therefore potential intervention targets) linking NSES to diabetes may require studies that take into consideration changes in both the exposure and the outcome side.

The World Health Organization (WHO) has identified social determinants as underlying many of the health inequities observed within countries[46], and resulting strategies to ameliorate social determinants through systems change are underway in countries including Spain[47]. For diabetes, an unhealthy diet, lack of physical activity, and subsequent obesity are some of the main modifiable risk factors that are adversely impacted by social determinants. Understanding the contextual contributors to the social patterning of diabetes we have described in this study can offer opportunities for prevention through structural changes[48]. Nonetheless, these strategies need not be restricted to macro-level changes. Globally, intensive lifestyle diabetes prevention programs[49] present an evidence-based opportunity that is not reliant on environmental structural change. Diabetes prevention programs using this model have proven effective in reducing diabetes incidence in persons in lower income communities in the U.S[50]. There is also initial evidence that patient diabetes self-management programs focused on barriers to care and social determinants can improve diabetes self-management skills, health behaviors, and HbA1c in low income patients and communities[51 52]. For reference, our results regarding the 2-year incidence of diabetes in high SES as compared to low SES areas (HR=0.80 and 0.69 in men and women, respectively) have an association with reduced diabetes incidence similar to a 1.2 kg and 2.1 kg reduction in body weight in the DPP trial[53]. Focusing diabetes prevention efforts in lower NSES areas may help in ameliorating health inequalities. Our study provides a framework to identify areas that may require more intensive efforts, by linking diabetes outcomes with readily measurable NSES.

Conclusion

To conclude, our study is the first to show a social gradient in diabetes burden by contextual measures of socioeconomic status in Southern Europe. The use of universal electronic health records of an entire population improves representability and statistical power, providing a rich representation of population health patterns. Future studies should provide targets for intervention to address this population health inequity.

Author Contributions: UB and MF conceptualized the study. UB conducted the statistical analysis and drafted the first version of the manuscript. UB, MF and FHB interpreted results and revised the first version of the manuscript. LSP and IdC organized and conducted health data collection. MF obtained funding for the study. All authors approved the final version of the manuscript.

Data Sharing: Neighborhood SES indicators are available online as detailed in the Appendix. Health data was obtained from the primary care system and cannot be shared due to privacy concerns.

Ethical Approval: This study was approved by the Madrid Primary Care Research Committee.

Patient and Public Involvement: patients and/or public were not directly involved in the design of this study.

Conflicts of Interest: The authors declare that they have no conflict of interest

Funding: MF was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007– 2013/ERC Starting Grant HeartHealthyHoods Agreement n. 336893). FHB was supported by the National Institute of Diabetes and Digestive and Kidney Diseases Diabetes Research Center (P30DK079637). The funding sources had no role in the analysis, writing or decision to submit the manuscript.

Acknowledgements: UB was supported by a Johns Hopkins Center for a Livable Future – Lerner Fellowship and a Postgraduate Fellowship from the *Obra Social La Caixa*.

Figure Captions

Figure 1 Estimated Diabetes Prevalence by levels of Neighborhood Socioeconomic Status

Figure 2 Estimated Diabetes Control by levels of Neighborhood Socioeconomic Status

Figure 3 Adjusted Kaplan-Meier Survival Curve of Diabetes Incidence by Neighborhood Socioeconomic Status

Figure 3 footnote: results predicted from models adjusted by age, sex and year and clustered on the census section. For prediction purposes age was set to the 3rd category (60 to 70 years of age)

References

1. NCD-RisC. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *The Lancet* 2016;387(10027):1513-30.

2. Zhang P, Zhang X, Brown J, et al. Global healthcare expenditure on diabetes for 2010 and 2030. *Diabetes Research and Clinical Practice* 2010;87(3):293-301.

3. Rose G. Sick individuals and sick populations. *International journal of epidemiology* 1985;14(1):32-8.

4. Corriere MD, Yao W, Xue QL, et al. The association of neighborhood characteristics with obesity and metabolic conditions in older women. *The journal of nutrition, health & aging* 2014;18(9):792-8.

