MASTER

Neighbourhood green space and health-seeking behaviour in the Netherlands
a multilevel analysis using open data

Wuestman, M.L.

Award date:
2016

Eindhoven, December 2015

Neighbourhood green space and health-seeking behaviour in the Netherlands: a multilevel analysis using open data

by Mignon L. Wuestman

identity number 0746806

in partial fulfilment of the requirements for the degree of

Master of Science
in Innovation Sciences

Supervisors:

Prof. dr. C.C.P. Snijders	TU/e: Human Technology Interaction
Dr. ir. R.N.A. Bekkers	TU/e: Technology, Innovation & Society
S. Maltha	Dialogic Innovatie & Interactie

Preface

This report consists of two parts. First, the paper that is the result of my thesis, and second, the literature review that inspired it. Both can be read as standalone documents, although I feel that both are required to understand the full depth of this very interesting field.

The fact that you are reading this preface means that you are either incredibly interested in the relation between neighbourhood green space and individual health, or in what I have done the last half year. In both cases, I thank you for the attempt, and hope my thesis can inspire you further. For me, my appreciation for this topic came with my wish to study big data for a humanitarian purpose. I took the time to find an interesting project; I never thought I would read journals on medicine and public health for my master thesis, and I am very happy I took the chance of working in a field that is completely new to me for this project.

A special thanks goes out to everyone at Dialogic, where I spent most of my time for six months, for giving me such a warm welcome, for helping me to escape from puzzling situations, for a never ending flow of puns, and for the interesting discussions over lunch. Thanks to Sven and Arthur for helping me find my way in the field of public health care, and to Menno, Jasper, Bram and Tommy for teaching this very tech-impatient girl how to keep your calm while using a computer. Of course, I also want to thank my first supervisor, Chris, for his knowledge and critical feedback even though the field was almost as new to him as it was for me, and my second supervisor, Rudi, for his last-minute enthusiasm and new perspective.

I promised a deserved spot in my preface to everyone who took my Facebook challenge, and who, doing this, helped me shape numerous theories about the strange distribution of average GP consult fees. So, thanks to Mark, Maurice, Rick, Alain, Daan, Joris, Lotte, Lieke, Freek, René and Linette, and a shout-out to social media!

Finally, a big thanks to my friends in the In ‘t Aufstudeerhok, de Baron or het Internaat, and of course to my boyfriend Jim, my parents Erick and Anita, and my brother Lars, who each have a much stronger appreciation of green space than I do and who stayed incredibly supportive and interested in my progress for all this time. I dedicate this project to Lars, who, while I study the effect of nature on well-being from books and figures, was actually putting my theories into practice by riding the waves in Egypt and the Philippines.

I hope you enjoy reading about my project as much as I enjoyed creating it!

Content

List of tables

List of figures

Neighbourhood green space and health-seeking behaviour in the Netherlands: a multilevel analysis using open data

M.L. Wuestman

Abstract

This explorative research studies associations between the relative amount of various types of neighbourhood green space and individual primary health-seeking behaviour. The study is structured around a conceptual framework for linking social inequality, neighbourhood context and health, and draws upon online open data sources related to annually declared total primary health care consult fees and geographic and socio-demographic neighbourhood characteristics in the Netherlands. Results suggest that there are strong interactions between neighbourhood level green space and individual health-seeking behaviour, but also that these interactions are not stable over neighbourhoods, green space types, or gender. Parks, forests and fields seem to act as resources in shaping incidental health-seeking behaviour, while bodies of water, parks, orchards, forest, fields and farms often, and especially for females, seem to be stressors.

Keywords: Neighbourhood, Public health, Public space, Nature, Multilevel analysis, the Netherlands

Introduction

An increasingly large body of literature has suggested that neighbourhood context may affect adults' and young people's health and health-related behaviours (Meijer et al., 2012; Stock & Ellaway, 2013; Yen, Michael, & Perdue, 2009). It might be relevant to include neighbourhood characteristics in local primary healthcare demand and supply monitoring, because healthcare demand monitoring is currently problematic in the Netherlands because of a lack of population-wide indicators (Rijksoverheid, 2013). If neighbourhood characteristics can be found to partly explain geographical variations in the use of health services, then this insight is useful for local policy makers aiming to improve the balance between supply and demand or the convenience and access to primary health care services, or, eventually, for those designing health-promoting neighbourhoods. The relationship between *neighbourhood-level* characteristics and health and health-related behaviours of those within its circumference is especially relevant, because many health services are organized at this same level: several authors have emphasized that national health care systems should be organized in a way that, besides minimizing costs and maximizing the likelihood of a positive health outcome, maximizes convenience and accessibility of services (Hall, 2006; de Graaf-Ruizendaal & de Bakker, 2013).

One aspect of the local environment that varies between neighbourhoods is the amount and quality of natural green space that can be used by inhabitants of the neighbourhood to recreate or enjoy the view. Although the effect of nearby green space on individual health and the geographical distribution of health has been studied before (Björk et al., 2008; Maas, 2008; Phillips et al., 2001), it has been underrepresented when compared to other neighbourhood characteristics such as average neighbourhood income, socio-economic status or unemployment rate (Meijer et al., 2012; Pickett & Pearl, 2001; Yen, Michael, & Perdue, 2009), and results are inconclusive.

This study explores associations between the relative amount of green space present in the neighbourhood and individual primary health-seeking behaviour. In terms of theory, this study will explore new questions regarding the association between neighbourhood and health: whereas other studies have mainly approached the physical neighbourhood environment as one of several general determinants of neighbourhood quality of life and area deprivation (Blakely et al., 2006; Naess et al., 2007), this study will focus on direct associations between the physical environment and health. It is among the first to distinguish between different types and qualities of 'green space' and relate these distinct types to neighbourhood green space and health-seeking behaviour theory separately. Moreover, even though health-seeking behaviour is a particularly interesting health indicator because it is closely related to health demand monitoring, it has, to the author's knowledge, not been studied in relation to green space before. It does seems likely that there is a relation between green space and health-seeking behaviour, because associations between health and green space have been found for several others indicators of health, such as self-rated health, stress, mental health and chronic conditions (Maas, 2008; Mitchell & Popham, 2007).

The outcomes of this study will contribute to the exploration of the possibilities of using online open data to study the associations mentioned above. With the increasing availability of high-quality representative open data, 'whole population analytics' has the potential to make sense of unstructured, complex and interdependent aspects of our world that would otherwise remain invisible, such as the density of green space or individual total health-related costs over the entire country, and the relation between the two.

Background

Development of the field

The first studies that have researched the relationship between neighbourhood and health appeared in the early 1990's (Meijer et al., 2012; Pickett & Pearl, 2001). In the first ten years, most studies were ecological and non-multilevel (Meijer et al., 2012). Reviewers in 2001, who focused on the association between social and physical neighbourhood characteristics and mortality (Ellen, Mijanovich, & Dillman, 2001; Pickett & Pearl, 2001), concluded that neighbourhood characteristics were modestly associated with individual mortality. The biggest issues at that time were the delineation of the neighbourhood, how neighbourhood influences should be measured, and how confounding factors should be eliminated (Meijer et al., 2012). Later reviews still recognized these concerns, but benefited from the increased application of multilevel studies, which were able to partial out individual and neighbourhood effects and study interactions between the individual and the neighbourhood level. Reviews from 2007 (Riva, Gauvin, & Barnett), 2009 (Yen, Michael, & Perdue) and 2011 (Nandi & Kawachi) concluded that there was consistent evidence for associations between several neighbourhood characteristics and individual health.

Until then, economic and social neighbourhood characteristics such as neighbourhood socio-economic status (SES) and average income had been the most investigated neighbourhood characteristics, but around 2006 researchers started to study the association between green space as a neighbourhood characteristic and health (cf. Maas, 2008). The evidence for an association between green space and health is mixed. Studies in the UK and Sweden found no significant association between green space and self-rated health (Björk et al., 2008) or mental health (Phillips et al., 2001). Björk et al (2008) specifically studied green space within a distance of 100 and 300 meters, and found no significant association for either. In contrast, in a Dutch study, Maas (2008) did find a significant positive association between the amount of green space and self-rated health within a distance of 1 and 3 kilometres, for both urban and rural areas. She also found that this relationship was stronger for people with low SES compared to those with high SES, and for young people and old people when compared to adults between 25 and 64 years old. Similar results were found in England by Mitchell and Popham (2007), who, besides this, also found that the association between self-rated health and green space was mediated by income and the degree of urbanization as well: they found no association between green space and self-rated health in high income suburban and rural areas. Other support for a positive association between green space and health comes from Australia, where Sugiyama et al. (2008) found that people who see their neighbourhood as very green were respectively 1.4 and 1.6 times more likely to have better physical and mental health. Findings supporting a positive association between health and green space suggest that this association persists for different indicators of health, such as specific health conditions, stress, self-rated health and mental health (Maas, 2008; Mitchell & Popham, 2007; Sugiyama et al., 2008). Given that positive associations were found for different indicators of health, it can be expected that an association between green space and health-seeking behaviour, in terms of using local primary health services, can be found as well, even though health-seeking behaviour has not, to the author's knowledge, been studied in relation to neighbourhood green space before. A study by Vallée and Chauvin (2012) did use multilevel modelling to relate health-seeking behaviour to neighbourhood health service density and found that women living in neighbourhoods with low medical density had a higher risk of delayed health screening, but only if they reported that their daily activities were centred within their neighbourhood of residence.

Current issues in the field of neighbourhood characteristics and health include the study of the direction and mechanisms underlying the associations found (Bernard et al., 2007; Frohlich, 2013; Macintyre, Ellaway, & Cummins, 2002; Vogitländer et al., 2013), the question whether found associations persist for different health indicators, different neighbourhoods and different cultures (Maas, 2008) and whether green space characteristics interact with other neighbourhood or individual characteristics (Mitchell & Popham, 2007), and the question how these new insights can be used to design health-promoting neighbourhoods and local health services adapted to the characteristics of the neighbourhood (de Graaf-Ruizendaal & de Bakker, 2013). Several authors highlight the potential of using new tools and methods, such as GIS (Schippereijn, Ejstrud, & Troelsen, 2013) and GPS (Maas et al., 2013).

Informing analysis: neighbourhood green space and health-seeking behaviour theory

It is generally accepted within the field of neighbourhood characteristics and individual health, that neighbourhood characteristics, and the position of the individual within the neighbourhood's structure of social inequality, can shape the stressors to which its inhabitants are exposed, and the resources available to handle these stressors (Bernard et al., 2007; Giddens, 1984; Macintyre, Ellaway, & Cummins, 2002). Compared to Frohlich's ISIS framework (2013), which explains spatial variations in health using individual economic, social, biological and cultural capital stock to describe how individuals identify, access and utilise resources in neighbourhoods to their health advantage, and Bernard's rules of access within the neighbourhood's configuration of resources (2007), a framework developed by Voigtländer and colleagues is especially insightful within the context of this study because it clearly highlights that individual health is the result of multi-level interactions. Moreover, the model is useful in this study because psychological processes and behaviour at the individual level, such as health seeking, are included in the model.

Voigtländer et al. (2013) identify interactions between the individual, the neighbourhood and the social inequalities existing within society that, together, describe how the neighbourhood influences health (Figure 1). They define neighbourhood as 'the structure of the social ties of residents in an area who live in proximity to each other and who – to some extent – use the same facilities or participate in the same organisations'. This definition emphasizes that neighbourhoods have a strong relational component, and that there is at least some interaction between an individual and his or her physical, social and institutional context.

The neighbourhood is referred to as the 'meso level' in the conceptual framework introduced by Voigtländer and colleagues; individual health is the product of internalisation of stressors and resources at the micro level; and the macro level consists of social inequalities within society, such as differences in sex, ethnicity, income and choice of residential location.

The physical environment is one of four categories of stressors and resources present at the meso level, besides markets, institutions and social capital. Green space is part of the physical environment. Next to these stressors and resources, the neighbourhood context is further characterized by its sociodemographic composition in terms of, for example, age, education, income and ethnicity and the changes that occur within this composition. Resources and stressors and the sociodemographic composition influence each other.

The interaction between the macro and meso level is embodied in the individual's choice of residential location and the neighbourhood composition that follows from the collective of these decisions. The meso level influences the micro level in three ways: a direct pathway from (dangerous) neighbourhood environments to direct health effects, such as a child falling from a tree in a forest nearby; an indirect pathway from neighbourhood environments, through individual perception and assessment, to individual health, such as enjoying the de-stressing effect of walking in the local park; and through personal resources which are the result of individual health and social position, and which lead via the indirect path to health as well. Finally, the micro level influences the macro level through the collective health statuses of individuals, constituting a part of social inequality.

Using this model to frame the association between neighbourhood green space and individual health-seeking behaviour, we can understand health-seeking behaviour as an iterative part of the model's 'behaviour'-block: health-seeking behaviour is the result of macro and meso processes as well as of our individual health status and our perception and assessment of this status. Keeping to this understanding of health-seeking behaviour and neighbourhood characteristics, we study to what extend individual health-seeking behaviour is attributable to neighbourhood characteristics, and to various types of neighbourhood green space in particular.

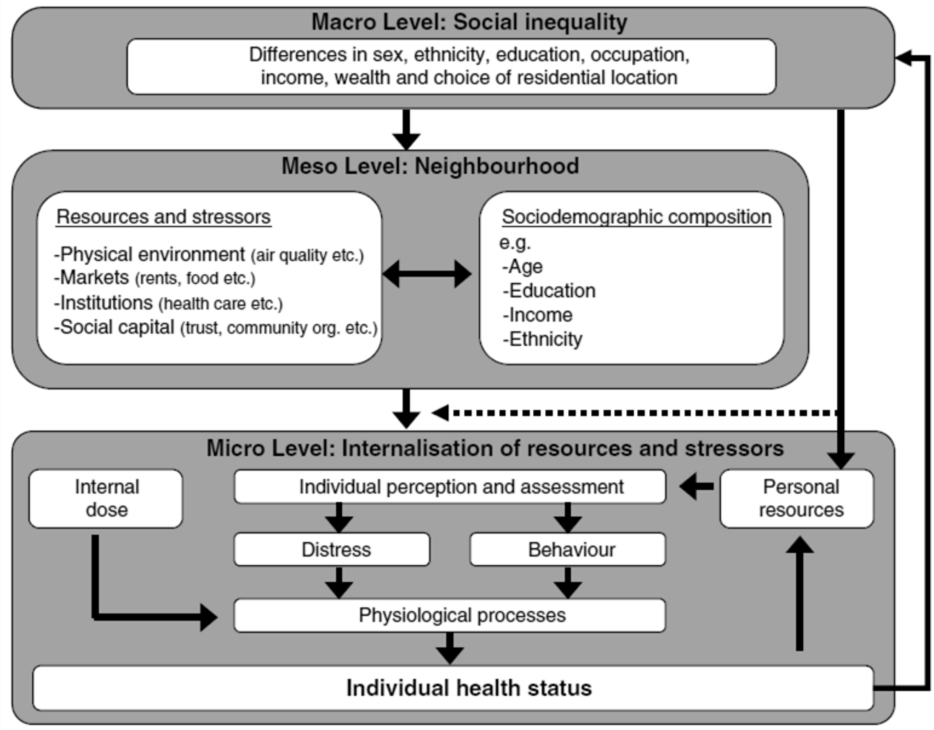

Figure 1. Conceptual framework linking social inequality, neighbourhood context and health (Voigtländer et al., 2013)

Method

Dataset

This research draws upon data from several online open data sources that relate to health-seeking behaviour in the Netherlands. A multilevel dataset was created by linking neighbourhood characteristics data with individual data, at the 3-digit zip code level. In the Netherlands, there are 799 3-digit zip codes, with a mean population of 20,629. A higher level of aggregation, such as municipalities, would have decreased homogeneity within the neighbourhood, while aggregation at a lower level would have resulted in many neighbourhoods with less than 100 people. Neighbourhood data and health care data are not available at a lower level due to privacy concerns.

Dependent variable

Health-seeking behaviour data was collected from Vektis' annual health insurance database 2012 (Vektis, 2015). Vektis collects this data as part of the health insurance act, which forces all Dutch health insurance companies to report declared healthcare costs. The dataset contains all those costs that were realized in 2012, that are part of basic health insurance, and that were declared by the health care provider and the insured between January 1st 2012 and January 1st 2014, for several health care categories. Provided data is extensively checked by Vektis, who claims that between 98% and 100% of all realized costs were declared. One row in the dataset represents all individuals with the same age, gender and 3-digit zip code, and contains these variables plus the number of insured individuals (of which the total, in the Netherlands, ought to be equal to the total population) weighted for their registration period in 2012 (to control for births, deaths and removals; referred to as 'insurance years'), and total declared costs per category. Due to privacy concerns, rows consisting of less than 10 insurance years were excluded from the dataset. In this research, health-seeking behaviour was operationalized using average general practitioner (GP) consultation fees per gender, age and zip code, calculated by dividing total GP consultation fees by the number of insurance years, under the assumption that individuals who seek health always first consult their GP. Average GP consultation fees are available for 16,482,606 individuals, which is over 98% of the Dutch population in 2012. The logarithm of average GP consultation fees was used for modelling because the distribution of the log-value was more normal and insightful that that of the original, which was highly skewed. The logarithms of average GP consultation fees will hereafter be referred to as 'costs'.

Explanatory variables

Level 1 variables (Individual level)

Individual-level variables were included in the analysis in two ways: to adjust for factors that may act as confounders in the primary association, and as factors that may modify, or interact with, that association. The identification of individual level characteristics – gender and age – was based on the Vektis dataset. A dummy variable was created, 'male', in which males were coded 1 and females 0. Age ranged between 0 and 90, where a value of 90 represented all those aged 90 and up. Besides 'age' and 'male', the deviation of the zip code mean age and gender was calculated for each individual to represent the individual's distance from the neighbourhood's average.

Level 2 variables (Neighbourhood level)

Within the neighbourhood level, in line with Voigtländer et al.'s model, we distinguish between green space determinants and sociodemographic composition. Green space variables were collected from Open Street Map (OSM), an openly licensed online world map created by volunteers using local knowledge, GPS tracks and donated sources (Open Street Map, 2015). A complete land-use map of the Netherlands was downloaded on August 9th 2015. This map contained both green and non-green space, but every shape on the map carried a type-label indicating what the area was used for. Using qGIS software version 2.10 Pisa (2014), green area was filtered by means of this label, so that only areas related to seven types of green space remained: fields, farming, leisure, orchards, forests, water and parks. Also using qGIS, intersections between the remaining areas and 3-digit zip code areas, obtained from Imergis (2015), were calculated, and the intersected green areas were grouped on 3-digit zip code and landuse type. For each zip-code, the percentage of the zip-code covered by the different types of green space was calculated.

Variables related to sociodemographic composition were collected from Statistics Netherlands (CBS, 2015), which, among other things, publishes reliable figures on yearly national demographics. Average household size and composition, absolute amount of (non-)Western immigrants and average household income for January 1st 2013 were presented at 4-digit zip code level and aggregated to 3-digit zip code level by summing (for absolute values) or calculating weighted averages (for relative values). Three variables describe the level of education within 4-digit zip codes: the percentage of inhabitants whose highest degree is in primary, secondary or tertiary education. These percentages, which were based on a census on September 30th 2011, were aggregated to the 3-digit level as well. Finally, data on neighbourhood-level socio-economic status (SES) was requested from EDM BV, who collected this data for the Netherlands Institute for Social Research (2015). This is the only data source used that required requesting, although the data is immediately made available to anyone who enlists an email address. Neighbourhood SES-score was derived from local levels of education, income, and labour position. Again, these values were aggregated from 4-digit to 3-digit zip code level by means of a weighted average.

As was mentioned earlier, mean age and gender were calculated for each 3-digit zip code to represent other aspects of the sociodemographic composition of the neighbourhood.

Level 3 variables (Province)

For each 3-digit zip code, a GIS-layer containing shapefiles of the 12 Dutch provinces, created by Kadaster and last updated February 26th 2015 (Imergis, 2015), was used to determine to which province it belongs. The variable that this resulted in was used as a third level variable. This third level was included to crudely control for the assumption that neighbourhoods in the same region resemble each other more than neighbourhoods in different regions.

Analysis

All analyses were conducted with Stata 12.1 (2011). Descriptive statistics were generated at both the individual level and the neighbourhood level by using insurance years to account for weighting of the dataset. Since this study is interested in neighbourhood effects over and above individual characteristics, a random intercept and random slope multilevel analysis was carried out (using robust standard errors), based on 16,482,606 individuals (level 1) nested within 799 neighbourhoods (level 2), within 12 provinces (level 3). We analysed the following four models:

Model 1: 3-level empty model of individuals nested within neighbourhoods nested within provinces, with no variables in the fixed and the random parts of the model, to determine what part of the total variance can be attributed to the neighbourhood level.
Model 2: Model including the same as above, but with fixed individual-level effects included; neighbourhood-level green space variables included as main effects and sociodemographic composition variables included as confounders.
Model 3: Model including the same as above, but with cross-level interactions between individual gender and neighbourhood green space measures, to test whether effects are modified by gender.
Model 4: Model including the same as above, but with random effects included, to test whether found effects are stable over neighbourhoods.

Maximum likelihood estimation was used to determine parameter estimates. Therefore, deviance, based on -2 log likelihoods and used to determine goodness of fit, can be applied to entire models. In the models discussed below, all control variables were included, and green space fixed and random effects were included stepwise if they improved model fit. Robust standard errors were used to allow for the fact that, in all likelihood, errors are both independent and normally distributed. Covariances are assumed to be unstructured.

Results

Base model

Exploration of the dependent variable, costs, and its distribution (Figure 2) reveals that the data seems to exist of three partly overlapping normal distributions, one centring around 2, the second around 4.5 and the third around 6. Ideally, if all influential variables are included, the modelling process should result in normally distributed residuals, so that these three peaks get dissolved. Model 5 estimates are presented in table 1, models 1 to 4 are omitted. According to this model, males and females do not differ significantly in terms of costs (B=0.01) (p>0.1). An individual's age's difference to the neighbourhood mean age is significantly associated to costs, even though neighbourhood mean age is not. Neither are population density and average household income. The relative amount of tertiary degree owners (B=-0.02) (p<0.05) and the relative amount of non-Western immigrants (B=0.02) (p<0.05) are the only significant sociodemographic composition variables; both are negatively associated with costs. In terms of green space, after allowing slopes to vary over neighbourhoods, only the relative area of fields remains statistically significant (B<0.01) (p<0.05). According to this model, people living in neighbourhoods with relatively large areas of fields declared higher GP consult fees than those who live in neighbourhoods with small areas of fields, even when controlling for population density. When allowing slopes to vary over different neighbourhoods, the association between fields and costs is no longer modified by gender.

The three peaks that were visible in the histogram of costs have not been dissolved in this model: they are still clearly visible on the histogram plot of the residuals (Figure 3).

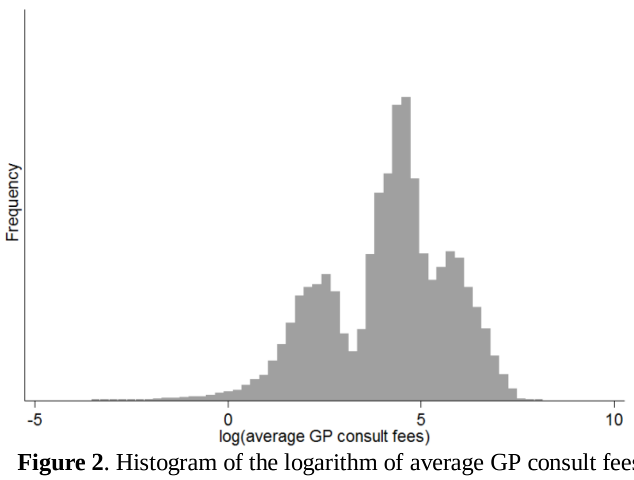

Figure 2. Histogram of the logarithm of average GP consult fees

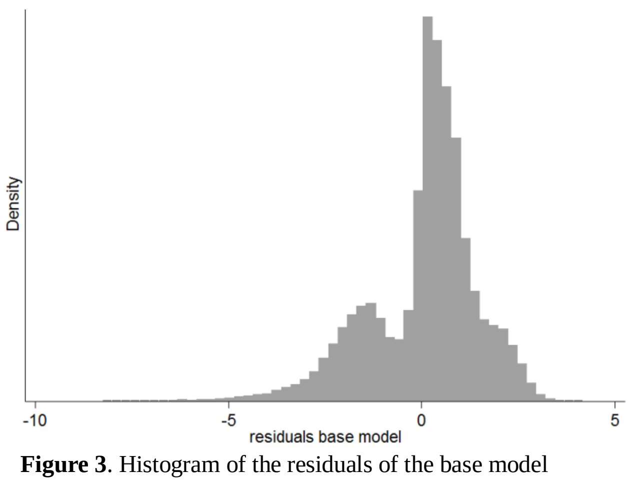

Figure 3. Histogram of the residuals of the base model

Tabel 1

Fixed effect (top) and variance (bottom) estimates for cost models

		Model 5
Estimates of fixed effects		
Level 1 (individual)		
	intercept	2.663 (1.040)**
	male	0.014 (0.042)
	sq. dev. of mean age	0.000 (0.000)**
	dev. of mean age	0.007 (0.001) **
Level 2 (neighbourhood)		
	mean age	0.026 (0.017)
	density	-36.568 (29.030)
	tertiary education	-0.015 (0.006)**
	income	0.000 (0.000)
	immigrants	-0.020 (0.005)**
	leisure	-0.025 (0.017)
	forests	0.003 (0.002)
	fields	0.002 (0.001)**
	male X fields	0.237 (1.185)
Variance of random effects		
Level 3 (province)		
	intercept	0.015 (0.015)**
Level 2 (neighbourhood)		
	intercept	0.294 (0.028)**
	forests	0.000 (0.000)**
	fields	0.000 (0.000)**
	male X fields	2640.102 (7040.825)**
Level 1 (individual)		
	residual	1.047 (0.088)

Note. * p<0.10 ** p<0.05. Standard errors are in parentheses. Intraclass Correlation Coëfficient (ICC) at province level are 0.007, 0.108 and 0.049 respectively; ICCs at neighbourhood level are 0.243, 0.628 and 0.572 respectively.

Discussion of base model results

The results of this model are surprising. The fact that there does not seem to be a difference in costs between males and females, between low and high income neighbourhoods and between densely and scarcely populated areas counters earlier findings (cf. Meijer et al., 2012; Pickett & Pearl, 2001; Yen, Michael & Perdue, 2009). Moreover, the distribution of the residuals leads us to conclude that declared GP consult fees are further associated to external factors; so that there are mechanisms that determine to which of the three normal distributions an individual contributes.

To get more insight into the mechanisms within these three normal distributions and the differences between them, the original data was divided into three clusters, each corresponding to one of the peaks, and separate models were built for the clusters. When plotting costs against age, we see a clear division between the three clusters (Figure 4). This scatterplot was used to determine the thresholds for each of the three clusters, and we find that the first normal distribution can be described as those y-values that lie below 2.8+0.0175*age, while the third normal distribution can be modelled as lying above 5+0.016*age (Figure 4). The three resulting y-variables will hereafter be referred to as 'low cost cluster', 'middle cost cluster' and 'high cost cluster'.

Table 2 summarizes the mean values of individual and neighbourhood-level variables for the three clusters. The low cost cluster is characterized with especially high average ages (42.950), high degrees of tertiary education degree holders (19.434) and non-Western immigrants (16.297), high population density (0.003) inhabitants per square meter), and high percentages of area covered by leisure-areas (0.802) and parks (2.602). The high cost cluster, representing 4.45% of the sample population, is characterized by low mean age (24.862), large areas covered by fields (38.366) and farms (20.549), low density (<0.001), high average household income (35820.42), very low proportions of non-Western immigrants (2.548), and is largely represented by males (0.669). Moreover, the high cost cluster's variable means have standard errors that are multitudes of those of the other two clusters. The middle cost cluster tends to have means that lie between that of the other two clusters, except that it has high percentages of forests (9.462) and a slightly higher neighbourhood mean age (43.982). Only the area of orchards did not differ significantly over the three clusters.

New models were built up for each of the three cost clusters, following the methodology described earlier.

Cluster models

Table 3 presents the results from the successive multilevel models. All three Model 1's show that while a large share of the variation in costs is attributable to the individual level, there is considerable variation at the neighbourhood level as well. Variation at the province level is smaller. These estimates yield intra class correlations, or the portion of the total variance that occurs between neighbourhoods, of 0.243, 0.627 and 0.572 at neighbourhood level, and 0.007, 0.107 and 0.049 at province level. The big differences between clusters are striking, especially at province level.

In the second model, which shows the association between neighbourhood characteristics and costs while

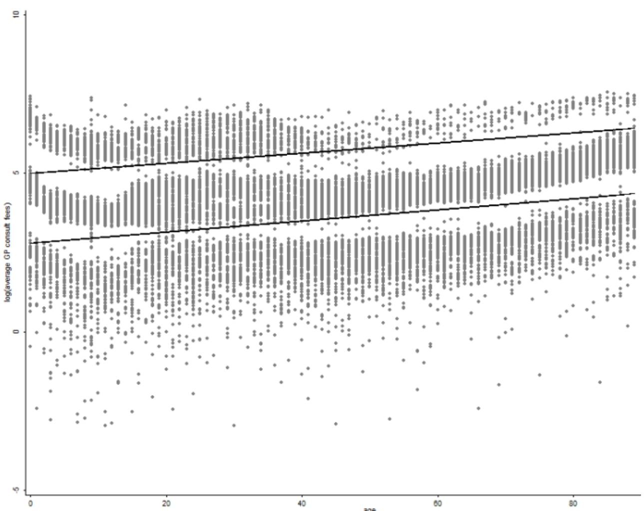

Figure 4. Scatterplot of costs against age, with cluster thresholds

Table 2
Summary of variables included in models, weighted by insurance years

	Low cost cluster		Middle cost cluster		High cost cluster		
N	6.156.605 (37.352%)		9.593.097 (58.201%)		732.903 (4.447%)		
	Mean	SE	Mean	SE	Mean	SE	p
forest	8.04	0.003	9.46	0.003	7.90	0.011	< 0.001
field	21.39	0.006	25.68	0.005	38.37	0.025	< 0.001
park	2.60	0.002	1.51	0.001	0.55	0.005	< 0.001
leisure	0.80	0.000	0.54	0.000	0.26	0.002	< 0.001
farm	9.30	0.004	13.46	0.004	20.55	0.020	< 0.001
orchard	0.73	0.000	1.04	0.000	1.31	0.003	0.539
water	9.05	0.005	9.09	0.004	9.53	0.020	< 0.001
household income	32673.67	1.957	34478.99	1.518	35820.42	6.012	< 0.001
age	42.95	0.009	39.46	0.007	24.86	0.022	< 0.001
male	0.40	0.000	0.54	0.000	0.67	0.001	< 0.001
SES	-0.22	0.000	0.07	0.000	0.18	0.001	< 0.001
tertiary education	19.43	0.004	18.09	0.002	12.95	0.008	< 0.001
population density	0.003	0.000	0.001	0.000	0.000	0.000	< 0.001
non-Western immigrants	16.30	0.006	9.41	0.003	2.55	0.004	< 0.001
mean age	43.91	0.000	43.98	0.000	42.10	0.002	< 0.001
logconsult	2.09	0.005	4.55	0.002	6.23	0.003	< 0.001

adjusting for sociodemographic composition and individual characteristics, we start to see differences between the three clusters**.** Because SES is calculated based on income and education levels, and because including average income and tertiary degree holders improved each model's fit more than including average SES did, SES-scores were excluded from the models. The three cost clusters are similar in that for each, individual level variables are significantly associated with costs: males and adults people have lower costs than women and very young and older people ($p<0.05$). For the low cost cluster, living in neighbourhoods with a high mean age ($B=0.026$; $p<0.05$), high percentage of non-Western immigrants ($B=0.009$; $p<0.05$) and a low percentage of area covered by parks ($B=-0.007$, $p<0.10$), forests ($B=-0.005$; $p<0.05$) and fields ($B=-0.003$; $p<0.05$) was significantly related to higher costs. No significant effects were found for population density, the percentage of inhabitants with a tertiary education degree, average household income, and for other types of land use.

For the middle cost cluster, a positive association between costs and neighbourhood characteristics was found for the percentage of the area covered by parks ($B=0.009$, $p<0.10$), orchards (0.011, $p<0.05$), forests ($B=0.009$, $p<0.05$), fields ($B=0.009$, $p<0.05$) and farms ($B=0.006$, $p<0.05$). A negative association was found for the percentage of inhabitants owning a tertiary education degree ($B=-0.010$, $p<0.10$) and non-Western immigrants ($B=-0.013$, $p<0.05$). No significant effect was found for mean age, population density, average household income and the percentage covered in leisure areas.

In terms of the found effects, the middle cost cluster resembles the high cost cluster. Again, living in a neighbourhood with a high percentage of inhabitants with a tertiary education degree ($B=-0.005$, $p<0.05$) or from a non-Western background ($B=-0.005$, $p<0.10$), and with a high percentage of water ($B=0.005$, $p<0.05$), parks ($B=0.009$, $p<0.05$), orchards ($B=0.008$, $p<0.05$), forests ($B=0.007$, $p<0.05$), fields ($B=0.006$, $p<0.05$) or farms ($B=0.006$, $p<0.05$) was significantly related to lower costs. However, within the high cost cluster, average household income ($B=0.000$, $p<0.05$) and population density ($B=45.254$, $p<0.05$) are positively and significantly associated with costs as well. For both the middle and the high cost clusters, no significant effect was found for mean age and for the percentage of the area filled with leisure facilities.

In the third model, interactions were added between gender and green space variables. This did not result in major changes in any of the previously found effects for the low cost cluster. The estimate of the effect of density increased from 2.579 to 3.388, but neither of these were significantly different from 0. A small effect for the interaction between gender and the percentage of the area covered in park did significantly improve the model's goodness of fit, but this effect was not significant either ($p>0.10$). The model does indicate that the negative effect of the percentage of the neighbourhood covered in forests on costs is weaker for males than for females ($B=0.003$, $p<0.05$). The same goes for the percentage of area covered by fields ($B=0.009$, $p<0.10$), indicating a positive association for males and a negative association for females. More differences between models 2 and 3 exist for the middle cost cluster. In model 3, population density ($B=26.676$, $p<0.05$) and average household income ($B<0.001$, $p<0.10$) do have a significant positive effect on costs where they previously had not. Significant effect modifiers for gender were found for all green space variables except the percentage of area covered by fields. The effect of water, orchards and farms are stronger for females, while the effect of parks, forests and fields are stronger for males. In case of the high cost group, the inclusion of effect modifiers again does not change the strength and direction of previously found effects. Similar to the middle cost cluster, significant differences between men and women were found for all types of green space except fields. The effects of the percentage of water and forest are stronger for females than males; the others are stronger for females.

The fourth model shows that most effects found persist when adding random slopes for the green space variables.

Table 3
Fixed effect estimates (top) and variance estimates (bottom) for models for consultcosts

	Low costs group				Middle costs group				High costs group			
Parameters	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Estimates of fixed effects												
Level 1 (individual)												
Intercept	1.197 (0.022)**	1.248 (0.530)**	1.250 (0.532)**	1.055 (0.543)*	4.606 (0.061)**	3.704 (0.380)**	3.697 (0.384)**	3.401 (0.232)**	5.854 (0.034)**	5.166 (0.358)**	5.167 (0.355)**	4.801 (0.317)**
male		-0.554 (0.016)**	-0.600 (0.017)**	-0.600 (0.017)**		-0.226 (0.005)**	-0.221 (0.007)**	0.122 (0.030)**		-0.168 (0.010)**	-0.174 (0.014)**	-0.118 (0.017)**
sq. dev. of mean age		0.000 (0.000)**	0.000 (0.000)**	0.000 (0.000)**		0.000 (0.000)**	0.000 (0.000)**	0.000 (0.000)**		0.000 (0.000)**	0.000 (0.000)**	0.000 (0.000)**
dev. of mean age		0.016 (0.000)**	0.016 (0.000)**	0.016 (0.000)**		0.011 (0.00)**	0.011 (0.000)**	0.011 (0.000)**		0.011 (0.00)**	0.011 (0.00)**	0.012 (0.00)**
Level 2 (neighbourhood)												
mean age		0.026 (0.011)**	0.026 (0.011)**	0.029 (0.011)**		0.006 (0.009)	0.006 (0.009)	0.015 (0.005)**		0.004 (0.008)	0.004 (0.008)	0.007 (0.007)
population density		2.579 (9.257)	3.388 (9.155)	6.827 (8.877)		26.897 (17.331)	26.676 (17.470)	16.881 (9.608)*		45.254 (16.775)**	45.360 (16.821)**	59.689 (16.213)**
tertiary education		0.001 (0.002)	0.001 (0.002)	0.001 (0002)		-0.010 (0.001)**	-0.010 (0.001)**	-0.007 (0.002)**		-0.005 (0.001)**	-0.005 (0.001)**	-0.007 (0.003)**
income		-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)		0.000 (0.000)	0.000 (0.000)	0.000 (0.000)*		0.000 (0.000)**	0.000 (0.000)**	0.000 (0.000)**
immigrants		0.009 (0.002)**	0.009 (0.002)**	0.009 (0.002)**		-0.013 (0.003)**	-0.013 (0.003)**	-0.011 (0.002)**		-0.005 (0.003)*	-0.005 (0.003)*	-0.007 (0.003)**
water						0.006 (0.003)**	0.06 (0.003)**	0.005 (0.000)**		0.005 (0.002)**	0.006 (0.002)**	0.007 (0.002)**
male X water							-0.000 (0.000)	-0.01 (0.01)**			-0.000 (0.000)	0.003 (0.003)
parks		-0.007 (0.004)*	-0.009 (0.003)**	-0.010 (0.002)**		0.009 (0.005)*	0.008 (0.006)	0.002 (0.003)		0.009 (0.004)**	0.008 (0.004)**	0.008 (0.004)**
male X parks			0.005 (0.004)	0.005 (0.004)			0.003 (0.001)**	-0.029 (0.008)**			0.000 (0.000)	0.002 (0.003)
orchards						0.011 (0.005)**	0.011 (0.005)**	0.010 (0.004)**		0.008 (0.004)**	0.006 (0.005)	0.010 (0.005)**
male X orchards							-0.002 (0.001)*	-0.013 (0.004)**			0.002 (0.003)	0.013 (0.006)**
forests		-0.005 (0.002)**	-0.007 (0.002)**	-0.006 (0.002)**		0.009 (0.003)**	0.009 (0.003)**	0.004 (0.001)**		0.007 (0.002)**	0.007 (0.002)**	0.009 (0.002)**
male X forests			0.003 (0.002)**	0.003 (0.002)**			0.000 (0.000)	-0.008 (0.002)**			-0.000 (0.000)**	0.002 (0.001)
leisure												
male X leisure												
fields		-0.003 (0.001)**	-0.003 (0.001)**	-0.003 (0.001)**		0.009 (0.002)**	0.009 (0.002)**	0.006 (0.000)**		0.006 (0.002)**	0.006 (0.002)**	0.007 (0.001)**
male X fields			0.009 (0.005)*	0.009 (0.005)*			0.000 (.)	0.000 (.)			0.000 (.)	0.000 (.)
farmland						0.006 (0.002)**	0.007 (0.002)**	0.006 (0.000)**		0.006 (0.002)**	0.006 (0.002)**	0.007 (0.001)**
male X farmland							-0.00 (0.00)**	-0.005 (0.001)**			0.000 (0.000)*	-0.000 (0.000)
Variance of random effects												
Level 3 (Province)												
Intercept	0.004 (0.002)**	0.006 (0.004)**	0.005 (0.004)**	0.004 (0.003)**	0.041 (0.019)**	0.008 (0.004)**	0.008 (0.004)**	0.002 (0.002)**	0.012 (0.006)**	0.002 (0.002)**	0.002 (0.002)**	0.007 (0.004)
Level 2 (Neighbourhood)												
Intercept	0.129 (0.007)**	0.122 (0.021)**	0.122 (0.021)**	0.101 (0.019)**	0.198 (0.010)**	0.090 (0.011)**	0.090 (0.011)**	0.061 (0.003)**	0.123 (0.007)**	0.062 (0.005)**	0.062 (0.005)**	0.092 (0.006)**
water												0.001 (0.000)**
male X water								0.000 (0.000)**				0.001 (0.001)**
male X park								0.005 (0.001)**				
male X orchard								0.000 (0.000)**				0.000 (0.000)**
forest				0.000 (0.000)**								
male X forest								0.000 (0.000)**				0.000 (0.000)**
male X farm								0.001 (0.001)**				0.000 (0.000)**
Level 1(individual)												
var(residual)	0.413 (0.000)**	0.265 (0.033)**	0.265 (0.033)**	0.265 (0.033)**	0.142 (0.000)**	0.039 (0.001)**	0.039 (0.001)**	0.037 (0.000)**	0.101 (0.000)**	0.063 (0.002)**	0.063 (0.002)*	0.060 (0.000)**

Note. * p<0.10 ** p<0.05. Standard errors are in parentheses. Intraclass Correlation Coëfficient (ICC) at province level are 0.007, 0.108 and 0.049 respectively; ICCs at neighbourhood level are 0.243, 0.628 and 0.572 respectively.