5. Garcia L, Lee A, Zeki Al Hazzouri A, et al. The Impact of Neighborhood Socioeconomic Position on Prevalence of Diabetes and Prediabetes in Older Latinos: The Sacramento Area Latino Study on Aging. *Hispanic health care international : the official journal of the National Association of Hispanic Nurses* 2015;13(2):77-85.

6. Gaskin DJ, Thorpe RJ, McGinty EE, et al. Disparities in diabetes: the nexus of race, poverty, and place. *American journal of public health* 2014;104(11):2147-55.

7. Johnson RC, Schoeni RF. Early-life origins of adult disease: national longitudinal population-based study of the United States. *American journal of public health* 2011;101(12):2317-24.

8. Krishnan S, Cozier YC, Rosenberg L, et al. Socioeconomic status and incidence of type 2 diabetes: results from the Black Women's Health Study. *American journal of epidemiology* 2010;171(5):564-70.

9. Piccolo RS, Duncan DT, Pearce N, et al. 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)* 2015;130:79-90.

10. Geraghty EM, Balsbaugh T, Nuovo J, et al. 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* 2010;23(1):88-96.

11. Connolly V, Unwin N, Sherriff P, et al. 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* 2000;54(3):173-77.

12. Hippisley-Cox J, O'Hanlon S, Coupland C. 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)* 2004;329(7477):1267-9.

13. Booth GL, Creatore MI, Moineddin R, et al. Unwalkable neighborhoods, poverty, and the risk of diabetes among recent immigrants to Canada compared with long-term residents. *Diabetes care* 2013;36(2):302-8.

14. Borkhoff CM, Saskin R, Rabeneck L, et al. Disparities in receipt of screening tests for cancer, diabetes and high cholesterol in Ontario, Canada: a population-based study using area-based methods. *Canadian journal of public health = Revue canadienne de santé publique* 2013;104(4):e284-90.

15. Cox M, Boyle PJ, Davey PG, et al. Locality deprivation and Type 2 diabetes incidence: a local test of relative inequalities. *Social science & medicine (1982)* 2007;65(9):1953-64.

16. Bebb C, Kendrick D, Stewart J, et al. Inequalities in glycaemic control in patients with Type 2 diabetes in primary care. *Diabetic medicine : a journal of the British Diabetic Association* 2005;22(10):1364-71.

17. Guthrie B, Emslie-Smith A, Morris AD. Which people with Type 2 diabetes achieve good control of intermediate outcomes? Population database study in a UK region. *Diabetic medicine : a journal of the British Diabetic Association* 2009;26(12):1269-76.

18. James GD, Baker P, Badrick E, et al. Ethnic and social disparity in glycaemic control in type 2 diabetes; cohort study in general practice 2004-9. *Journal of the Royal Society of Medicine* 2012;105(7):300-8.

19. Millett C, Car J, Eldred D, et al. Diabetes prevalence, process of care and outcomes in relation to practice size, caseload and deprivation: national cross-sectional study in primary care. *Journal of the Royal Society of Medicine* 2007;100(6):275-83.

20. Mezuk B, Chaikiat Å, Li X, et al. Depression, neighborhood deprivation and risk of type 2 diabetes. *Health & place* 2013;23:63-9.

21. White JS, Hamad R, Li X, et al. Long-term effects of neighbourhood deprivation on diabetes risk: quasi-experimental evidence from a refugee dispersal policy in Sweden. *The Lancet Diabetes & Endocrinology* 2016;4(6):517-24.

22. Chaix B, Billaudeau N, Thomas F, et al. Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology (Cambridge, Mass)* 2011;22(1):18-26.

23. Cubbin C, Sundquist K, Ahlén H, et al. Neighborhood deprivation and cardiovascular disease risk factors: protective and harmful effects. *Scandinavian journal of public health* 2006;34(3):228-37.

24. Müller G, Kluttig A, Greiser KH, et al. 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* 2013;178(2):221-30.

25. Müller G, Wellmann J, Hartwig S, et al. Association of neighbourhood unemployment rate with incident Type 2 diabetes mellitus in five German regions. *Diabetic medicine : a journal of the British Diabetic Association* 2015;32(8):1017-22.