For the low cost cluster, the model was improved by allowing the slope of the percentage of area covered by forests to vary over different neighbourhoods. The variance within these slopes is smaller than 0.001 ($p<0.05$). This did not have a big effect on the strength or direction of the main effect, or any of the other effects found. The middle cost cluster model was improved by sequentially adding random slopes for the interaction effects between gender and water (var<0.001), parks (var=0.001), forests (var<0.001) and farmland (var=0.001), and for the effect of the percentage of area used as orchards (var<0.00). The p-value for all of these variances was smaller than 0.05. Allowing slopes to vary causes the main effect for the percentage of park to decrease so that it is not significantly different from 0. Also, the interaction effect between gender and forest, which used to have a positive coefficient, is now negative and stronger than the main effect ($B=-0.008$, $p<0.05$), indicating that the effect of the percentage of forests on costs is positive for females, but negative for males, when controlling for variations in the main effect between neighbourhoods. Similarly, allowing slopes to vary between neighbourhoods has changed the main ($B=0.006$, $p<0.05$) and interaction effect ($B=-0.001$, $p<0.05$) for the percentage of farmland, so that now there seems to be a positive association between costs and the percentage of farms for females, but no association or a very weak association for males. For the percentage of land used as orchards, there seems to be a positive association for women, and a negative association for men.

For the high cost cluster, finally, random slopes were added for the percentage of area covered by water and orchards, and for the interaction effect between gender and water, parks, forests and farms. These random slopes improved the fit of the model, but changed all interaction effects found in model 3. None of them are significant in this model, except for the interaction effect between gender and orchards. This effect indicates that the positive association between costs and orchards is over twice as strong for men as it is for women.

Discussion

Three cost clusters

There are several potential external factors that might explain the differences between the three clusters. These external factors can be related to demand and supply of health services, or to methodological choices made in the data collection process.

First, the big differences between the clusters' mean population density leads us to assume that a neighbourhood's population density might play a role in determining to which of the three clusters its inhabitants contribute. Plotting average GP consult fees over age while distinguishing between seven groups of increasing density shows that high densities are especially apparent in the low and middle cost clusters (figure 5), so that it is likely that population density, at least partly, influences cluster distribution. However, individuals from the same zip code can be part of different cost clusters, so it is unlikely that this theory fully explains the pattern. One way in which population density might cause a differentiation between high, middle and low cost clusters is that urban GPs tend to be busier and more likely to refer patients to medical specialists, while rural GPs tend to perform simple specialist procedures themselves. Another explanation is that rural individuals tend to live further away from their GP and might have a different mind-set when it comes to health-seeking, so that they wait until their demand is more urgent (and expensive) before visiting their GP, compared to urban individuals.

Besides differences in population density, the difference between low, middle and high cost clusters could lie in structural differences between the nature of health-seekers' complaints for the three clusters, hereafter referred to as the specific-complaint-explanation. Plotting average GP consultation fees over age (figure 5), we see that costs in the low cost cluster increase only slightly with age, while costs in the middle cost cluster increase exponentially, with another peak for young children. A possible explanation of these shapes is that the lowest group consists of generally

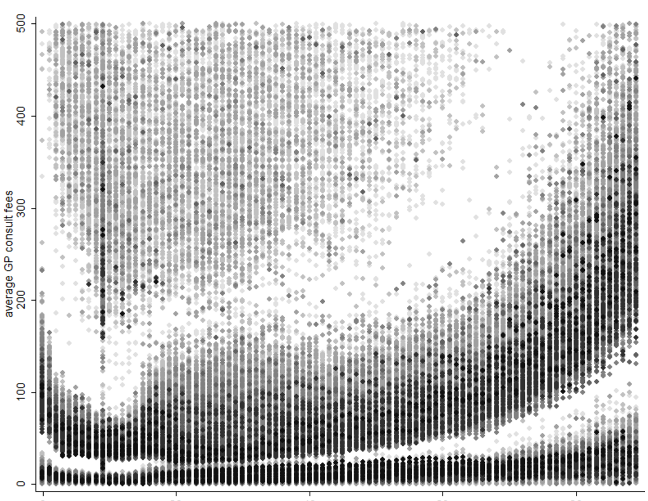

Figure 5. Average GP consult fees against age, by population density

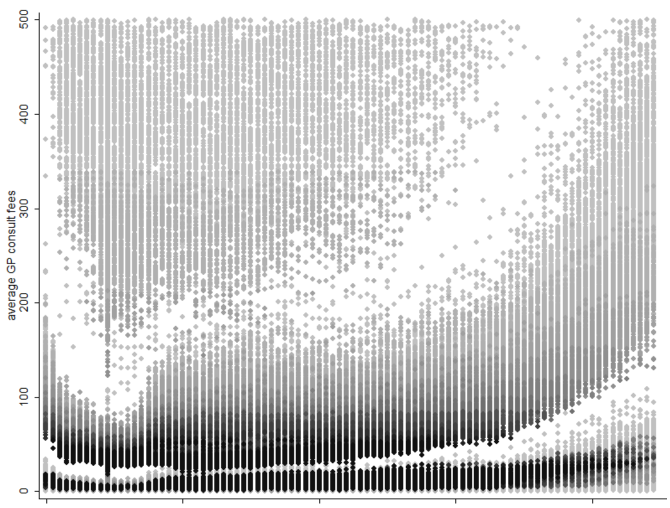

Figure 6. Average GP consult fees against age, by row weight

healthy individuals, who only visit their GP incidentally, while the second group consists of patients with structural health problems. The third group, with its striking differences in gender, peaks at a younger age and might be related to fertility-related health-seeking behaviour.

A third external factor that might determine to which group an individual contributes is the costs GPs declare after a consult. Consults where the patient visits the practice are cheapest (€9,04 if the consult is shorter than 20 minutes), followed by consults where the GP visits the patient, which tends to occur more often as the patient gets older (€13,56 for the same consult), again followed by consults by patients not registered at the practice they visit, which tends to occur more often with students around the age of 20-25 (€27,19 for the same consult). The low cost cluster might consist of regular consults, the middle cost cluster of consults where the doctor visits the patient, and the high cost cluster might consist of people who are not registered at the GP practice they visit.

A weakness of each of these theories is that the values in the data do not belong to individuals, but to groups of individuals who share some characteristics. Cost values are not individual, but group averages. This methodological choice leads to a loss of information at the individual level, and the varying weight of each row in the data distorts the pattern. Plotting costs over age, distinguishing between row weights in steps of thirty (figure 6), we see that the high cost cluster exists solemnly of low weights, so that extreme individual values have a large effect on the average. For the middle and low cost clusters, we see a similar effect, where the exponential increase in average costs over age in the middle cluster might be due to the fact that weight logically decreases as more individuals from the same group decease. The ‘clean’ gradient in this figure leads us to assume that this theory is very likely to at least partly explain the pattern.

Not all theories mentioned above can be tested with the data available. To test whether health service supply-side differences explain the pattern, in addition to the dataset used here, information about the particular GP visited by individuals, such as the density of the direct environment of the practice, the procedures done by the GP, and the nature of the consult (normal, GP visit, passing patient), are necessary. To test health demand-side differences, additional information on individuals’ complaints, their distance to the nearest GP, or even their tendency to visit a GP for minor complaints, are needed. Further research on the differences between the cost clusters is required to test what causes those differences. It would be particularly interesting to design a model that, on the one hand, tests to which cost cluster an individual contributes (using variables such as insurance years, which have little value in estimating actual health-seeking costs), while on the other hand it estimates the individual’s costs, using independent predictors. However, for the remainder of this study, the methodological explanation and the specific-complaint-explanation will be assumed to, together, determine the shape of the data, because these can explain the behaviour of average consultation fees over age, and because the specific-complaint-explanation can, sensibly, be thought of to be related to neighbourhood green space, because different environments might cause different complaints.

Model findings

Individual health-seeking behaviour is partly attributable to neighbourhood characteristics. The results of this study indicate that 24.3%, 62.7% and 57.2% of the variance within the three clusters is accounted for by neighbourhoods, rather than by individuals. This is large compared to other findings, which indicated ICC values of 0.14 in New Zealand among neighbourhoods with a mean population of 2000, where self-reported health was studied (Aminzadeh, et al., 2013), and 0.035 in a Dutch study on neighbourhood social capital and self-rated health in 4-digit zip code neighbourhoods (Mohnen et al., 2011). In terms of Voigtländer et al.’s framework, this finding indicates that although the micro level is strong, meso level influences are of considerable importance in the determination of average GP consult fees, even when ‘neighbourhood’ is considered to be large enough to contain several thousand inhabitants.

We further found that distinguishing between different types of green space is useful in determining how green space influences individual health. Especially the nearby presence of large areas of parks, forests, fields and orchards seems to be associated to individual health-seeking behaviour. The associations between parks and health-seeking behaviour for the different cost clusters seems to be similar to the one between forests and health-seeking behaviour, suggesting that individuals interact with forests and parks in the same way (e.g. strolling or enjoying the view it provides), benefitting or suffering from its effects in much the same way as well. The same can be said for fields and farms in rural areas. However, this might also indicate that OSM map builders use ‘field’ and ‘farmland’ as interchangeable labels.

This study does not provide reason to assume that green space is either a resource or a stressor, because results reveal that neighbourhood green space can be both positively and negatively associated with health-seeking behaviour, dependent on population and gender. Assuming that the differences between the three cost clusters are at least partially caused by the specific-complaint-explanation, where the low cost cluster consists of incidental health-seeking behaviour and the middle cost cluster of structural health-seeking behaviour, results suggest that neighbourhood green space is, in general, negatively associated to the costs of the former and positively associated to the costs of the latter, though more so for women than for men. For future research, it would be interesting to study whether this is a causal relation, implying that green space is a resource for incidental health-seekers – e.g. due to the indirect destressing effect of being able to habitually walk in a park or forest – and a stressor for structural health-seekers – e.g. due to a direct effect of being prone to jogging injuries. One potential explanation for the difference between men and women is that they might be different in their ways of enjoying green space, so that what is a resource for one may be a stressor for the other. Another explanation might be that one is more sensitive to certain contextual influences at the meso level than the other. Further research is required to test these theories. The finding that, besides income and urbanization (Mitchell & Popham, 2007) and SES and age (Maas, 2008), gender also modifies the association between neighbourhood green space and individual health supports Voigtländer et al.'s claim that macro level aspects of social inequality, of which gender is one, influences individual health via the neighbourhood.

The random slopes that significantly improved the models, of which most were related to gender as an effect modifier, indicates that the effects for both men and women are not stable over the entire Dutch population. For most random slopes included, there were some neighbourhoods whose random slope opposes and exceeds the fixed slope, so that even the sign of the slope (+/-) is instable. Random slopes for the interactions between gender and relative water, park, orchard, forest and farm area improved the model. For future research, it might be useful to study the sources of these instabilities.

Strengths and weaknesses

An important strength of this study lies in its use of theory. The theoretical approach used played an important guiding role in deciding how to best use the data that was available, deciding upon appropriate interactions, and making appropriate inferences.

Relying on open data is both a strength and a weakness of this study. On the one hand, the reliability and coverage of the data sources used ensure a certain quality. All datasets used in this study are extensively checked, either by government officials (in the case of Statistics Netherlands), insurance companies and law firms (in the case of Vektis) or enthusiastic hobbyists (in the case of Open Street Maps). These sources provide this research with a large and nationally representative sample: both the dependent and the independent variables included have a near-to-full coverage of the Netherlands. Combining several datasets also ensures that neighbourhood exposures and health-seeking behaviour were measured separately, reducing the chances of correlated measurement error. On the other hand, using open data sources limits the researcher to the structure and granularity decided upon by the (many different) creators of the data. The limited granularity of the Vektis dataset has certainly limited explanatory value of the models. While the dataset includes data on almost all Dutch individuals, data is aggregated so that we do not know how GP consultation fees vary within groups with the same age, gender and 3-digit zip code. Because of this, it is unclear whether the three peaks in our dependent variable are the result of actual differences in health-seeking behaviour across the clusters, or by methodological decisions such as calculating group averages, as was discussed earlier. Vektis' decision to publish the data at 3-digit zip code level implicitly meant that all other data sources used in this study, which were often available at 4-digit zip code, had to be aggregated at a higher level, sacrificing potentially relevant micro structures. Indeed, in the case of average household income, aggregating 4-digit zip code data to 3-digit zip code level revealed that the standard deviation from this aggregated value was larger than 10% of this value for 53% of all 4-digit zip codes, and larger than 40% for 4.2% of 4-digit zip codes. For the percentage of tertiary degree holders, 15.6% of 4-digit zip codes had a standard deviation larger than 40%. Due to privacy considerations, Vektis did not publish any data on age/gender/zip code groups that consisted of less than 10 individuals. This limits generalizability of the results, since rural areas might be underrepresented.

In terms of methodology, a weakness of this study is that average declared GP consult costs were used both as a Y-variable and as an explanatory variable, because it was used in the construction of the three cost clusters, thus limiting its independence. The quality of the Y-variable is further restricted by the fact that it only includes GP consult fees that were declared either by the GP or by the individual, but not those that were paid for by the patient immediately after the consult, and no health-seeking behaviour in other primary health sectors such as physiotherapy were included. However, in the Netherlands, the GP is framed as the central primary health care gatekeeper, so that individuals are stimulated to always contact the local GP first (Schellevis & Westert, 2004). Moreover, GPs in the Netherlands are considered accessible and approachable, so that using GP consultation fees as a proxy for health-seeking behaviour was deemed reasonable.

For practical reasons this study was limited to the use of geographically defined zip code data to proxy neighbourhood, while administrative boundaries often lack intrinsic meaning, so that neighbourhoods might, in reality, not be constrained by such boundaries (Chaix et al., 2009; Riva, Gauvin, & Barnett, 2007). While 4-digit zip codes in the Netherlands were laid out to the benefit of the local postman, so that areas are compact and do not overlay barriers such as rivers or rails, this benefit is reduced when aggregating zip codes to a 3-digit level. Also, individuals are often part of many (overlapping) (geo-)contexts for living, working and leisure, so that the vicinity of the home may not be the only interesting meso level context (Vallée & Chauvin, 2012). Including more personal geo contexts, through for example using ego-centred boundaries instead of administrative boundaries, would dramatically increase explanatory and implicational value of studies looking at neighbourhood influences on health (Chaix et al., 2009). Similarly, accounting for the fact that it is likely that neighbouring neighbourhood resemble each other more than others can be done more elegantly than through including provinces as third levels. One way of modelling this is by using spatial autocorrelation.

Moreover, the cross-sectional nature of the data included limits the ability to ascertain the direction of the associations found. Therefore, no causal inferences can be made, which limits the policy implications that can be derived from this study. Being able to determine to what extend health-seeking behaviour within the neighbourhood is determined by composition or by context through a longitudinal study would provide more insight into the underlying mechanisms and the direction of the arrows in Voigtländer et al.'s framework. Finally, while controlling for several macro- and micro level variables and variables related to sociodemographic composition may lessen ecological bias, results may still be modified or confounded by exogenous effects such as an individual's mobility, level of education or ethnic background. An especially important missing individual level variable is the (private) garden access. It seems likely that having access to a garden modifies the effect that public green space has on individual health, since green space related resources and stressors could be obtained from having a private garden.

Conclusion

The results of this study demonstrate that there are strong interactions between meso level differences between types and areas of green space and individual health-seeking behaviour, but also that these interactions are not stable over neighbourhoods, green space types, or gender. Results suggest that parks, forests and fields might act as resources in shaping incidental health-seeking behaviour, while bodies of water, parks, orchards, forest, fields and farms often, and especially for females, seem to act as stressors in shaping structural health-seeking behaviour. However, it should be emphasized that this is an exploratory study, and that the interpretation of the models created is largely based on assumptions that require further consideration.

References

Aminzadeh, K., Denny, S., Utter, J., Milfont, T., Amertunga, S., Teevale, T., et al. (2013). Neighbourhood Social Capital and adolexcent self-reported wellbeing in New Zealand: A multilevel analysis. *Social Science & Medicine 84*, 13-21.

Bernard, P. et al. (2007). Health inequalities and place: A theoretical conception of neighbourhood. *Social Science & Medicine 65*, 1839-1852.

Björk, J., Albin, M., Grahn, P., Jacobson, H., & Ardö, J. (2008). Recreational values of the natural environment in relation to neighbourhood satisfaction, physical activity, obesity and wellbeing. *JECH*, 62.

Blakely, T., Atkinson, J., Ivory, V., Collings, S., Wilton, J., & Howden-Chapman, P. (2006). No association of neighbourhood volunteerism with mortality in New Zealand: A national multilevel cohort study. *International Journal of Epidemiology 35(4)*, 981-9.

CBS. (2015). *Bevolking en huishoudens; viercijferige postcode, 1 januari 2013*. Retrieved August 15, 2015, from http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=82245NED&LA=NL

Chaix, B., & ET AL. (2009). Neighbourhoods in eco-epidemic research: Delimiting personal exposure areas; a response to Riva, Gauvin, Apparicio and Brodeur. *Social Science & Medicine 69*, 1306-1310.

de Graaf-Ruizendaal, W., & de Bakker, D. (2013). The construction of a decision tool to analyse local demand and local supply for GP care using a synthetic estimation model. *Human Resources for Health*, 11-55.

DeGuzman, P., Merwin, E., & Bourguignon, C. (2013). Population Density, Distance to Public Transportation, and Health of Women in Low-Income Neighbourhoods. *Public Health Nursing Vol 30. No 6.*, 478-490.

Ellen, I., Mijanovich, T., & Dillman, K. (2001). Neighborhood effects on health: exploring the links and assessing the evidence. *Journal of Urban Affairs 23(3-4)*, 391-408.