26. Sundquist K, Eriksson U, Mezuk B, et al. Neighborhood walkability, deprivation and incidence of type 2 diabetes: A population-based study on 512,061 Swedish adults. *Health & Place* 2015;31:24-30.

27. Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes--a randomized social experiment. *The New England journal of medicine* 2011;365(16):1509-19.

28. Larrañaga I, Arteagoitia JM, Rodriguez JL, et al. 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* 2005;22(8):1047-53.

29. Reques L, Giráldez-García C, Miqueleiz E, et al. Educational differences in mortality and the relative importance of different causes of death: a 7-year follow-up study of Spanish adults. *Journal of Epidemiology and Community Health* 2014;68(12):1151-60.
30. Tammaru T, van Ham M, Marcińczak S, et al. Socio-Economic Segregation in European Capital Cities: East Meets West: Taylor & Francis 2015.
31. Weisskopf MG, Sparrow D, Hu H, et al. Biased Exposure-Health Effect Estimates from Selection in Cohort Studies: Are Environmental Studies at Particular Risk? *Environmental health perspectives* 2015;123(11):1113-22.
32. Bilal U, Diez J, Alfayate S, et al. Population cardiovascular health and urban environments: the Heart Healthy Hoods exploratory study in Madrid, Spain. *BMC medical research methodology* 2016;16:104.
33. Borrell C, Malmusi D, Muntaner C. Introduction to the “Evaluating the Impact of Structural Policies on Health Inequalities and Their Social Determinants and Fostering Change”(SOPHIE) Project. *International Journal of Health Services* 2017;47(1):10-17.
34. de Burgos-Lunar C, Salinero-Fort MA, Cardenas-Valladolid J, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. *BMC medical research methodology* 2011;11:146.
35. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994;81(3):515-26.

36. Harrell F. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer 2015.

37. Wilf-Miron R, Peled R, Yaari E, et al. Disparities in diabetes care: role of the patient's socio-demographic characteristics. *BMC public health* 2010;10:729.

38. Glass TA, Bilal U. Are neighborhoods causal? Complications arising from the 'stickiness' of ZNA. *Soc Sci Med* 2016;166:244-53.

39. Auchincloss AH, Roux A, Mujahid MS, et al. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: The multi-ethnic study of atherosclerosis. *Archives of internal medicine* 2009;169(18):1698-704.

40. Christine PJ, Auchincloss AH, Bertoni AG, et al. Longitudinal associations between neighborhood physical and social environments and incident type 2 diabetes mellitus: The multi-ethnic study of atherosclerosis (mesa). *JAMA Internal Medicine* 2015;175(8):1311-20.

41. Díez J, Bilal U, Cebrecos A, et al. Understanding differences in the local food environment across countries: A case study in Madrid (Spain) and Baltimore (USA). *Preventive medicine* 2016;89:237-44.

42. Bilal U, Jones-Smith J, Diez J, et al. Neighborhood Social and Economic Change and Retail Food Environment Change in Madrid (Spain): The Heart Healthy Hoods Study. *Health & Place* 2018

43. Díez J, Conde P, Sandin M, et al. Understanding the local food environment: A participatory photovoice project in a low-income area in Madrid, Spain. *Health & Place* 2017;43:95-103.

44. Díez J, Valiente R, Ramos C, et al. The mismatch between observational measures and residents' perspectives on the retail food environment: a mixed-methods approach in the Heart Healthy Hoods study. *Public health nutrition* 2017:1-10.

45. Gullon P, Bilal U, Cebrecos A, et al. Intersection of neighborhood dynamics and socioeconomic status in small-area walkability: the Heart Healthy Hoods project. *Int J Health Geogr* 2017;16(1):21.

46. Marmot M. Social determinants of health inequalities. *The Lancet* 2005;365(9464):1099-104.

47. Merino Merino B, Vega Morales J, Gil Luciano A, et al. Integration of social determinants of health and equity into health strategies, programmes and activities: health equity training process in Spain. 2013

48. Franco M, Bilal U, Diez-Roux AV. Preventing non-communicable diseases through structural changes in urban environments. *J Epidemiol Community Health* 2015;69(6):509-11.