Frohlich, K. (2013). Area Effects on Behaviour and Lifestyle. In *Neighbourhood Structure and Health Promotion* (pp. 39-60). New York: Springer Science+Business Media.

Giddens, A. (1984). *The constitution of society*. Berkeley: University of California Press.

Hall, R. (2006). *Patient Flow: Reducing Delay in Healthcare Delivery*. Los Angeles: Springer Science+Business Media.

Imergis. (2015). *Geografische Open-data GIS bestanden*. Retrieved April 30, 2015, from www.imergis.nl/asp/47.asp

Maas, J. (2008). Vitamin G: Green Environments - Healthy environments. Utrecht: NIVEL.

Maas, J., Sterkenburg, R., de Vries, S., & Pierik, F. (2013). Using GPS to Measure the Interaction Between Individuals and Their Neighbourhood. In *Neighbourhood Structure and Health Promotion* (pp. 153-175). New York: Springer Science+Business Media.

Macintyre, S., Ellaway, A., & Cummins, S. (2002). Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science & Medicine 55*, 125-139.

Meijer, M., Röhl, J., Bloomfield, K., & Grittner, U. (2012). Do neighborhoods affect individual mortality? A systematic review and meta-analysis of multilevel studies. *Elsevier Social Science & Medicine 74*, 1204-1212.

Mitchell, R., & Popham, F. (2007). Green space, urbanity and health: relationships in England. *JECH 61*, 681-3.

Mohnen, S., Groenewegen, P., Völker, B., & Flap, H. (2011). Neighborhood social capital and individual health. *Social Science & Medicine 72*, 660-667.

Naess, O., Piro, F., Nafstad, P., Smith, G., & Leyland, A. (2007). Air Pollution, social deprivation, and mortality: a multilevel cohort study. *PubMed Epidemiology 18(6)*, 686-94.

Nandi, A., & Kawachi, I. (2011). Neighborhood Effects on Mortality. In R. Rogers, & E. Crimmins, *Handbook of adult mortality* (pp. 413-439). London: Springer.

Open Street Map. (2015). *OpenStreetMap Data Extracts*. Retrieved August 9, 2015, from http://download.geofabrik.de/

Parra, D., Gomez, L., Sarmiento, O., Buchner, D., Brownson, R., Schimd, T., et al. (2010). Perceived and objective neighbourhood environment attributes and health related quality of life among the elderly in Bogotá, Colombia. *Social Science & Medicine 70*, 1070-1076.

Phillips, M., Hall, T., Esmen, N., Lynch, R., & Johnson, D. (2001). Use of global positioning system technology to track subject's location during environmental exposure sampling. *Journal for Exposure Analysis Environ Epidemol 11(3)*, 207-215.

Pickett, K., & Pearl, M. (2001). Multilevel analysis of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology and Community Health 55(2)*, 111-122.

QGIS Development Team. (2014). QGIS 2.10 Geographic Information System API Documentation. Open Source Geospatial Foundation Project.

Rijksoverheid. (2013). *Eerstelijnszorg*. Retrieved August 07, 2015, from http://www.rijksoverheid.nl/onderwerpen/eerstelijnszorg/grotere-rol-voor-eerstelijnszorg

Riva, M., Gauvin, L., & Barnett, T. (2007). Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. *Journal of Epidemiology and Community Health 61(10)*, 853-861.

Schellevis, F., & Westert, G. (2004). Nog altijd poortwachter. Rol en positie huisarts opnieuw in kaart gebracht. *Medisch Contact*, 59:622-5.

Schippereijn, J., Ejstrud, B., & Troelsen, J. (2013). GIS: A spatial turn in the health sciences? . In *Neighbourhood Structure and Health Promotion* (pp. 127-152). New York: Springer Science+Business Media.

Sociaal en Cultureel Planbureau. (2015). *Statusscores*. Retrieved August 18, 2015, from http://www.scp.nl/Onderzoek/Lopend_onderzoek/A_Z_alle_lopende_onderzoeken/Statusscores

StataCorp. (2011). *Stata Statistical Software: Release 12.* College Station, TX: StataCorp LP.

Stock, C., & Ellaway, A. (2013). Neighbourhood Structure and Health Promotion: An Introduction. In *Neighbourhood Structure and Health Promotion* (pp. 1-10). New York: Springer Science+Business Media.

Sugiyama, T., Leslie, E., Giles-Corti, B., & Owen, N. (2008). Associations of neighbourhood greenness with physical and mental health: do walking, social coherence and local social interaction explain the relationships? *JHEC 62*, 9.

Vallée, J., & Chauvin, P. (2012). Investigating the effect of medical density on health-seeking behaviours using a multiscale approach to residential and activity spaces: Results from a prospective cohort study in the Paris metropolitan area, France. *International Journal of Health Geographics 11(1)*, 54-65.

Vektis. (2015). *Vektis Open Data Databestand Zorgverzekeringswet*. Retrieved July 5, 2015, from http://www.vektis.nl/index.php/vektis-open-data

Vogitländer, S., Razum, O., & Berger, U. (2013). Methods to measure neighbourhoods and analyse their impact on health: an overview. In *Neighbourhood Structure and Health Promotion* (pp. 93-116). New York: Springer Science+Business Media.

Yen, I., Michael, Y., & Perdue, L. (2009). Neighbourhood Environment in Studies of Health of Older Adults: a systematic review. *American Journal of Preventive Medicine 37(5)*, 455-463.

Using Big Data to Study Neighbourhood Effects on Health Demand: Literature Review

M.L. Wuestman

Introduction

Hall et al. (2006) stated that the regional health care system should be designed with three goals in mind: (a) costs of providing desired services should be minimized; (b) convenience and access to services that individuals need should be maximized; and (c) the likelihood of a positive health outcome from the services should be maximized. The Dutch ministry of public health has expressed a similar ideal of efficient, local and accessible primary healthcare, organized on the level of the neighbourhood (de Graaf-Ruizendaal & de Bakker, 2013) (Rijksoverheid, 2013). Furthermore, local actors have a wish to know more about the current healthcare situation in their area, in order to identify shortages or surpluses of supply (Rijksoverheid, 2013). Functional local healthcare demand and supply monitoring is a first necessary step towards this goal. Luckily, patterns of local health care demand, even unscheduled demand, are rather predictable (Hall, 2006). In the Netherlands in 2013, a decision tool was constructed for this purpose. The demand-side was modelled using national sample-based medical record data and sociodemographic characteristics of neighbourhoods such as composition of ages, genders, households and ethnicity (de Graaf-Ruizendaal & de Bakker, 2013), and the estimates that resulted from this analysis were confronted with actual health care supply in order to estimate shortages and surpluses. This is an insightful approach, although there is also reason to assume that there are neighbourhood effects on health that cannot be determined by looking only at neighbourhood composition, such as physical attractiveness of a neighbourhood, availability of facilities for physical activity and neighbourhood cohesion (Stock & Ellaway, 2013). There is clear evidence that, even when controlling for individual characteristics, such contextual social and physical neighbourhood characteristics affect all-cause mortality, mental health, chronic conditions and health-related behaviours within a neighbourhood's inhabitants (Stock & Ellaway, 2013) and also that actual health situation is an important determinant of health care demand (Hall, 2006). Surely, these kinds of neighbourhood characteristics are not easy to measure, either because they require very intensive use of traditional data collection methods on a national scale or because traditional data collection methods are not well enough equipped to handle these kinds of data in the first place (Zimmerman, 2015). Using Big Data, and user generated online data in particular, might be one way to measure neighbourhood characteristics, because methodologies related to this kind of data are particularly well suited for studying network effects, perception and sentiment (Lin, et al., 2014) (Molla, et al., 2014). If it is possible to harness internet data to measure neighbourhood characteristics for Dutch neighbourhoods, and if these characteristics can be found to be associated to local health demand, then this insight can be of use for policy makers aiming to improve convenience and access to health care services in the Netherlands.

In this review, I will discuss neighbourhood effects on health and how these relate to Big Data methodology. I will start by explaining potential pathways from neighbourhood to health, followed by a structural review of studies that have tried to relate economic, social, institutional and/or physical neighbourhood characteristics to overall health and mortality, chronic conditions and disease prevalence, mental health and/or health-related behaviour. The first chapter will be concluded with a section on methodological challenges and opportunities in measuring neighbourhood effects on health and some final remarks on the main findings of this review, suggestions for future research and limitations of the approach chosen. This first chapter will reveal that there is a lot of movement in this field, but also that there are still quite some uncertainties, in terms of semantics, operationalisation and methods. The second chapter will introduce the concept of Big Data and identify the main controversies and opportunities surrounding it. The rest of this second chapter will be dedicated to relating Big Data to the social sciences and policy making, where I will zoom in on measuring population characteristics using internet data in particular. Again, this chapter will be concluded with a reflection on the results of this review and suggestions for future efforts in this field. The final section of this review is dedicated to linking the findings of the earlier sections, and deriving opportunities for future research from these findings.

Neighbourhood effects on Health

A growing body of literature suggests that the neighbourhood and the health of its inhabitants are connected (Stock & Ellaway, 2013). The relationship between health outcomes and living in socially and economically deprived neighbourhoods is especially striking, since studies have shown that residents of these neighbourhoods suffer from higher rates of respiratory diseases, cancer, heart disease, hospitalization of infants and overall mortality (Ellen, et al., 2001). Researchers have become interested in the mechanisms behind this relationship. Several models have been developed in order to explore the interactions between neighbourhood and health, and these models were put to the test in a vast amount of case studies. The next section contains a review of several theories and models developed to study these relationships. Next, I shall discuss the main findings of case studies conducted in this field. From these two reviews, I will draw conclusions that will highlight potential areas for future research.

Models explaining neighbourhood effects on health

Through the development of geographical variations in health as a research topic, three types of explanations of these variations have dominated the field. The first, the composition explanation, uses the concentration of individuals with similar socio-economic status in specific residential locations as an explanation for the occurring variation (Larsen, 2013). This explanation suggests that people with similar characteristics, for whatever reason, tend to live in each other's neighbourhood. This was the dominant approach until way into the 1990s, even though sociologists in Chicago already demonstrated in 1948 that negative effects on health (such as infant mortality, tuberculosis and physical abuse) of neighbourhoods with high rates of poverty, residential instability and poor housing persisted over time, despite the movement of different population groups from them (Frohlich, 2013). It appeared like neighbourhoods possess some enduring features that transcend the characteristics of their inhabitants (Shaw & McKay, 1942) (Frohlich, 2013). Halfway the 1990s this 'contextual' explanation, which suggests that there exist ecological attributes of spatially defined areas that affect whole groups so that the variation in places explains geographical health variation (Bernard, et al., 2007) (Larsen, 2013), was adopted by the majority of researchers (Macintyre, et al., 2002). Researchers working with the context explanation are interested in studying "those factors influencing human behaviours or health which remain once every imaginable individual characteristic is taken into account" (Macintyre, et al., 2002, p. 129), such as available neighbourhood services, quality of infrastructure or neighbourhood safety. Shortly after, it was found that collective properties of local residents are part of the context facing any individual living in that place, which means that it is not possible nor useful to continue seeing the compositional and contextual explanations as mutually exclusive and culturally universal (Macintyre, et al., 2002). The third explanation, the one researchers in recent projects believed to be more fruitful, acknowledges a reciprocal relationship between people and place that co-determines health (Larsen, 2013) (Macintyre, et al., 2002) (Bernard, et al., 2007). According to this approach, health and neighbourhood are both complex concepts that reinforce each other within the context of culture and cannot be studied without considering the interactions between the individual, the community and health. Through consumption, service use, politics and social interactions, neighbourhood actors shape and reproduce their context, while their lifestyle and health are affected by those goods consumed and services used (Bernard, et al., 2007). The contemporary sociologist Anthony Giddens (1984) relates to this reciprocal relationship between configurations and individual behaviour as the interaction between 'structure' and 'agency'. As recent studies have acknowledged this approach, it serves as the foundation of each of the models discussed hereafter.

Besides recognizing reciprocity between people and place, each of these models also distinguishes between social and physical environments within the neighbourhood. The social environment of urban areas has been described as the total of values and norms that are shared by members of social groups, and the relationships and interactions shared among residents and communities within the area (Galea, 2007). The physical environment includes both the built and the natural environment, where the former includes trees, bodies of water and geological and climatic conditions and the latter consists of transport routes and networks, housing, shops, parks and public areas (Galea, 2007). The elements in both the social and the physical environment are interdependent and can be both 'pathogenic' and 'salutogenic' for residents in its proximity (Frohlich, 2013). In a later section, the usefulness of this distinction will be discussed.

Opportunity Structures Framework. Macintyre, Ellaway and Cummins (2002) made an early attempt in creating a framework explaining the geographical variations in health through what they call 'opportunity structures'. These are 'socially constructed and socially patterned features of the physical and social environment which may promote or damage health either directly, or indirectly through the possibilities they provide for people to live healthy lives' (p. 132). Through these features, possibilities for people to live more or less healthy lives become socially distributed. An example of direct damaging of health by opportunity structures would be if polluted air compromises residents' health; an example of creating opportunities through indirect promotion of health is the local availability of affordable and nutritious food (Macintyre, et al., 2002). Opportunity structures arise through five features of local areas which might promote or damage health: (1) physical features of the environment shared by all residents in a locality; (2) availability of health environments at home, work and play; (3) services provided, publicly or privately to support people in their daily lives; (4) socio-cultural features of a neighbourhood and (5) the reputation of an area.

The framework's authors discuss one major limitation of this framework: it does not specify what exactly would need to be studied within each of the five features in order to improve our understanding of the importance of the different aspects of the social and physical environment (Macintyre, et al., 2002). Another limitation lies in the fact that this framework does not give insight into what determines the opportunity structure on an individual level. Although opportunity structures are explained as a neighbourhood characteristic, not every individual is equally likely to suffer or benefit from the neighbourhood's health promoting or damaging features. It would be useful to expand the framework by including a model to explain how differences in, for example, an individual's personality, education, network, social background, income, et cetera influence their personal experience of the neighbourhood's opportunity structure.

Rules of Access. Where the opportunity structures framework discussed above uses a rather sociological approach, the next borrows its main concepts from the field of economics. This framework views differences in the distribution of resources within the context of the neighbourhood as the main source of local associations with health (Bernard, et al., 2007). The local configuration of resources thus affects local social interactions and health while, on the other hand, those social interactions shape the local configuration and

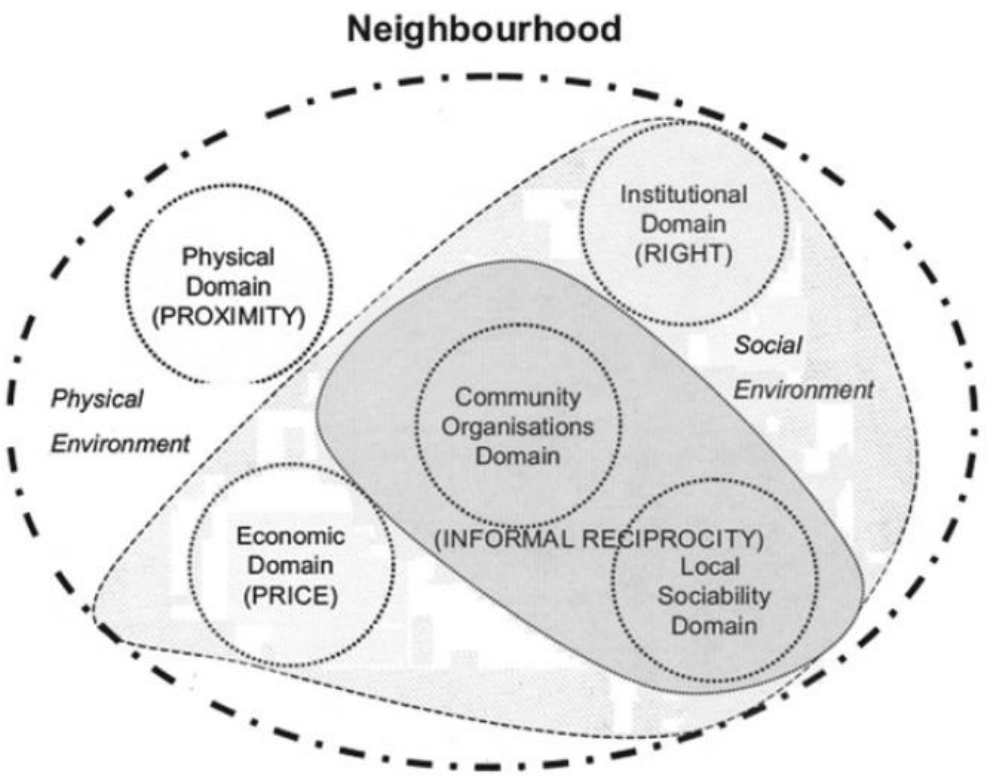

Figure 7. Neighbourhood domains and rules of access (Bernard et al., 2007)

constitute spatially defined distribution networks through which the resources are accessible (Bernard, et al., 2007). Bernard et al. use the Opportunity Structures framework to explain how the local configuration of resources creates opportunities and constraints that shape and orient inhabitants' behaviour (Giddens, 1984), thus promoting or damaging health.

For a categorization of the different types of resources that are distributed and exchanged, the authors refer to the work of Godbout (2003). Godbout distinguishes between three sets of rules for the circulation of resources: (1) market rules, which apply to those resources that can be obtained by paying a price for them, (2) state/institutional rules, regulating access to resources which citizens are entitled to according to publicly enacted rules, and (3) rules of informal reciprocity, which involve self-regulating participants and creating social ties through voluntary and community organisations (Godbout, 2003). Informal reciprocity is based on the notions of gift and trust, which create social ties because they bring the system in a perpetual state of imbalance (Bernard, et al., 2007). Bernard et al. suggest that access to local resources is determined by these three rules and adds a fourth: proximity. While proximity is the result of the physical domain, the first three are considered part of the social environment, which can be subdivided into four domains: the economic domain, based on price; the institutional domain, based on right; the community organisations domain and the local sociability domain, both based on informal reciprocity (Bernard, et al., 2007). A visualisation of the rules and their domains can be found in figure 7. The difference between the community organisations domain and the local sociability domain lies in the way social networks are used: while community organisations have the role of mobilizing networks and provide leverage in order to pursue collective goals and change the physical, economic and institutional environments, local sociability is aimed at procuring individual benefits such as particular information and social support (Bernard, et al., 2007) . The local configuration of resources can be shaped and reshaped when conflicts occur between the different domains, for example when health services are provided by members of different domains simultaneously.

A major limitation of this framework is that it limits itself to the community level, and does not explain how the interactions between these resources, domains and rules bring about health inequalities at an individual level. Also, Bernard et al. do not distinguish between access to local resources and use of local resources. While they do add proximity as a rule, referring to the social or physical proximity of resources, they do not seem to recognize that actual (health-seeking) behaviour is the result of personal motivation and external triggers as well as accessibility/ability (Fogg, 2009). Intuitively, it could be helpful to include a fifth rule that concerns an individual's attachment to the neighbourhood, since it is very well possible that individuals use resources made available through another environment, such as work, family, or former neighbourhood.

The ISIS-framework. While Bernard and colleagues' original model did not include any explanation on how the resources are transformed into health and health inequalities (Bernard, et al., 2007), a later paper by Katherine L. Frohlich (2013) expanded the model by adding an explanation of inequalities in health on the individual level by Abel (2008) in what she called the Interdisciplinary Study of Inequalities in Smoking Framework (ISIS-framework). This framework can be found in figure 8. As the name suggests, this model was derived from a study on social inequities in smoking, but the theory is formulated more generally. Frohlich suggests that 'health is produced not only with (or without) the structural constraints and opportunities offered at the local level but through individuals' capital stock which permits them to identify, access and utilise (or not) resources in neighbourhoods to their health advantage' (Frohlich, 2013, p. 53). Abel and Frohlich recognize four different kinds of individual capital, based on an earlier classification by Bourdieu (1986): (1) economic capital, representing money and material assets; (2) social capital, representing the resources (material or immaterial) which can be accessed through interpersonal social relationships; (3) biological capital, representing natural assets and (4) cultural capital, representing people's symbolic and informational resources for action and which

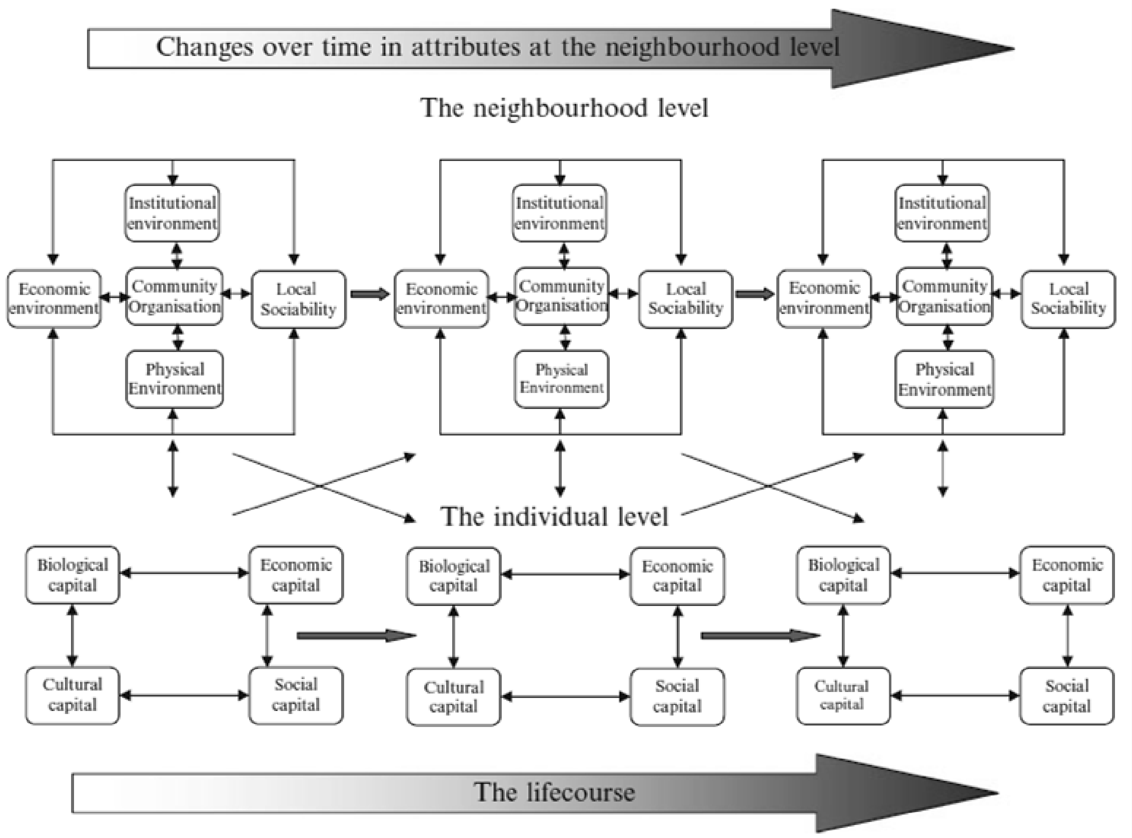

Figure 8. the ISIS framework for explaining how inequities in health are produced in neighbourhoods (Frohlich, 2013)

can be divided into incorporated (skills and knowledge), objectivised (books, tools) and institutionalised (educational degrees) cultural capital. The acquisition and development of and the interactions between these capitals is what constitutes individual and collective agency. At the individual level, Frohlich suggests, capitals provide the agency potential for health, but this potential is dependent on resource availability and accessibility within the neighbourhood (Frohlich, 2013).

Frohlich also recognizes that both neighbourhoods and individuals change over time. Her model includes those changes on both the collective and the individual level, and points out that there are interaction effects and path dependencies in the development of people and place (Frohlich, 2013). This is important, because this insight suggests that it might not be sufficient to observe individual and collective characteristics in order to study health inequality, because the foundations for these inequalities might have been determined in an earlier phase. A limitation of this framework is that it does not explain how the interactions within and between the different types of capital and different neighbourhood level domains actually lead to health inequality at an individual level. While it is clear that both neighbourhood level and individual level effects influence individual health, it seems like a more psychological or behavioural explanation of how these effects change behaviour and how behaviour changes health was not studied. Again, it would be useful to include a measure at the individual level for an individual's motivation to use his or her capitals, perhaps expressed in a sense of attachment to the neighbourhood.

Voigtländer and colleagues' conceptual framework. Another framework that used Bernard et al.'s framework to explain differences in individual health status is that of Voigtländer et al., depicted in figure 9. This conceptual framework links the macro, meso and micro levels present in society together, so that the position of the individual within the structure of social inequality influences individual health, through exposure to stressors and resources within social contexts, such as neighbourhoods (Voigtländer, et al., 2013). I will first discuss what each of these three levels entails, followed by a description of the interactions between them.

The macro level consists of the social inequalities existing within a society. These inequalities are apparent in, for example, differences in sex, ethnicity, income and choice of residential location but also in the health status of individuals. The meso level represents the neighbourhood. Voigtländer et al. identify four categories of stressors and resources in the neighbourhood: (1) the physical environment, (2) markets, (3) institutions and (4) social capital. These four are derived from Bernard et al.'s domains through which the distribution of resources are determined; (collective) social capital represents the two informal reciprocity domains (Voigtländer, et al., 2013). Besides these, the neighbourhood context is characterized by its sociodemographic composition in terms of, for example, age, education, income and ethnicity and the changes that occur within this composition. Resources and stressors and the sociodemographic composition influence each other. Finally, the micro level consists of the individual and the effects which the neighbourhood's resources and stressors has on him.

The first interaction, between the macro and the meso level, works in two ways. Firstly, an individual's choice of residential locations and, as such, the potential neighbourhood and its composition, is determined by his or her social position within the structure of social inequality. Secondly, social inequality 'spatializes' itself, resulting in a spatial concentration of people, organisations and other contextual characteristics within neighbourhoods. For the second interaction, the neighbourhood influences the individual in three ways: a direct pathway from dangerous neighbourhood environments to direct health effects; an indirect pathway from neighbourhood environments, through individual perception and assessment of these environments and the distress or behavioural change, to individual health and, finally, individual health is determined by personal resources which are the result of individual health and social position, and which leads via the indirect pathway to individual health as well (Voigtländer, et al., 2013). The third and final interaction between levels works rather intuitively: the micro level influences the macro level, because the collective health status of individuals constitutes a part of social inequality.

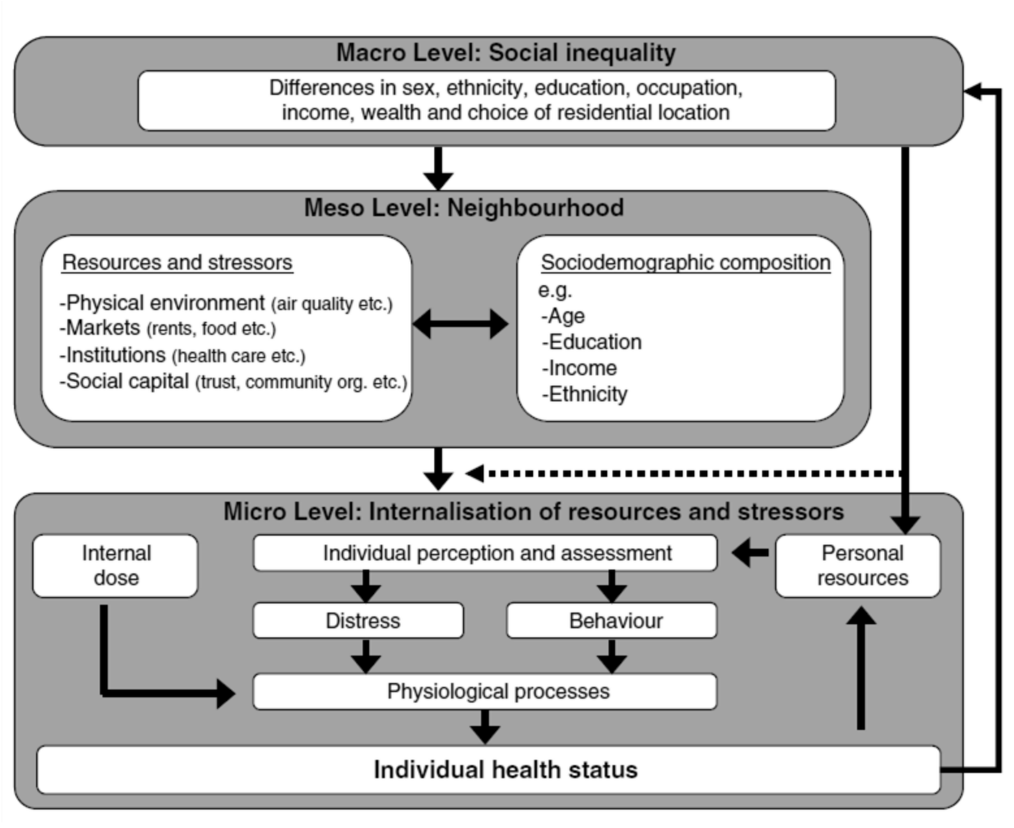

Figure 9. Conceptual framework linking social inequality, neighbourhood context and health (Voigtländer et al., 2013)

This model's strength lies in its multilevel perspective and its focus on interactions. This clarifies the distinction between individual characteristics and community characteristics, and is a good example of a reciprocal explanation as opposed to a composition- or contextual explanation of health inequality. However, the way the model explains the mechanisms within the micro level comes across as quite a chaotic process. Voigtländer's use of the term 'social capital' is also slightly confusing, since the ISIS-framework uses it in a different context. However, in literature on social capital, it is generally accepted that social capital can be both an individual and a group attribute. Kawachi (2006) conceptualizes social capital as a group attribute as 'the resources available to members of social groups', so that an uncooperative and mistrusting member of a group can still reside in a trusting and helpful community. Another explanation is that (individual) social capital refers to all resources potentially owned by social network members, which may become available to a focal individual as a result of mutual investments in a shared past (van der Gaag & Snijders, 2004). Both individual and collective social capital can either be bonding (resources accessed within groups whose members are alike) or bridging (resources accessed through connections that cross boundaries of social identity) (Kawachi, 2006), and can be both positively and negatively associated with health (Portes, 1998).

Conclusion. When comparing these models, it is clear that all of them recognize a clear distinction between the social and the physical world. However, whereas for some researchers local institutions and market mechanisms are part of the social world (Macintyre, et al., 2002) (Frohlich, 2013), for others these domains deserve to be seen as another world (Bernard, et al., 2007) (Voigtländer, et al., 2013) (Galea & Vlahov, 2005). This illustrates the inherent difficulty with this distinction between the social and the physical environment. In practice, these two environments are not as distinct as these models make them appear. Neighbourhood education level, for example, is a result of both the social world (interactions between teachers and students, the local perceived importance of education) and the physical world (the quality of school buildings, access to facilities). Therefore, when measuring education in terms of the social and physical environment separately, correlations are likely to be very high, so that the models discussed above, all using this distinction, are empirically not very practical. A second ambiguity lies in the different models' conceptions of social capital: whereas the ISIS framework sees social capital as one of the capitals that together produce health within the individual realm, Voigtländer et al. use the same term to address what Bernard et al., and Frohlich call 'informal reciprocity' and thus sees social capital as a group attribute. This kind of semantic confusion is also apparent in the fact that none of the authors of the models discussed above use a clearly delineated definition of what they mean by 'the neighbourhood'. A third major limitation is that these models do not supply researchers with clear predictions, because they are not normative and they do not give any insight into whether associations between neighbourhood effects and health are causal, or even whether they are positive or negative. These models cannot easily serve as policy instruments; they do not imply that, for example, improving the social environment increases the health of a neighbourhood's inhabitants.

For future research it would be interesting and insightful to test these models and further develop them into useful, applicable tools for managing health at a neighbourhood level. This would involve useful and clear definitions, increasing attempts to make normative models, and suggesting causal pathways from individual susceptibility to neighbourhood effects, to neighbourhood effects and their consequences for individual health. It would be useful to integrate the strengths of the different models: where the ISIS-framework, including Bernard and colleagues' work and the opportunity structures framework, is strong in its recognition of different levels, its systematic approach to the processes in these two levels and its recognition of temporal changes, Voigtländer's approach is strong in including psychological and behavioural processes and including a third level, both of which are missing in the ISIS-framework. In practice, an option could be to add a third 'landscape'-level to the ISIS framework in order to put the neighbourhood in its context and make neighbourhoods comparable, and a fourth level that describes individual psychological processes and personality in order to understand how capitals are established and how they interact with the neighbourhood's opportunity structure, and how these processes result in health outcomes. More attention to positive and negative causal pathways between different aspects of the neighbourhood environment and the individual would be a useful step towards a normative, interpretative framework.

Review of experiments

This section will be dedicated to getting a grasp of the different attempts that have been made to reveal associations between neighbourhood and health. Since the early 90's, researchers have studied potential associations increasingly, and their achievements, and the trends that can be perceived over their studies, will be summarized shortly.

In order to systematically review the current body of literature on concrete attempts at finding associations between health outcomes, individual and neighbourhood characteristics in a specific area, it is useful to distinguish between the different possible health outcomes, and the nature of the possible neighbourhood characteristics to which these health outcomes are linked. The result, which can be seen in table 4, can be viewed as a four-by-four table, upon which all studies included in this review can be mapped. The categories on each of the axes of the table can be derived from theory. Yen et al. (2009) distinguish four categories of health outcomes that have been studied in relation to their possible association with health: overall health and mortality, chronic conditions or disease prevalence, mental health outcomes, and health behaviours (such as smoking, physical activity or seeking medical attention). This is a useful distinction to make, because it helps us to see to which kinds of health care most attention has been paid in recent studies, so these four categories can be found on the x-axis of this system. Different types of neighbourhood characteristics can be found on the y-axis and consist of the four categories of neighbourhood characteristics identified by Bernard et al. (2007); the physical, economic and institutional environment and the environment of informal reciprocity. These four categories were chosen because it gives more insight than merely distinguishing between social and physical environments.

Studies were found in previous reviews (Yen, et al., 2009) (Meijer, et al., 2012) (Pickett & Pearl, 2001). The three reviews used all include studies that are published in English peer-reviewed journals, report data from primary studies, are based on populations from developed countries and a random sample of the adult population, used multilevel modelling, and used at least one individual and at least on residential area-level variable. In total, 95 studies were included between 1983 and 2008, of which 47 were conducted in the United States, 14 in Scandinavian countries, 15 in the United Kingdom and 4 in the Netherlands. To get a more recent overview of the work done, additional relevant papers published between 2008 and 2015 were added as well. All of these works, including those identified by the authors of previous reviews, have the following structure: each have (multiple) health outcomes as dependent variable, (multiple) neighbourhood characteristics as independent variable, and multiple individual characteristics as control variables (this implies that one work can be found in several cells in the framework used here). Furthermore, each of these studies have used a multilevel approach, which will be discussed in more detail in the section on methodological challenges which will follow later in this chapter, and all of them are scholarly and peer reviewed publications. The additional, most recent, works were identified using this search query in the Focus search engine:

(TitleCombined:(neighbourhood) OR TitleCombined:(neighborhood) OR TitleCombined:(area)) AND (TitleCombined:(health)) AND (TitleCombined:(effect) OR TitleCombined:(effects) OR TitleCombined:(determinants)) AND (Abstract:(multi-level) OR Abstract:(multilevel) OR TitleCombined:(multi-level) OR TitleCombined:(multilevel))

This resulted in 28 papers, of which two were double and one was not relevant, so finally 25 papers were included. In Appendix A, a table can be found that describes, for each of these, its first author, publication year and study period, studied country, sample size, the different (individual/neighbourhood) characteristics included and finally the main results of the study (table 5). A similar description of the works included in the other three reviews that were included in this analysis can be found in the corresponding papers. All information mentioned below is derived from the three pre-existing reviews and the 25 new papers. Details, including full references, can be found in the original reviews.

Results. *First row: economic environment* . Within the domain of the economic environment, it appears that most attention has been paid to overall mortality and chronic conditions and disease prevalence. The economic characteristics that have been studied most often across all four health outcomes are average income in the neighbourhood (Ko, et al., 2014) (O'campo, et al., 1997) (Reijneveld, 1998) (Fisher, et al., 2004) (Aneshensel, et al., 2007) (Kubzansky, et al., 2005) (Ostir, et al., 2003) (Walters, et al., 2004) (Deeg & Thomese, 2005) (Anderson, et al., 1997) (Sloggett & Joshi, 1998) (Turrell, et al., 2007) (Blakely, et al., 2003) (Curtis, et al., 2004) (Dahl, et al., 2006) (Lochner, et al., 2001) (Kravdal, 2007) (Voigtländer, et al., 2010), neighbourhood (un)employment rate (Borrell, et al., 2002) (O'campo, et al., 1997) (Mari-Dell'Olmo, et al., 2007) (Humphreys & Carr-Hill, 1991) (Roberts, 1997) (Bosma, et al., 2001) and the, more general, average socio-economic status (SES) within a neighbourhood (Bentley, et al., 2008) (Webster, et al., 2008) (Robert & Li, 2001) (Nordstrom, et al., 2004) (Merkin, et al., 2007) (Diez-Roux, et al., 1997) (Jerrett, et al., 2005) (Naess, et al., 2007) (Roberts, 1997). Economic environment and overall health studies have focused on two main health outcomes: overall mortality and self-rated health. Associations between overall mortality and low average household income have been found by Blakely et al. (2003), Ko et al. (2014), Sloggett & Joshi (1998), Turrell et al. (2007) and Curtis et al. (2004); Anderson found this association between average household income of the neighbourhood and an individual inhabitant’s health to be particularly strong for older men. However, Dahl et al. (2006), Wong et al. (2009), Lochner et al. (2001) and Kravdal (2007) found no such association, and stated that income inequality within a neighbourhood may have a larger effect on individual inhabitants’ health than neighbourhood average income. Evidence for an association between neighbourhood average income inequality and individual health was broadly found, especially for people younger than 65 (Backlund, et al., 2007) (Waitzman & Smith, 1998) (Lochner, et al., 2001) (Dahl, et al., 2006), and especially in areas with many manual labourers (Henriksson, et al., 2007). Roberts (1997) also studied the effect of age on the association between area-level SES and individual self-rated health, and found that this effect is weakest for young adults and 70+ adults, and strongest for adults between 60 and 69 years old. Furthermore, some studies suggest associations between self-rated health and low average household income in the neighbourhood (Cagney, et al., 2005) (Patel, et al., 2003) (Robert & Li, 2001) (Subramanian, et al., 2006), and others found associations between mortality and neighbourhood unemployment rate (Bosma, et al., 2001) and neighbourhood SES composition (Jerrett, et al., 2003) (Naess, et al., 2007)

Economic environment and chronic conditions/disease prevalence have been studied with reference to neighbourhood SES and unemployment rate as well. Associations between chronic conditions and diseases and neighbourhood SES were found for, among others, cancer (Bentley, et al., 2008), women's blood pressure (Diez-Roux, et al., 1997), type 2 diabetes (Maier, et al., 2014), cardiovascular disease (Jackson, et al., 2008) and kidney disease (Merkin, et al., 2007), while (male) unemployment rate has been found to be associated with HIV (Mari-Dell'Olmo, et al., 2007), respiratory disease (Humphreys & Carr-Hill, 1991) and low birth weight (O'campo, et al., 1997) (Roberts, 1997). Other economic neighbourhood factors that have been studied are, for example, percentage of rental housing, the proportion of a neighbourhood's population without access to a car (Humphreys & Carr-Hill, 1991), and median house value (Franzini & Spears, 2003).