49. Diabetes Prevention Program Research Group. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *The Lancet* 2009;374(9702):1677-86.

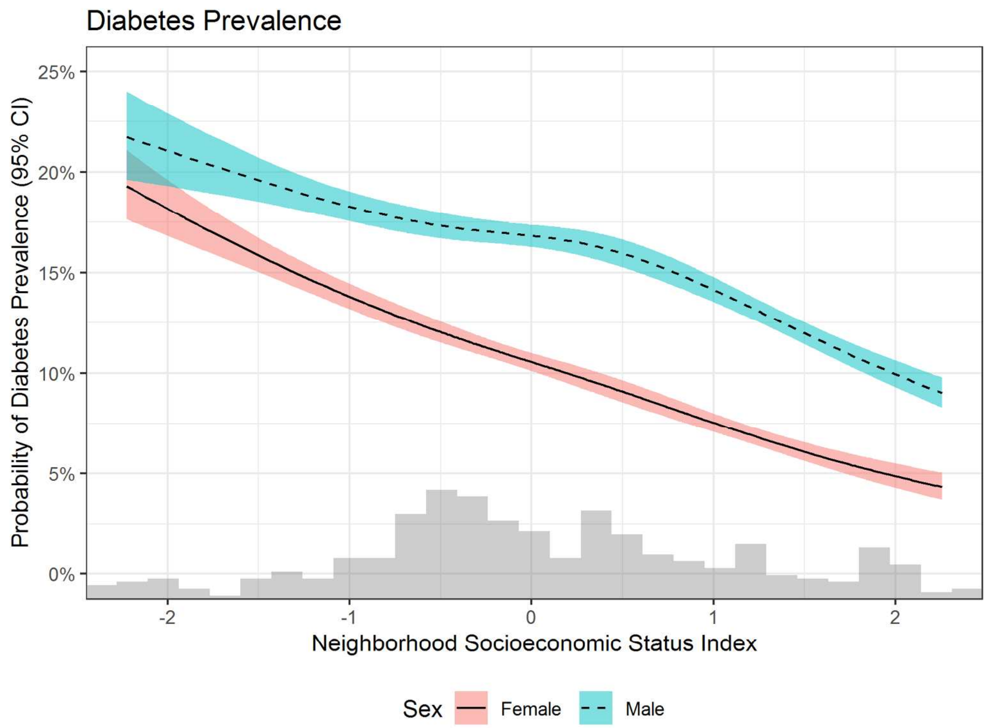
50. Ackermann RT, Finch EA, Brizendine E, et al. Translating the Diabetes Prevention Program into the Community: The DEPLOY Pilot Study. *American journal of preventive medicine* 2008;35(4):357-63.

51. Fitzpatrick SL, Golden SH, Stewart K, et al. Effect of DECIDE (Decision-making Education for Choices In Diabetes Everyday) Program Delivery Modalities on

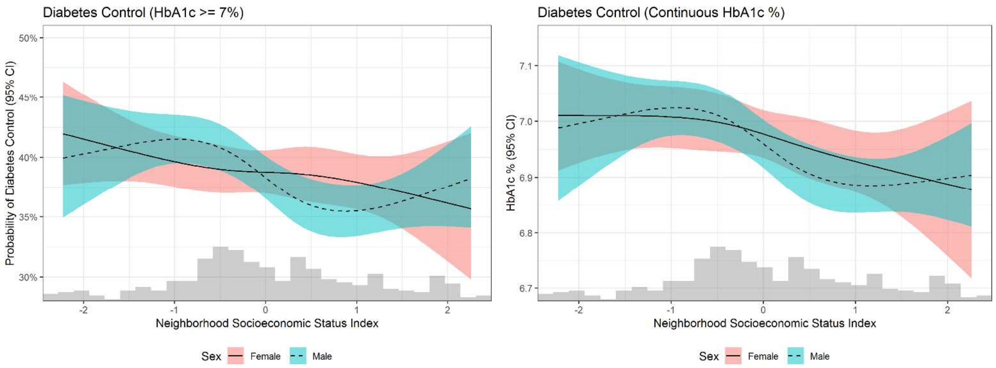
Clinical and Behavioral Outcomes in Urban African Americans With Type 2 Diabetes: A Randomized Trial. *Diabetes Care* 2016

52. Hill-Briggs F, Lazo M, Peyrot M, et al. Effect of Problem-Solving-Based Diabetes Self-Management Training on Diabetes Control in a Low Income Patient Sample. *Journal of general internal medicine* 2011;26(9):972.