Associations between mental health and the economic environment have been studied to a lesser extent. Here, the focus has largely been on low average neighbourhood income, which was found to be related to depressions (Aneshensel, et al., 2007) (Kubzansky, et al., 2005) (Ostir, et al., 2003) (Phillips, et al., 2011), loneliness (Deeg & Thomese, 2005) and over-all mental health perception (Kwag, et al., 2011). However, Walters found no association between either of these three. Albor et al. (2014) found that not average income but income inequality was associated with the odds of having more neighbourhood friends and the odds of depression or anxiety for high income mothers. Behanova et al. (2013) found that area unemployment and mental health were strongly associated in the Netherlands, but not in Slovakia.

Finally, some studies suggest that there is an association between the economic environment and health-related behaviour. For example, low average income in the neighbourhood was found to be associated with violence (O'Campo, et al., 1995), smoking (Reijneveld, 1998) (Adams, et al., 2009), obesity (Maier, et al., 2014) (Adams, et al., 2009) and walking behaviour (Fisher, et al., 2004). Borrell et al. (2002) found a relation between unemployment rate and injury, and O'Campo et al. (1995) found a similar association between unemployment rate and violence. It can, however, be questioned whether unemployment in this sense should be considered part of the economic domain or part of the domain of informal reciprocity.

Table 4. Characterization of relevant studies.

	Overall Health/Mortality				Chronic Condition/Disease Prevalence				Mental Health		Health-related Behaviours	
Economic Environment	Haan, et al.	1987	Jerrett, et al.	2003	Humphreys & Carr-Hill	1991	Martikainen, et al.	2003	Ostir, et al.	2003	Krieger	1992
	Humphreys & Carr-Hill	1991	Martikainen, et al.	2003	Krieger	1992	Eschbach, et al.	2004	Walters, et al.	2004	Curry, et al.	1993
	Sloggett & Joshi	1994	Patel, et al.	2003	Jones & Duncan	1995	Nordstrom, et al.	2004	Kubzansky, et al.	2005	O'Campo, et al.	1995
	Jones & Duncan	1995	Curtis, et al.	2004	O'campo, et al.	1997	Wen & Christakis	2005	Deeg & Thomese	2005	Karvonen & Rimpela	1996
	Anderson, et al.	1997	Roos, et al.	2004	Roberts	1997	Deeg & Thomese	2005	Wight, et al.	2006	Reijneveld	1998
	Diez-Roux, et al.	1997	Cagney, et al.	2005	Diez-Roux, et al.	1997	Blakely, et al.	2006	Hybels, et al.	2006	Borrell, et al.	2002
	LeClere, et al.	1998	Deeg & Thomese	2005	Waitzman & Smith	1998	Chaix, et al.	2006	Aneshensel, et al.	2007	Martikainen, et al.	2003
	Sloggett & Joshi	1998	Naess, et al.	2005	Reijneveld	1998	Merkin, et al.	2007	Naess, et al.	2007	Fisher, et al.	2004
	Waitzman & Smith	1998	Dahl, et al.	2006	Robert	1998	Chaix, et al.	2007			Blomgren, et al.	2004
	Reijneveld	1998	Henriksson, et al.	2006	LeClere, et al.	1998	Mari-Dell'Olmo, et al.	2007	Phillips, et al.	2011	Berke, et al.	2007
	Robert	1998	Robert & Ruel	2006	Robert & Li	2001	Naess, et al.	2007	Kwag, et al.	2011	Naess, et al.	2007
	Waitzman, et al.	1999	Subramanian, et al.	2006	Borrell, et al.	2002	Bentley, et al.	2008	Behanova, et al.	2013		
	Bosma, et al.	2001	Bowling, et al.	2006	Blakely, et al.	2003	Chaix, et al.	2008	Albor, et al.	2014	*Adams, et al.*	*2009*
	Lochner, et al.	2001	Backlund, et al.	2007	Franzini & Spears	2003	Webster, et al.	2008			*Halonen, et al.*	*2012*
	Lochner, et al.	2001	Henriksson, et al.	2007							*Maier, et al.*	*2014*
	Veugelers, et al.	2001	Kravdal	2007	*Jackson, et al.*	*2008*						
	Robert & Li	2001	Naess, et al.	2007	*Maier, et al.*	*2014*						
	Blakely, et al.	2003	Turrell, et al.	2007								
	Wong, et al.	*2009*	*Kwag, et al.*	*2011*								
	Voigtländer, et al.	*2010*	*Ko, et al.*	*2014*								
Social Environment	Jones & Duncan	1995	Cagney et al.	2005	Humphreys & Carr-Hill	1991	Schieman & Meersman	2004	Ostir, et al.	2003	Booth, et al.	2000
	Robert & Li	2001	Deeg & Thomese	2005	Jones & Duncan	1995	Robert, et al.	2004	Walters, et al.	2004	Balfour & Kaplan	2002
	Bosma et al.	2001	Jaffe et al.	2005	Shouls, et al.	1996	Wen & Christakis	2005	Schieman & Meersman	2004	Borrell, et al.	2002
	Malmstrom et al.	2001	Jaffe et al.	2005	O'campo, et al.	1997	Deeg & Thomese	2005	Kubzansky, et al.	2005	Martikainen, et al.	2003
	Veugelers et al.	2001	Robert & Ruel	2006	Roberts	1997	Jaffe, et al.	2005	Deeg & Thomese	2005	Fisher, et al.	2004
	Patel et al.	2003	Subramanian et al.	2006	Krause	1998	Jerrett, et al.	2005	Wight, et al.	2006	Blomgren, et al.	2004
	Blakely et al.	2003	Bowling et al.	2006	LeClere, et al.	1998	Bowling, et al.	2006	Hybels, et al.	2006	Li, et al.	2005
	Jerrett et al.	2003	Backlund et al.	2007	LeClere, et al.	1998	Chaix, et al.	2006	Aneshensel, et al.	2007	Michael, et al.	2006
	Martikainen et al.	2003	Kravdal	2007	Robert & Li	2001	Chaix, et al.	2007	Naess, et al.	2007	Naess, et al.	2007
	Curtis et al.	2004	Naess et al.	2007	Borrell, et al.	2002	Merkin, et al.	2007				
	Roos et al.	2004			Blakely, et al.	2003	Chaix, et al.	2007	*Hull, et al.*	*2008*	*Uthman & Kongyuy*	*2008*
					Franzini & Spears	2003	Chaix, et al.	2007	*Wu, et al.*	*2010*	*Murayama, et al.*	*2012*
	Li & Chuang	*2009*	*Crammet al.*	*2012*	Martikainen, et al.	2003	Naess, et al.	2007	*Riva, et al.*	*2011*		*2015*
	Murayamaet al.	*2012*	*Bécareset al.*	*2013*	Diez Roux	2004	Chaix, et al.	2008	*Bécares, et al.*	*2013*	*Cunningham-Myrie, et al.*	
					Eschbach, et al.	2004						
					Nordstrom, et al.	2004	*Mohnen, et al.*	*2014*				

Table 4. continued

Institutional Environment	Jerrett et al.	2003			*Bécares, et al.*	*2013*	*Vallée, et al.*	*2010*
							Vallée & Chauvin	*2012*
	Bécares et al.	*2013*						
Physical Environment	Jones & Duncan	1995	Jones & Duncan	1995	Kubzansky, et al.	2005	Fisher, et al.	2004
	Blakely, et al.	2003	Blakely, et al.	2003	Deeg & Thomese	2005	Blomgren, et al.	2004
	Jerrett, et al.	2003	Patterson & Chapman	2004	Naess, et al.	2005	Li, et al.	2005
	Curtis, et al.	2004	Clarke & George	2005	Berke, et al.	2007	Michael, et al.	2006
	Deeg & Thomese	2005	Deeg & Thomese	2005			Naess, et al.	2007
	Subramanian, et al.	2006	Jerrett, et al.	2005	*Phillips, et al.*	*2011*		
	Naess, et al.	2007	Naess, et al.	2007			*Cunningham-Myrie, et al.*	*2015*
			Chaix, et al.	2008				
	Voigtländer, et al.	*2010*	Webster, et al.	2008				

Second row: environment of informal reciprocity. Next, the main findings on the effects of the domain of informal reciprocity on health will be discussed. Informal reciprocity health effects are largely related to the quantity and quality of services available in the neighbourhood and (individual and community-level) social capital. Other social characteristics that cannot be understood as part of the economic or institutional environment (divorce rate, religious affirmation in the neighbourhood, share of households lead by females, average maternal age, ethnic groups present in the neighbourhood) are considered part of this environment as well. In terms of overall health, most research has found that there is a (weak) association between poor quality facilities (Murayama, et al., 2012) (Subramanian, et al., 2006) (Bowling, et al., 2006) and between low neighbourhood-level social capital (Bowling, et al., 2006) (Blakely, et al., 2006) (Mohnen, et al., 2014) and self-rated health. One interesting result was by Murayama et al. (2012), who found that neighbourhood-level institutional mistrust was associated with self-reported poor health. Others found a strong association between self-rated health and having low status in a high-status neighbourhood (Roos, et al., 2004) (Deeg & Thomese, 2005) (Kravdal, 2007) (Backlund, et al., 2007), percentage of manual workers (Martikainen, et al., 2003), divorce rate (Kravdal, 2007), low religious affirmation (Jaffe, et al., 2005) and rate of female family leaders (LeClere, et al., 1998) (Li & Chuang, 2009).

Associations between informal reciprocity and chronic conditions and disease prevalence were found. Socio-economic status of the neighbourhood seems to be associated to subclinical disease (Nordstrom, et al., 2004), cardiovascular disease (Diez Roux, 2004) and breast cancer (Roos, et al., 2004). Crime rate has also been studied in this respect (O'campo, et al., 1997). Some researchers have focused on differences between ethnic groups (Bécares, et al., 2013) (Diez Roux, 2004) (Eschbach, et al., 2004), but results from this have been inconclusive.

Results on the effect of informal reciprocity on mental health are less clear than the results mentioned above. While some researchers found a relationship between mental health problems and neighbourhood education level (O'campo, et al., 1997) (Wu, et al., 2010), neighbourhood status (Deeg & Thomese, 2005) (Aneshensel, et al., 2007) (Kubzansky, et al., 2005) (Ostir, et al., 2003) and religious/sportive participation (Hull, et al., 2008), Walters et al. (2004) and Hybels et al. (2006) studied these characteristics as well and concluded by stating that no such relationship could be found. Walters did find a relationship between mental health and population density and reported neighbourhood problems. Schieman & Meersman (2004) found the latter to be particularly important for men, Bécares et al. (2013) studied this in relation to ethnic minorities and Riva approached this question in terms of rural and urban areas

As was mentioned earlier, unemployment rate and the type of work done by a neighbourhood's inhabitants can be considered to be part of the environment of informal reciprocity as well. Associations between the percentage of people doing manual work (Blomgren, et al., 2004) and the unemployment rate (Borrell, et al., 2002) and health-related behaviour such as alcohol related mortality and injury were indeed found. Also, a relationship was found between a community's level of education and shared attitude towards smoking and smoking behaviour (Curry, et al., 1993) and, not very surprisingly, between neighbourhood physical activity and obesity (Cunningham-Myrie, et al., 2015).

Third row: institutional environment. When looking at the table, it is clear that the institutional environment has not received as much attention as the other environments. Those who did study a neighbourhood's institutional environment focused on the way health care supply was organized in an area. Jerrett et al. (2003, 2005) studied the relationship between all-cause mortality and the amount of doctors per hospital bed but did not find a strong association. Such an association was found by Subramanian et al. (2006), who found that individuals who reside in neighbourhoods with fewer services that promote social organization were less likely to report poor health. Vallée et al. (Vallée & Chauvin, 2012) (Vallée, et al., 2010) also found a positive association between the administrative neighbourhood of residence, general practitioner density in particular, and the health-seeking behaviour in women, but only for those women who concentrated their daily activities within their neighbourhood of residence (as opposed to those who spend their time largely within their work environment, for example). When repeating this study two years later, this result was confirmed. Finally, Bécares et al. (Bécares, et al., 2013) studied the effect of racist practices embodied within institutional structures on self-rated health and mental disorders and found a small effect.

Fourth row: physical environment. Although the physical environment has received more attention than the institutional environment, it is still not a very common approach. The physical environment is often seen as one of several determinants of neighbourhood quality of life and area deprivation, and is usually not seen as an interesting topic in itself (Blakely, et al., 2006). For example, Naess et al. (2007) found that deprived neighbourhoods were more often exposed to air pollution.

In relation to overall health and mortality, Jerrett et al. (2003) (2005) found that high levels of air pollution were associated with greater all-cause mortality in inhabitants of the neighbourhood. Clarke & George (2005) studied the relationship between land-use and health-related behaviours, and found that neither housing density nor land-use density was associated with health-related behaviours, although housing density and land-use diversity moderated the association between deprivation and health-related behaviours. Blomgren et al. (2004) also studied health-related behaviours, and found that urbanization rate has a protective effect for areas with many manual workers, where alcohol-related mortality tends to be higher. From these results, it appears that the physical environment often has an indirect effect on health. However, Jones and Duncan (1995) did find a direct association between smoking and urban environment, and Fisher et al. (2004) also found associations between aspects of the physical environment (physical activity facilities per mile, density of place, number of street intersections, amount of green and open recreational spaces, facility accessibility) and increased walking behaviour in individuals. These results were contested by Michael et al. (2006), who found no such association between walking activity and shop accessibility, physical activity resources, sidewalk presence and quality and the presence of graffiti or litter. Finally, Berke et al. (2007) found that men living in more-walkable neighbourhoods had reduced odds of significant depressive symptoms compared to men in less-walkable neighbourhoods.

Conclusion. When taking some distance and looking at the table at large, several observations stand out. First of all, it is clear that research has concentrated on the upper left quartile; within the environments of informal reciprocity and economy and overall health and chronic conditions/disease prevalence. Mental health in relation to neighbourhood effects is a relatively new field of study, with the oldest study coming from 2003. The physical and especially the institutional environments seem to have been underrepresented in recent research. The study of the institutional environment in relation to neighbourhood effects is rather new as well, though it is hard to say whether this field will grow in the future. It is surprising that not much research has focused on the physical environment, because all theories discussed earlier in this chapter are careful to consider the physical environment as an important determinant of health next to the social environment. However, not all theories consider the institutional environment as a distinct environment, which might explain why not many researchers have considered it at all. The emphasis on the environments of informal reciprocity and economy might be explained by the fact that a large share of the research done in these fields study neighbourhood SES, which can be derived quite easily from figures published by national governments and statistics offices such as the Dutch CBS (CBS, 2014). Neighbourhood SES, although technically a neighbourhood characteristic, is based on a compositional explanation rather than the contextual or reciprocal explanation, and therefore do not provide insight into the neighbourhood through more than a summation of individual characteristics (Diez Roux, 2002). This can be problematic. If, for example, there is a positive association found between unemployment rate and depression, it is not clear whether this means that, for whatever reason, both employed and unemployed inhabitants tend to be more depressed in a specific neighbourhood with a high unemployment rate, or that unemployed people – of which there are a lot in this neighbourhood – tend to be more depressed in general, in which case we cannot speak of a neighbourhood effect at all.

Limitations of the approach used in this review are mostly related to the inclusion and exclusion of papers and the way the framework used was organized. First of all, some papers are hard to position within the framework because their topics are not easily translated into the terms used here: are injuries, accidents, suicides and STDs part of disease prevalence, or are they behaviour-related health outcomes? Secondly, There is a bias in the selection of literature, because the reviews used to find studies all have their own selection processes. For example, Meijer's focuses on socio-economic status and health, which might, at least partly, explain the crowdedness of the upper left corner. Pickett & Pearl's selection process was similar to the one used here, but is older than Meijer's, and Yen et al.'s focused on studies that particularly addressed health studies which included older adults. Thirdly, papers might have been missed because of the terminology used here. Since this is a relatively new field, there is no consensus yet on the terms used, so it is very well possible that relevant papers were not included because they used different words to describe their research. This limitation was addressed by Riva et al. (2007) as well. Finally, most studies did not just focus on one cell in the table. Actually, some studies cover an entire column or row. Although all studies included in the table used individual characteristics as control variables, it still is wrong to see the different cells as independent of each other. For example, some researchers seem to have used informal reciprocity-related neighbourhood characteristics as proxies for economic characteristics. Similarly, the exclusion or inclusion of multiple variables in the model can generate very different results as opposed to a model that focuses on one characteristic in one cell only. Associations found between a certain neighbourhood characteristic and a health outcome in one cell should not be interpreted without considering results from other cells that are part of the same study as well.

Future research should be aimed at dissolving the issues mentioned above: investigating whether the institutional and physical environments are relevant neighbourhood characteristics in terms of health effects and whether this distinction between different environments is empirically useful in the first place, and, if so, how related characteristics influence health; focusing on context-based variables beyond SES and aggregated social capital in the economic and informal reciprocity domain (Macintyre, et al., 2002); and gaining more insight on neighbourhood effects on mental health and health (seeking) behaviour. Indeed, other reviewers have also highlighted the importance of enlarging insight into the effects of stress and other psychological processes in neighbourhood health effects (Diez Roux & Mair, 2010, p. 136). It is striking that although theorists focus on the 'reciprocal' explanation, almost all practical studies are examples of the 'contextual' explanation only. Because of this, it is hard to learn whether interactions implied by the 'reciprocal' models are real. More controlled experiments, or at least studies in a quasi-experimental setting that use natural (though probably not completely random) processes such as moving of people in and out of neighbourhoods, would help to isolate true neighbourhood effects and causalities. Finally, more attention should be paid to studying and formalizing interactions between the different environments, between the different health outcomes and between different population groups (gender, ethnic groups, different countries).

Challenges of measuring neighbourhood effects on health

Several researchers have tried to identify the main challenges faced by academics measuring neighbourhood effects on health. These challenges lie in the data collection and analysis as well as in theorizing the process itself, and can be summarized in five categories which I will discuss hereafter.

Defining and delineating neighbourhood. Defining 'neighbourhood' is problematic at both the conceptional and the operational level (Voigtländer, et al., 2013) (Pickett & Pearl, 2001) (Diez Roux & Mair, 2010). Researchers do not commonly express what their conceptional definition of the neighbourhood is, and where some see the neighbourhood first and foremost as a geographic area, others consider it to have a strong sociocultural aspect as well (Riva, et al., 2007). Chaix (2009), for example, defines the neighbourhood as an 'exposure area' that captures the potentially heterogeneous conditions of one's local environment, while Voigtländer et al. (2013) take a different, more relational approach by defining neighbourhood as 'the structure of the social ties of residents in an area who live in proximity to each other and who – to some extent – use the same facilities or participate in the same organisations'. On the operational level, most researchers of quantitative empirical studies use administrative boundaries of variable sizes, such as post codes, census tracts, school districts or even 'perceived local area within 20-minute drive from your home' (Bowling, 2006) to delineate neighbourhood (Voigtländer, et al., 2013) (Chaix, 2009). This is problematic for three reasons: first, the choice of operationalization is often made out of convenience and is not always based on theory (Voigtländer, et al., 2013); second and related to this, it is difficult to compare the results of different studies that use different operationalisations; and thirdly, administrative boundaries often lack intrinsic meaning because these boundaries may not reflect the interaction space of individuals, because interaction spaces cut across administrative borders or the scale is too large so that existing differences are dissolved (Chaix, 2009) (Riva, et al., 2007). However, having experience with doing research on these many different operationalisations allows future researchers to study how exactly geographic variations in health are influenced by size and intrinsic meaning of the area, in order to find out which delineations of 'the neighbourhood' are relevant for what purposes and under what conditions. An alternative to using administrative boundaries could be to use ego-centred boundaries, which aim to approximate neighbourhoods by drawing a buffer zone around individual's residences (Chaix, et al., 2009), either through a radius of a particular distance or through street network-based buffers.