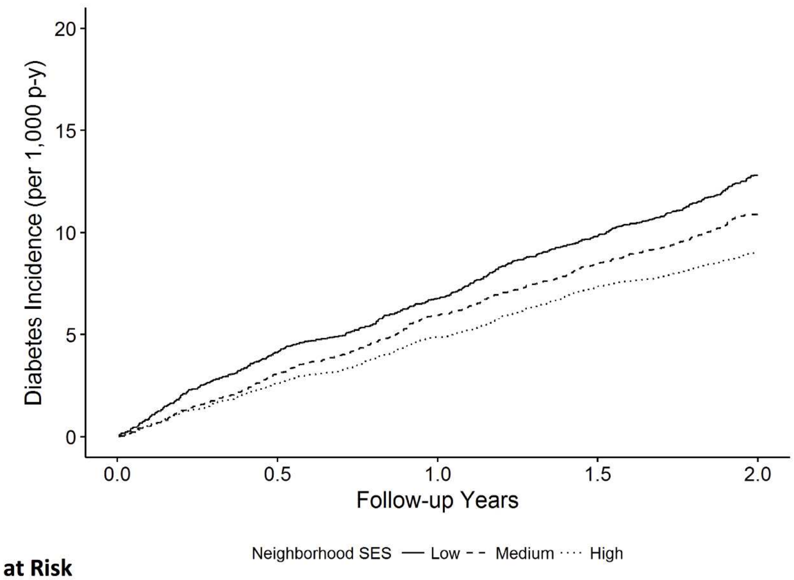
53. Hamman RF, Wing RR, Edelstein SL, et al. Effect of Weight Loss With Lifestyle Intervention on Risk of Diabetes. *Diabetes Care* 2006;29(9):2102-07.



Estimated Diabetes Prevalence by levels of Neighborhood Socioeconomic Status



Estimated Diabetes Control by levels of Neighborhood Socioeconomic Status



People at Risk					
Low	60206	59957	59791	57506	56699
Medium	73293	73050	72834	70437	69551
High	112641	112325	112069	108687	107647

Adjusted Kaplan-Meier Survival Curve of Diabetes Incidence by Neighborhood Socioeconomic Status

254x190mm (300 x 300 DPI)

Neighborhood Socioeconomic Status: Operationalization of Indicators

To measure neighborhood socioeconomic status we explored all, to our knowledge, available data sources on social, economic and contextual factors in Madrid, Spain. We looked for readily available indicators (to ease replicability), that were measured at the neighborhood or census section level (to improve granularity) and that were available for several years (to allow for further studies looking at longitudinal changes). After a literature review and data exploration we used seven indicators in four domains.

Operationalization of indicators**:**

1. Education:
    a. Primary education: people with primary studies or below / people aged 25 or above
    b. University education: people with university education or above / people aged 25 or above
2. Wealth:
    a. Property value: average sales price of housing properties in EUR per m^2
3. Occupation:
    a. Part-time employment: workers in part-time employment / all workers
    b. Temporary employment: workers in temporary employment / all workers
    c. Manual occupational class: workers in manual or unqualified jobs / all workers
4. Living conditions:
    a. Unemployment rate: individuals registered as unemployed / all people aged 16 to 64

Data Sources:

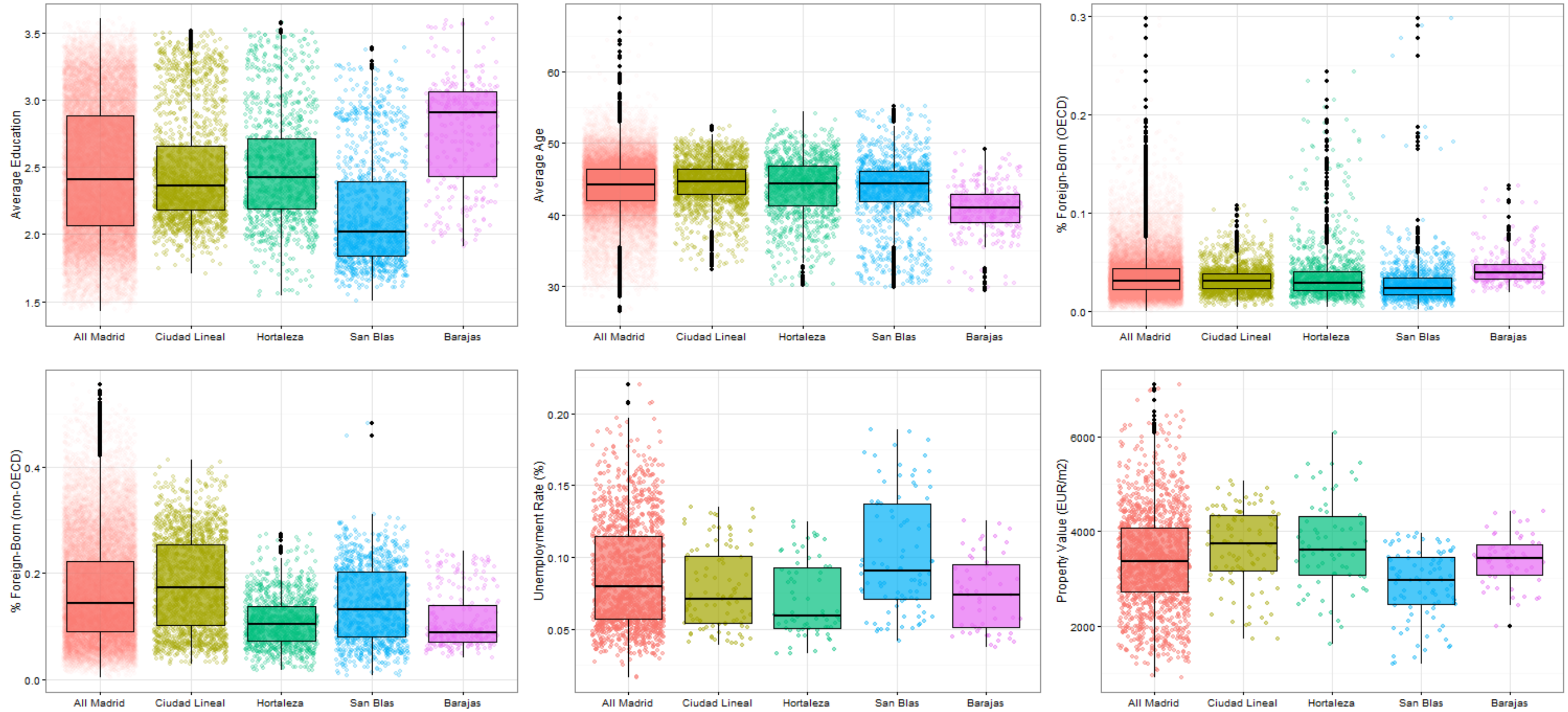
1. Education: The education indicators were obtained from the Padron, a continuous universal census of the entire population used for administrative purposes. It includes data on education level which we recategorized into the four typically used levels in Spain: no formal studies, primary education, secondary education, and university education. We also obtained proportion of people above age 25 to use as the denominator.
2. Wealth: Property value was obtained from the Idealista Report, a yearly study of neighborhood-level sale prices of all housing sold through the biggest real state corporation in Spain (Idealista). All data was