Conceptualizing the causal pathway. As was mentioned above, in not all studies the choice of operationalization of neighbourhood was based on theory. This is a tendency that is apparent in different phases of the process as well (Riva, et al., 2007). Diez Roux (2004) and Diez Roux and Mair (2010) claimed that the use of the word 'effect' in describing 'neighbourhood effects on health' suggests a causality that might not be real, since the relationship between the two is not straightforward: on the one hand, clearly, neighbourhoods might affect health, but on the other hand, it is also realistic to believe that people select their neighbourhoods on the basis of a predisposition to certain (health related) behaviours. Diez Roux (2004) refers to this as 'self-selection'. Therefore, she suggests that more insight is required into the potential causal pathways between neighbourhood characteristics, neighbourhood composition and individual health – and the interactions and feedback loops between them. Such potential pathways exist for individual studies, but most of them have not directly been tested, nor are they combined with existing frameworks. For example, some authors suggest that neighbourhood-level social capital is positively associated with good health because neighbours can rely upon each other for mental support, thus reducing mental health problems, while other authors claim that neighbourhood-level social capital places excessive demands to provide support to others upon some members of cohesive groups, which causes stress and therefore has a negative influence on mental health (Kawachi, 2006). Another example of a causal pathway is that green space in the living environment is supposed to influence health in three ways: (1) through supporting recovery from stress and fatigue when exposed to green space, (2) by stimulating physical behaviour and social contacts and (3) through selective migration, related to SES (Maas, 2008). More a priori theories, hypotheses and experimental research are needed to allow for more rigorous evaluation of the associations between neighbourhood and health (Yen, et al., 2009).

Measurement of neighbourhood context. Measuring neighbourhood exposure, and the related data collection process, is inherently complex because of the complex nature and structure of neighbourhoods themselves. Earlier sections have highlighted that neighbourhoods consist of a physical environment, local markets, institutions and social capital; all of which have to be taken into consideration to fully understand health impact of neighbourhoods (Bernard, et al., 2007). Diez Roux and Mair (2010) have even called for attention to the synergetic effects between the physical and the social environment, which has been largely unexplored. In terms of operationalizing these characteristics, Diez Roux (2002) distinguishes between derived variables (summary or average characteristics of individuals in the same neighbourhood, such as mean income and unemployment rate) and integral variables (describing neighbourhood-level characteristics, often shaped by derived variables. Examples are quality of green space and social cohesion). Derived variables are closely related to the compositional explanation, whereas integral variables follow from contextual explanations. The reciprocal explanation can be operationalized by combining integral and derived variables. Measuring derived variables is less difficult than measuring integral ones, since derived variables can be constructed based on administrative sources such as population registries and census data, whereas quantitative assessment of the neighbourhood is required for constructing. Four important methods for constructing integral variables are: (1) self-reporting, (2) systematic social observation, (3) resident surveys, and (4) estimation based on georeferenced data (Raudenbush & Sampson, 1999). Self-reported health status comes with a clear bias which limits objectivity (Diez Roux & Mair, 2010). This is less so for the other three methods. Systematic observation is limited to those constructs which do not require resident's perspectives, and may thus be less useful for aspects regarding the social environment (Chaix, 2009). Surveys can be used in such cases. Georeferenced data is usually used for distance- and density measures regarding the physical environment, such as air quality and distance to and quality of green space (Chaix, 2009). Riva and colleagues (2007) claim previous research has focused largely on derived variables and call for an increase in the amount and quality of studies regarding integral variables. This is also apparent from the results of the review of experiments above, in which indeed most neighbourhood characteristics are derived contextual variables, as opposed to the integral, reciprocal variables suggested by Riva et al. (2007) and Macintyre et al. (2002).

Measuring individuals and households. While studying neighbourhood effects on health, it is still necessary to include variables on the individual- and household level as well, because such variables can work as confounders, effect modifiers or intermediary steps (Voigtländer, et al., 2013). Clearly, individual characteristics such as age and gender have an effect on the susceptibility of health risks that come with living in a certain neighbourhood, which means that neglecting these characteristics would not lead to reliable results. Similarly, including potentially confounding variables is necessary in order to control for self-selection (Riva, et al., 2007). Riva et al. (2007) point out that there is no consensus or consistency within this field on whether to address individual characteristics as confounders, modifiers or intermediaries, and call for a more theoretically founded approach to such variables. The role that is played by individual characteristics can also be studied in more detail through controlled experiments and longitudinal studies, in which the role of individual characteristics is one of the questions to answer in order to derive a causal pathway.

Modelling neighbourhood effects on health. The previous paragraph highlighted the importance of including individual and household variables when studying neighbourhood effects on health. This implies that variables of two, or even three, levels must be included in the modelling process. Another reason for including variables of different levels is related to what is in Voigtländer et al.'s work (2013) referred to as the 'ecological fallacy': effects of certain variables, which can be found in aggregated data at neighbourhood level, may or may not be the same as the corresponding effects at individual level. The process if further complicated by the fact that elements within and between these levels are interdependent: it is likely that there is a spatial correlation between neighbourhoods, just like it is likely that the characteristics of one household or individual are not independent from their neighbours' (Voigtländer, et al., 2013) (Pickett & Pearl, 2001) (Diez Roux, 2001). A useful way of approaching this difficulty is by using multilevel modelling. This type of modelling is useful when the data structure requires two assumptions to be made: (1) individual data is clustered within neighbourhoods, and (2) there is a correlation between persons living in the same neighbourhood (Voigtländer, et al., 2013). The framework of mixed regression models, in which a multilevel model can be embodied, can be expanded in order to include non-linear relations, heteroscedasticity, spatial models and effect modifications (where it is assumed that the effect of certain predictors varies according to other predictors of the same level) (Voigtländer, et al., 2013). Another aspect of the modelling process is related to statistical power and sample size. Riva et al. (2007) and Diez Roux and Mair (2010) point out that these have often been neglected, and should be considered more carefully in the future, so as not to hamper the ability to detect interactions in future works and allow future studies to compare the results of previous studies.

Neighbourhoods, individuals and time. Both neighbourhoods and the individuals that live inside them are not static. Individuals change and they move between neighbourhoods, which clearly has an effect on these neighbourhoods over time. Time, therefore, is an important aspect that ought to be considered when studying neighbourhood effects on health. The interactions between neighbourhoods, individuals and time can work in three distinct ways (Kuh & Ben-Shlomo, 2004): (1) accumulation, arguing that disease risks may accumulate over time and that, therefore, cross-sectional studies underestimate neighbourhood effects, (2) critical or sensitive periods, suggesting that exposures may have long-term health effects only if they occur within a specific period, creating a bias effect in cross-sectional studies because only a proportion of inhabitants will be exposed to neighbourhood effects that are similar to those in their critical period, and (3) trigger effects, describing a chain of risk where only the final link has a health effect. The bias created by these three pathways when doing a cross-sectional study, and the presence of multiple feedback loops between neighbourhood and individual over time (Diez Roux & Mair, 2010), highlights the potential role for longitudinal studies and studies that include historical neighbourhood data (Diez Roux & Mair, 2010) as opposed to cross-sectional studies (Riva, et al., 2007). Changes over time, such as the ageing of people or people moving from one neighbourhood to another, can be used to single out the effect of those (changing) circumstances on (changing) individuals. When it appears like the health of individuals moving from a neighbourhood with a certain characteristic tends to drop when moving to an area without this characteristic, this can be a reason to assume that there is an association between the characteristic and individual health. This can be an interesting new approach to studying neighbourhood effects.

Conclusion and Discussion

In this chapter, I have discussed three things. First of all, several useful frameworks that explain the theoretical pathway between neighbourhood characteristics and health outcomes were discussed. The ISIS-framework is a particularly useful framework, because it distinguishes between four different neighbourhood characteristics, it includes a temporal aspect and it links individual characteristics to the neighbourhood level. Additional layers explaining contextual trends and individual processes and a stronger focus on normative causal pathways would be useful additions to the framework. Secondly, I looked at how neighbourhood effects have been found to affect health in practice. I found that most attention has been paid to socioeconomic neighbourhood characteristics, while the other two environments identified in the ISIS framework, the institutional and physical environments, are often not considered separate environments. At this point, there seems to be a contradiction between theory and practice. It can be questioned whether these distinctions between environments are useful in the first place, because of their reciprocal and interconnected nature. Furthermore, general health outcomes and specific physical diseases have received a lot more consideration than mental health and health-related behaviour. Within health-related behaviour, health-seeking is particularly underrepresented, as the study done by Vallée and Chauvin (2012) was the only one that included in these reviews that did consider health-seeking behaviour. Finally, I found that there is a lack of attention to integral variables, as opposed to derived variables. The third and final topic I discussed here are the challenges that researchers in this field must face. From these findings, the following suggestions for future research can be derived.

In order to build up to a full understanding of neighbourhood effects on health that can be applied in questions of policy, future research should be based on a meaningful delineation of the neighbourhood (supported by theories such as the ISIS-framework) and work towards useful, causal, normative models. A way to study the consequences of the choice of delineation of the neighbourhood is to use several different delineations (postal code, school area, municipality) in the same study, and compare the results of the different delineations to see which effects are visible on which level. One way to start working on a causal pathway would be to study the movement of people from neighbourhood to neighbourhood and the consequence this movement has on their health. Finally, it would be interesting to study topics in this field that have currently not been dealt with sufficiently. Topics such as mental health outcomes, health-related (health-seeking) behaviour, social capital for community level and individual level separately, differences between groups within a neighbourhood, such as ethnic-, age- and gender groups, and finally international differences should be taken into consideration so that, in due time, it will be possible to create insights from different countries, and use this insight to generate useful, practical results that can be used as input for local and general policy processes.

Big Data and Neighbourhood Characteristics

That the term ‘Big Data’ refers to a relatively new field in science is illustrated by the fact that it did not appear in an abstract until 2009 (White & Breckenridge, 2014). Since then, the field has developed into what many authors call a hype (Zikopoulos, et al., 2015) (Gartner, 2015). While some authors claim that Big Data is past its peak (Gartner, 2015), others state that Big Data analytics mark an entirely new paradigm of computing and empirical research (Zikopoulos, et al., 2015) (Boyd & Crawford, 2012) (Shaw & McKay, 1942) marked by learning and interactive systems, data-driven epistemological research and ethics, as opposed to theory-driven, traditional research in previous eras (Zikopoulos, et al., 2015). This shift, and the rising importance of big-data computing, stems from advances in many different technologies, among which sensors, computer networks, data storage, cluster computer systems, algorithms and cloud computing facilities, but is intertwined with social, economic and political developments as well (Bryant, et al., 2008). Even though the field of Big Data is growing fast, there is a strong ongoing discussion on what this field actually entails and whether or not ‘Big Data’ as a term is helpful at all (Boyd & Crawford, 2012) (Power, 2014) (Russom, 2011). Critics state that Big Data as a technological category is becoming an increasingly meaningless term, because of its many different interpretations, and suggest that terms such as ‘unstructured data’, ‘process data’ (Power, 2014), ‘discovery analytics’, ‘exploratory analytics’ (Russom, 2011) would be better able to cover the load and meet expectations. This chapter will discuss Big Data within the context of social sciences, and healthcare policy more specifically. In the first section, I will discuss what Big Data is, or rather how Big Data is viewed by different influential authors. Secondly, I will characterize different kinds of Big Data studies in terms of data types, analysis types and outcome types. In this section, the focus will increasingly be on Big Data for healthcare policy. Thirdly, the most important methodological and societal strengths and weaknesses of using Big Data in social science will be discussed and fourthly, the theoretical insights from the first three sections will be used to find the practical potential of different online user generated data types for the analysis- and outcome types relevant in the context of this review. Finally, this chapter will be concluded by summarizing some of the tensions within the field of Big Data for social science.

One important controversy in the term is that it, indirectly, suggests that there was no such thing as ‘big’ data before this term was introduced (Russom, 2011). However, ‘big’ is a very relative term, especially when taking technological developments and Moore’s law into account (Russom, 2011). Another weakness of the term ‘Big Data’ is that it seems to suggest that the factor that differentiates ‘Big Data’ from ‘normal data’ lies in its size, whereas most academic definitions that are used in recent papers emphasize that really Big Data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets (Boyd & Crawford, 2012) (Zikopoulos, et al., 2015) (White & Breckenridge, 2014). The different definitions of the term that circulate today are on a continuum between these two extremes, but also on a continuum between seeing Big Data in a narrow, technical perspective on the one hand, and a more inclusive perspective on the other. For example, the definition of Big Data used by McKinsey (2011) is 'datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse', which is rather technical and focused on size. At the other side of the spectrum is Gartner's definition (2013), stating that 'Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making'. One example of a definition which recognizes the complexity of the data, but sees it as a technical phenomenon only is the definition used by IBM (Zikopoulos, et al., 2015), referring to Big Data as 'data sets whose volume, variety, velocity and complexity make it impossible for current databases and architectures to store and manage'. Boyd & Crawford (2012) give an example of a definition which does highlight the societal effects of Big Data, but only focuses on data volume: "We define Big Data as a cultural, technological and scholarly phenomenon that rests on the interplay of technology, analysis and mythology [a widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity and accuracy]". In the remainder of this chapter, Big Data will be considered as a cultural, technological and scholarly phenomenon – in accordance with Boyd & Crawford's definition - in which data is large in volume, variety, velocity and complexity – following IBM's definition. In the next section, the implications of this approach will be discussed in terms of different data, analysis and outcome types.

Big Data in social sciences

Even though this ongoing discussion on terminology makes careful analysis of literature more difficult, most scholars seem to agree that the data used in Big Data can be of several types (Russom, 2011): (1) structured (in tables, relational databases, record formats, etc.); (2) semi-structured (data that lacks a strict data model structure, but elements in the data can be identified by using tags of other types of markers); (3) unstructured (such as text expressing human language); (4) web data (blogs, tweets, social networks, logs, clickstreams); and (5) real-time (event data, spatial data, RFID, etc.). It must be noted that these categories are not mutually exclusive, especially because data types are often multidimensional and data types are often mixed in order to develop new analytic applications (Russom, 2011). Multidimensional databases can be used for three major types of analysis (Power, 2014): (1) retrospective data analyses, in which historical data and quantitative tools are used to understand patterns and results to make inferences about the future; (2) predictive data analyses, where simulation models are used to generate scenarios based on historical data to understand the future and making it known in advance; and (3) prescriptive data analyses, using planned, quantitative analyses of real-time data that may trigger events to recommend action. Within the social sciences, this combination of data types and analyses is often used for measuring population change, discovering sources of disparities, studying communication, language and linguistics and uncovering (the effect of) technologies, new media and social networks (Ovadia, 2013) (Gutmann & Friedlander, 2011). Cook & Collins (2015) identify five uses of data that are rather similar to these, although specified for improving policy making in healthcare: (1) providing population characteristics; (2) identifying risk factors and developing prediction models; (3) conducting observational studies comparing different interventions; (4) exploring variation between parties (healthcare providers) in order to detect outliers; and (5) as a supplementary source of data for another study.

Strengths and Weaknesses. The most apparent strengths of using Big Data in the social sciences are already implied in the ways the term is defined. Big Data has the potential to make sense of unstructured, complex and interdependent aspects of our world that require too much computational power or data collection intensity for 'traditional' social science, real-time, global, and at an unprecedented pace. This could increase the potential for social scientist to re-evaluate and re-operationalise social issues such as race, class, employment and networks in order to enhance and strengthen existing data, theory, methods and interpretations (Zikopoulos, et al., 2015). For example, the recommender systems employed by Amazon and Netflix are proof that Big Data analysis can reveal a lot of information on the preferences and behaviours of large populations, and even make predictions based on this. Some authors claim that Big Data can fundamentally change our definition of knowledge and lead to a computational turn in thought and research, transforming the activities of companies, researchers, medical practitioners, defence and intelligence, and the general population (Boyd & Crawford, 2012). These researchers claim that Big Data will overthrow the usual approach to science (hypothesizing, modelling and testing) and end the reign of theory, because algorithms can provide accurate forecasts at the cost of theoretical justification (Shah, et al., 2015). A more nuanced and perhaps more realistic vision is that, under the Big Data paradigm, theory will be generated through a combination of both inductive and deductive approaches (Boyd & Crawford, 2012) (Anderson, 2008).

On a more practical level, some authors state that Big Data has the potential of progressively eliminating selection bias, because the monitored (online) population tends to become equal to the general population, thus allowing for 'whole population analytics' with truly random and representative samples when there is full access to data (Cook & Collins, 2015) (Zimmerman, 2015). Similarly, behaviours that used to be offline are, increasingly, entering the domain of Big Data, thus allowing researchers to study interactions and behaviours and their variations over time and space that previously were not feasible (Zimmerman, 2015).

A related advantage is that often, these data sources involve naturally occurring social and digital media, which is not the case for surveys and experiments that are generally involved in social science research. This means that Big Data has the potential of providing researchers with data without any personal intervention by the researcher, thus potentially reducing researcher bias (Shah, et al., 2015) (although, of course, this only holds as long as those monitored do not know that they are being monitored). This is what Zimmerman (2015, p. 3) highlights when stating that 'the beauty of Big Data is that, unlike traditional survey data that are collected upon the consent of the individual and may suffer from several biases, they reveal the […] choices people make in the privacy of their home, and while they think they are under no observation'.

The advantages mentioned above, however, cannot be mentioned without recognizing that they are rather opportunistic, and indirectly point to important risks of using Big Data. Boyd & Crawford (2012) discuss several pitfalls related to using and interpreting Big Data analysis: (1) claims to objectivity and accuracy are misleading, since working with data is still subjective; (2) bigger data are not always better data, because it does not mean that methodological issues are no longer relevant; and (3) taken out of its context, Big Data loses its meaning. Data has no value in and of itself (Power, 2014). In terms of methodology, this implies that researchers still need to carefully consider data quality in terms of completeness, representativeness, comprehensibility and accuracy in terms of validity and reliability (Cook & Collins, 2015) in light of the purpose for which the data set was created – which is, in the case of Big Data, often not the purpose for which the data set is analysed (Cook & Collins, 2015). Even within one medium not every behaviour is equally representative, depending on the purpose of the analysis. Twitter, for example, is used as a way to interact with friends by some, and for news or professional purposes by others. Also, even more than in neighbourhoods, online populations migrate frequently, for example from MySpace to Facebook, which makes longitudinal studies of one medium difficult but also creates an opportunity to study structural differences between media.

Furthermore, extra caution must be taken regarding (causal) interpretation of data, especially since Big Data analytics tend to have a different relationship with theory and the boundaries between prediction and explanation, correlation and causation, have not always been clear in Big Data research before (Cook & Collins, 2015) (Shah, et al., 2015). Certainly, ethical and societal considerations should not be neglected when doing Big Data-related research. Boyd and Crawford (2014, p. 671) remark that "just because [data] is accessible does not make it ethical." The well-known trade-off between privacy and accessibility (Ovadia, 2013) requires careful social and political consideration on data governance, because it creates a risk of creating and maintaining new digital divides between the data-rich and the data-poor or the data-buyers and the data-sellers (Boyd & Crawford, 2012). The issue of consent and confidentiality in research is similarly striking, because researchers might not be in control of original data-collection activities, and respondents' consent may not have been obtained in a manner that is typically required for 'traditional' social science (White & Breckenridge, 2014). Privacy, accessibility, data ownership and consent each play an important role in the governance and analysis of, for example, cell phone data. Owners of this kind of data have a very valuable asset which can be used to study movement and interactions of large populations, for the purpose of knowledge, marketing, politics, et cetera (Leber, 2013), leaving very little agency in the hands of the individuals who create the data. Finally, it is important to realize that this field is not fully crystallized and that digital technologies quickly follow up on each other, which means that research on Big Data sources such as social media are at constant points of inflection, so that findings are quickly outdated (IBM) and current path-dependent decisions will shape the future (Boyd & Crawford, 2012).