downloaded from the statistics website of the City Government of Madrid. Property value data from the IDEALISTA Report contains data for all houses listed for sale in their website on the first day of each year. The report contains data at the neighborhood level (n=128 each year). To translate this to the census section level, we obtained data from the IDEALISTA API (http://developers.idealista.com/access-request) on April 18th 2016. We collected all housing units for sale on that day, including their price, size and geocoded location. We overlayed a census section polygon file and assigned each housing unit to a census section. With this, we constructed a measure of average property value per census section for 2016. We then used a weighted linear mixed model with property value at the census section as the dependent variable, and property value at the neighborhood level (from the IDEALISTA Report 2016 data) as a fixed and random coefficient (at the neighborhood level ,with an unstructured covariance structure), and the following fixed effects for each census section: % primary education, % university education, % immigration from non-oecd countries, % people below age 25, % people above age 25, and a quadratic fixed term for each indicator. Each observation was weighted by the number of housing units on sale on each census section. We then predicted the property value in each census section in 2013 by replacing the data above with the respective data from 2014. To diagnose this imputation we correlated the predicted values for 2016 with the observed values in 2016, finding a pearson correlation coefficient of 0.93.

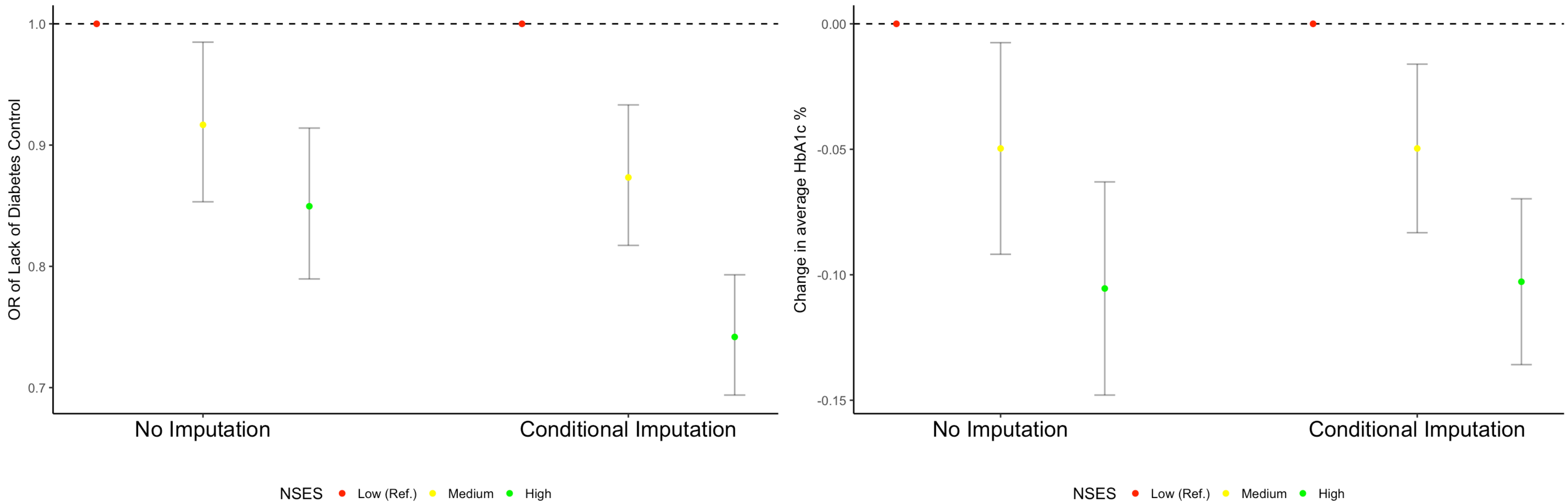
3. Occupation: The total number of workers, and the number of workers in part-time and temporary employment along with the occupational class were obtained from the Social Security registries. These were downloaded from the statistics website of the City Government of Madrid.

4. Living conditions: Registered unemployment was obtained from the statistics of the Employment Service (SEPE), downloaded from the statistics website of the City Government of Madrid. The denominator was, given the lack of a better measure for the active population at this geographical level, the amount of people between 16 and 64 years of age in the neighborhood, obtained from the Padron.

Appendix Figure 1. Distribution of key sociodemographic and socioeconomic variables in the four districts as compared to the entire city of Madrid



Appendix Figure 2: Comparison of the OR of Lack of Diabetes Control and the Change in Average HbA1c % in models using complete case analysis (ignoring missing data) and in models using conditional mean imputation of missing HbA1c %



STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1	Association of Neighborhood Socioeconomic Status and Diabetes Burden using Electronic Health Records in Madrid (Spain): The Heart Healthy Hoods Study
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2	See abstract
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4	See introduction
Objectives	3	State specific objectives, including any prespecified hypotheses	5	Taking the above into consideration, we studied the association between neighborhood socioeconomic status and diabetes prevalence, incidence and control in an electronic health record-based cohort of the entire population of Northeastern Madrid that includes data on more than 640,000 people.
Methods				
Study design	4	Present key elements of study design early in the paper	6	See methods section
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-8	See methods section
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	6-9	Individual-level data were obtained from Electronic Health

		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants		Records (EHR) including 640,217 individuals registered in any Health Center of the area. These EHR contain data on patient age, sex, residential location, clinical diagnoses, and laboratory values (lipids and HbA1c). Since this screening for cardiovascular risk factors is limited to people 40 years and older32, we restricted our dataset to people born after January 1, 1973 (aged 40 or above by 2013). Our final study sample was composed of 270,660 individuals, of which 23,908 had a diagnosis of diabetes. Primary Care EHR include 99.5% of the individuals living in the area per the Census
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/A	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6,7	See methods section
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*6-8*	*See methods section*
Bias	9	Describe any efforts to address potential sources of bias	8-9	See methods section
Study size	10	Explain how the study size was arrived at	6	See methods section

Continued on next page

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8	See methods section
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8	See methods section
		(*b*) Describe any methods used to examine subgroups and interactions	8	A previous report in the Spanish setting highlighted a significant interaction by sex of contextual socioeconomic status and diabetes28, so we explored whether this interaction existed in our analysis and displayed stratified results if this was the case.
		(*c*) Explain how missing data were addressed	9	Around 21% of the sample that had prevalent diabetes had no HbA1c % measured in 2013 or 2014. To assess whether this missing data phenomenon affected our inferences, we did a sensitivity analysis using a conditional mean imputation of HbA1c % in people with diabetes. In this model, we predicted the HbA1c % value using age, sex, health care center, NSES index, and diagnosis of other cardiovascular risk factors or conditions (hypertension, dyslipidemia, prevalent cardiovascular disease, chronic kidney disease and retinopathy). We then compared the point estimates of the association between prevalent lack of control

				and average HbA1c % obtained with and without conditional mean imputation.
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	9	See methods section
		(*e*) Describe any sensitivity analyses	9	
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6,11, Table 1	Our final study sample was composed of 270,660 individuals, of which 23,908 had a diagnosis of diabetes. Primary Care EHR include 99.5% of the individuals living in the area per the Census32.
		(b) Give reasons for non-participation at each stage		
		(c) Consider use of a flow diagram		
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1	See table 1
		(b) Indicate number of participants with missing data for each variable of interest	9	Around 21% of the sample that had prevalent diabetes had no HbA1c % measured in 2013 or 2014.
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	11	See table 1
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*11*	*8.8% of the population over 40 years of age had diabetes, 1.0% developed diabetes during follow up, and the average HbA1c in diabetic people was 6.7 (IQR=6.2 to 7.5).*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	*NA*	
		Cross-sectional study—Report numbers of outcome events or summary measures	*NA*	

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 2	See table 2
		(*b*) Report category boundaries when continuous variables were categorized	Table 2	See table 2
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	*NA*	

Continued on next page

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	14,15	See results section
Discussion				
Key results	18	Summarise key results with reference to study objectives	18	See discussion
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	19	See discussion
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	18,19,20	See discussion
Generalisability	21	Discuss the generalisability (external validity) of the study results	18,20	See discussion
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	23	Funding: MF was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013/ERC Starting Grant HeartHealthyHoods Agreement n. 336893). FHB was supported by the National Institute of Diabetes and Digestive and Kidney Diseases Diabetes Research Center (P30DK079637). The funding sources had no role in the analysis, writing or decision to submit the manuscript.

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.