Internet data and neighbourhood characteristics

This chapter focuses on the use of Big Data for social sciences, and the use of internet data within the context of neighbourhood characteristics in particular. Therefore, it is useful to know the kinds of questions related to this that can be answered by Big Data, and to what extent this has already been done. The following section will apply theoretical insights mentioned above in order to give a more practical overview of the potential of internet data for healthcare policy and neighbourhood characteristics. As neighbourhoods can be considered socially and geographically delineated populations, finding neighbourhood characteristics and using these to find health outcomes is part of Cook & Collin's first and second use of data for healthcare policy: providing population characteristics and identifying risk factors and developing prediction models. This section will discuss examples of questions and answers from these two areas, for different kinds of data discussed by Russom (2011) separately, in order to identify which data sources can be used to find relevant neighbourhood characteristics. A note must be made here that this distinction between data sources is somewhat arbitrary since no data sources are really 'unstructured', because even textual data has a certain inherent, linguistic structure that can, theoretically, be analysed, and the distinction between 'structured', 'semi-structured' and 'unstructured' is thus not very clear. This distinction is mainly used because it provides us with a useful way to identify and characterize different data sources, even within one medium.

Structured data. Most European governments, and an increasing amount of non-European ones, (Open Government Partnership, 2015) provide their citizens with open government data on a diverse set of topics that are more or less under government jurisdiction. The Dutch CBS (CBS, 2014) provides, among other things, details on the status of the economy, education, crime, energy use. Within the context of this review, its dataset on the composition of the Dutch population in terms of age, gender, ethnicity, household composition and employment for each post code in the country (CBS, 2014) is the most useful. Each of the 'neighbourhood effects on health'- studies described in the previous chapter uses a dataset of this kind to control for individual factors contributing to health. Other useful structured data sources for finding neighbourhood characteristics in the Netherlands are the property and geographical data provided by the Dutch land registry (Kadasterdata, 2015), which has previously been used in studies on the effect of land use, trees and green areas, on health (Maas, 2008); neighbourhood- and building-level energy use data provided by the Dutch ministry of infrastructure and environment's executive organisation (Rijkswaterstaat Leefomgeving, 2013) and data on public transport provision and district access (OpenStreetMap, 2015). No examples related to neighbourhood characteristics and social studies with respect to the latter two sources were found.

Unstructured data. Clearly, the internet, just like our platforms for offline interactions, is packed with unstructured data. Offline social interactions are unstructured in a way that they can really only be analysed through interviews, experiments and other intensive methods, but online interactions at least have a commonality in the way they are stored, which makes gathering and analysing their content a bit easier. The textual or visual content of social networking sites such as Facebook, Twitter and Instagram have been used to generate population characteristics and prediction models on various issues. For example, text- and sentiment analysis on Twitter has on several occasions served as a local source for ‘now-casting’ flu outbreaks (Bernardo, et al., 2013), election outcomes (McKelvey, et al., 2014) and earthquake effects (Crooks, et al., 2012). Although pictures posted on Instagram are even more difficult to analyse than textual updates, one research has used image recognition techniques to visualize sunrise and sunset patterns around the globe (Chandra, 2015). Another source of online unstructured data that can be used to find population characteristics and prediction models is online advertisement content. An analysis of online advertisements on websites like Craigslist has helped researchers from the University of Southern California to locate high-risk areas for the trafficking of minors (Wang, et al., 2012). These examples make clear that it is hardly useful to study unstructured textual or visual data without linking this to a more structured (geographic or relational) dataset, whether or not from the same medium, in order to clarify the context of the text or picture analysed.

Semi-structured data. Social networking sites have an inherent structure which can be of help when analysing its content. One fairly recent research attempted to take trend analysis of epidemics on Twitter one step further by combining textual analysis of Tweets with locational and network analysis based on the geotag of the Tweet and the network of the tweeter, in order to predict future victims of the flu by studying their relation to current victims (Sadilek, et al., 2012). Similarly, Facebook profiles, networks and updates were used to identify and track terrorism centrality nodes within regions (Ressler, 2006). Online sharing economy initiatives constitute a potentially interesting data source as well, although they are relatively unexplored within the context of local policy. Websites such as Peerby (a neighbourhood lending system where neighbours can offer and find services and products), ThuisAfgehaald and WeHelpen (similar services where neighbours can offer or ask for food leftovers or general services respectively) and Snappcar (a car sharing platform) might reflect or even build social capital within a neighbourhood (SocialTech, 2014) and can thus be used to associate lending activity with neighbourhood characteristics. All of these websites have a geographical and a textual component, linking services and goods provided to a particular postcode or address. Similarly, internet-based geographical information systems such as Google Maps and OpenStreetMap (OSM) link tagged points, lines and areas to their geographic location. Among other things, local facilities, land use, neighbourhood delineations and infrastructure can be derived from these sources, with clear potential for generating neighbourhood characteristics. Finally, cell phone data can give information on movement and location of individuals, and interaction between them. This kind of data has an increasingly large online component because smartphones make increasing use of apps that require WiFi, 3G or 4G connection and that store user data.

Web Data. As we are particularly interested in using internet data, all of the examples mentioned above are examples of web data as well. However, some aspects of web data which has hitherto been ignored are browsing behaviour and clickstreams on websites such as Google, Amazon and Netflix. Google search queries have been used to find flu-trends within (nation-level) areas (Bernardo, et al., 2013) (Eysenbach, 2009) and Amazon and Netflix have the data to generate useful user profiles which can be used for numerous goals. However, a downside of these kinds of data is that they are less accessible as they cannot be seen, or can only be seen until a certain limit, by users of the website that are not related to the maintainers and managers of the website.

Real-time data. All data sources mentioned here are updated within a certain interval. However, this interval may vary from several weeks or even years in the case of, for example, CBS data, to real-time, continuously updated data in the case of social networking sites.

Conclusion and Discussion

Within the field of Big Data, just like in the fields of many other innovations that were made possible by technological developments, there is still a lot of movement between the technological aspects of the innovation and the people concerned with these developments and the social world with which the innovation interacts. Here, developments in the field of computational and statistical technology are intertwined with social and economic implications, business, policy making and data governance. Big Data has been a part of efficient business organisation for a little while, but adoption by policy makers and social scientists seems to lag behind. A cause (and a consequence) of this might be that there simply are no established formal regulations, frameworks and status quo's yet on how Big Data technologies should interact with the social world, there is no consensus on the definition, scope and potential of Big Data and there is not enough experience with Big Data in the social sciences and policymaking to establish what Big Data can and cannot do, how such a project ought to be organized and how results found should be integrated in actual policy. This problem will not be easy to solve because of the big divide between the actors of the engineering world and those of the social world, where it seems like most engineers tend to focus on the possibilities whereas social actors seem to be rather shivery of Big Data's threat to privacy and consent. This is perfectly illustrated by Zimmerman's quote on the beauty of data, mentioned earlier in this review. The land between the technical and the social world remains to be undercultivated. This is surprising, because the creation of data is such a large part of our lives today and because some authors' claims towards a data-based paradigm shift. However, the absence of formal frameworks and regulations also leaves a lot of room for experimentation and creativity, visible from the large amount of Big Data startups (Withing, 2015), data-art (Urist, 2015), the development of new business models (ATKearney, 2013) and Big Data hackathons such as Orange's Data4Development challenge (Orange, 2015). If the goal is to organize this industry and develop insightful theories and models but also keep the entrepreneurial spirit and creativity, the technological and the social actors need to work together in Big Data research, policy making, development and entrepreneurship as well.

Big Data allows us to study phenomena that were invisible to the human eye before because of increased accessibility and representativeness of data, decreased bias and the transfer of 'offline' behaviour to 'online' behaviour. These new methodologies demand a new mind-set of the researcher employing them, and careful consideration is required in terms of data collection and ownership and the (often social) context in which the data is meaningful. Besides developing better methodologies, yet unanswered questions on sustainable data ownership regulations need to be addressed in future research as well.

In this chapter, I discussed several of the controversies around the topic of Big Data within the social sciences. I zoomed in from discussing Big Data within the social sciences in general, to specifically discussing internet data for providing population characteristics and using these to develop prediction models. A major limitation of the work done in this review is that, because of the lack of consensus on definitions and approaches, I cannot be sure that all relevant aspects were included in this review. Similarly, this review mainly, though not exclusively, includes scholarly work, even though a large share of the Big Data industry resides in the business world. This means that developments from businesses might be underrepresented in this review.

Conclusion and Discussion

Although none of the topics addressed in this review are particularly new, the combination of neighbourhood characteristics and health outcome and the combination of Big Data and social science and policy are both novel and relatively unexplored. The former chapters have tried to review the main themes in each of these new fields. It was found that neighbourhood effects on health are neither solemnly compositional nor contextual, but rather the physical, economic, institutional and informal reciprocal environments and overall health, mental health, health behaviours and disease prevalence are both nested and hierarchical domains that reinforce each other. Furthermore, it was found that although Big Data is a very lively and creative sociotechnical phenomenon, there are several main controversies around using Big Data in social sciences. There is no consensus on what it means to use Big Data, and there are trade-offs between privacy/consent and accessibility, between data ownership and data use, and between business, academia and policy. It seems like, in the field of Big Data in the social sciences, there is a lot of practice but hardly any established models of connections to policy, because knowledge is distributed, while for the field of neighbourhood characteristics and health there is not much interaction between theoretical models and practical experiments in the first place. The remainder of this chapter will be dedicated to reflecting on the findings of earlier chapters and on finding potential common ground for future works in the two fields explored in this review.

Both fields struggle with a lack of properly delineated definitions, theories, methods and models. On the one hand, this allows creative research and development in both fields. On the other hand, this prevents strategic and efficient knowledge creation. There are several opportunities for the two main fields to help each other to solve this problem. First of all, Riva et al. (2007) called for more attention to integral rather than derived neighbourhood characteristics, and Raudenbush and Sampson (1999) mentioned several (problematic) approaches to measuring such characteristics (self-reporting, which often comes with a objectivity-limiting bias; systematic social observation, which is limited to constructs that do not require residents' perspectives; resident surveys; and estimation based on georeferenced data, often limited to physical characteristics). There could be a rather clear role for Big Data in making these kinds of data collection less problematic. Social media or cell phone data, for example, has the potential of revealing self-reported, socially observed and georeferenced patterns of behaviour and attitudes without giving the appearance of being monitored (Shah, et al., 2015) and without researcher bias (Zimmerman, 2015). Similarly, Google Maps and OpenStreetMap can provide physical neighbourhood characteristics a lot quicker than any offline monitoring method could, and can even be used to track changes in neighbourhood characteristics over time, so that we can use this to single out specific neighbourhood effects. Online mapping tools, land registries, transport provision data and even pictures posted on social networking sites can provide physical environment characteristics, while textual posts, user networks and user profiles on social networking sites and content of online advertising and product sharing services can provide a lot of insight into the environment of informal reciprocity within a neighbourhood. Furthermore, online searching behaviour could be used to identify potential issues within an area's social or physical environment, and national open sociodemographic data can relatively simply provide a researcher with individual characteristics and derived neighbourhood characteristics. Hardly any of these sources have been used in any of the studies on neighbourhood characteristics discussed in this report. These applications can also be helpful in different fields of study, such as measuring the association between neighbourhood characteristics and sustainable behaviour by using online energy use registries. A clear recommendation for future research would be to use internet data sources such as those mentioned here to generate neighbourhood characteristics that are potentially associated with health. One way to do this, is to study health seeking behaviour in neighbourhoods (one of the underrepresented fields) by employing Big Data to derive neighbourhood characteristics that may be associated with health seeking behaviour, such as a neighbourhood's general willingness to support each other, expressed through its activity rate on sharing initiatives such as Peerby, or a neighbourhood's rate of green areas. This would involve a multilevel analysis with individual health-seeking behaviour as dependent variable, the neighbourhood's activity level on Peerby and the kind of products and services shared as independent variable, and individual characteristics such as income, background and status as confounding variables. Another approach to this question could be to track individuals who move from a neighbourhood with a certain level and type of Peerby activity to a neighbourhood in which this is different, and study how this changing environment affects them over a term of several years. This, of course, can only be done if the sample of moving individuals is large and diverse enough to actually single out neighbourhood effects.

Insights from these studies would be useful for the integration of social Big Data research and for local healthcare policy makers, but also for the further development and formalization of the studies of neighbourhood effects on health and Big Data in the social sciences.

References

Abel, T., 2008. Cultural capital and social inequality in health. *Journal of Epidemiology and Community Health, 62*, p. e13.

Anderson, C., 2008. The End of Theory: Will the Data Deluge Make the Scientific Method Obsolete?. *Edge: The Third Culture*, 30 June.

ATKearney, 2013. *Big Data and the Creative Destruction of Today's Business Models.* [Online] Available at: https://www.atkearney.com/strategic-it/ideas-insights/article/-/asset_publisher/LCcgOeS4t85g/content/big-data-and-the-creative-destruction-of-today-s-business-models/10192 [Geopend 02 September 2015].

Bernardo, T. et al., 2013. Scoping Review on Search Queries and Social Media for Disease Surveillance: A Chronology of Innovation. *Journal of Medical Internet Research, Volume 15, Issue 7,* pp. 1-13.

Bernard, P. et al., 2007. Health inequalities and place: A theoretical conception of neighbourhood. *Social Science & Medicine, 65,* pp. 1839-1852.

Bourdieu, P., 1986. The forms of capital. In: *Handbook of theory and research for the sociology of education.* Connecticut: Greenwood Press.

Boyd, D. & Crawford, K., 2012. Critical Questions for Big Data. *Information, Communication & Society, Vol. 15, No. 5,* pp. 662-679.

Bryant, R., Katz, R. & Lazowska, E., 2008. Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society. *Computing Research Association, Computer Community Consortium*, pp. 1-7.

CBS, 2014. *Statline: Bevolking en huishoudens; viercijferige postcode, 1 januari 2013.* [Online] Available at: http://statline.cbs.nl/Statweb/selection/?VW=T&DM=SLNL&PA=82245NED&D1=0,63-70&D2=0-80,4169,4485&HDR=T&STB=G1 [Geopend 06 August 2015].

Chaix, B., 2009. Geographic life environments and coronary heart disease: A literature review, theoretical contributions, methodological updates, and a research agenda. *Annual Review of Public Health, 30,* pp. 81-105.

Chaix, B. et al., 2009. Neighbourhoods in eco-epidemic research: Delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. *Social Science & Medicine 69,* pp. 1306-1310.

Chandra, M., 2015. *Synchronicity.* [Online]
Available at: http://michellechandra.github.io/synchronicity.html [Geopend 07 August 2015].

Cook, J. & Collins, G., 2015. The Rise of Big Clinical Databases. *Wiley Online Library,* pp. 93-102.

Crooks, A., Croitoru, A., Stefanidis, A. & Radzikowski, J., 2012. #Eartquake: Twitter as a Distributed Sensor System. *Transactions in GIS,* 17(1), pp. 124-147.

de Graaf-Ruizendaal, W. & de Bakker, D., 2013. The construction of a decision tool to analyse local demand and local supply for GP care using a synthetic estimation model. *Human Resources for Health,* p. 11:55.

Diez Roux, A., 2001. Investigating neighborhood and area effects on health. *American Journal of Public Health, 91,* pp. 1783-1789.

Diez Roux, A., 2002. A glossary for multilevel analysis. *Journal of Epidemiology and Community Health, 56,* pp. 588-594.

Diez Roux, A., 2004. Estimating neighborhood health effects: causal inference for a practicable social epidemiology. *Journal of Social Science and Medicine, 58,* pp. 1953-1960.

Diez Roux, A. & Mair, C., 2010. Neighbourhoods and Health. *Annals of the New York Academy of Sciences: the Biology of Disadvantage 1186,* pp. 125-145.

Ellen, I., Mijanovich, T. & Dillman, K., 2001. Neighborhood effects on health: Exploring the links and assessing the evidence. *Journal of Urban Affairs, 23(3-4),* pp. 391-408.

Eysenbach, G., 2009. Infodemiology and Infoveillance: Framwork for an emerging set of public health informatics methods to analyze search, communication and publication behaviour on the internet. *American Journal of Preventive Medicine, Volume 40, Issue 5,* pp. S154-S158.

Fogg, B., 2009. *A Behavior Model for Persuasive Design: proceedings of the 4th International Conference on Persuasive Technology.* New York, NY, ACM.

Frohlich, K., 2013. Area Effects on Behaviour and Lifestyle: The Spatiality of Injustice. In: *Neighbourhood Structure and Health.* New York: Springer Science+Business Media, pp. 39-60.

Galea, S., 2007. *Macrosocial determinants of population health.* New York: Springer.

Galea, S. & Vlahov, D., 2005. Urban Health: Evidence, challenges and directions.. *Annual Review of Public Health, 26,* pp. 341-365.

Gartner, 2013. *IT Glossary 2012: Big Data.* [Online] Available at: http://www.gartner.com/it-glossary/big-data/ [Geopend 03 August 2015].

Gartner, 2015. *Hype Cycle for Emerging Technologies 2015,* Stamford: Gartner.

Giddens, A., 1984. *The constitution of society.* Berkeley: University of California Press.

Godbout, J., 2003. *The world of the gift.* Montréal: McGill-Queen's University Press.

Gutmann, M. & Friedlander, A., 2011. *Rebuilding the mosaic: Fostering research in the social, behavioral, and economic sciences at the National Science Foundation in the next decade.,* Arlington: National Science Foundation, Directorate for Social, Behavioural, and Economic Sciences.

Hall, R., 2006. *Patient Flow: Reducing Delay in Healthcare Delivery.* Los Angeles: Springer Science+Business Media.

Kadasterdata, 2015. *Ons Productaanbod.* [Online] Available at: http://www.kadasterdata.nl/ [Geopend 03 August 2015].

Kawachi, I., 2006. Commentary: Social Capital and Health - making connections one step at a time. *International Journal of Epidemiology,* Volume 35(4), pp. 989-993.

Kuh, D. & Ben-Shlomo, Y., 2004. Introduction. In: *A life course approach to chronic disease epidemiology.* Oxford: Oxford University Press, pp. 2-14.

Larsen, E., 2013. People and Place: The Interrelated Connections between Interaction, Perception and Space. In: *Neighbourhood Structure and Health Promotion.* New York: Springer Science+Business Media, pp. 77-90.

Leber, J., 2013. *Technology Review: How Wireless Carriers are Monetizing your Movements.* [Online] Available at: http://www.technologyreview.com/news/513016/how-wireless-carriers-are-monetizing-your-movements/
[Geopend 03 September 2015].

Lin, K., Wu, S., Chen, L. & Ku, T., 2014. *Mining the user clusters on Facebook fan pages based on topic and sentiment analysis.* Redwood City, CA, IEEE 15th International Conference on Information Reuse and Integration (IRI).

Maas, J., 2008. *Vitamin G: Green environments - Healthy environments,* Utrecht: NIVEL.

Macintyre, S., Ellaway, A. & Cummins, S., 2002. Place effects in health: How can we conceptualise, operationalise and measure them?. *Social Science & Medicine, 55,* pp. 125-139.

Manyika, J. et al., 2011. *Big Data: the next frontier for innovation, competition and productivity,* sl: McKinsey & Company.

McKelvey, K., DiGrazia, J. & Rojas, F., 2014. Twitter publics: How online political communities signaled electoral outcomes in the 2010 US house election. *Information, Communication & Society,* 17(4), pp. 436-450.

Meijer, M., Röhl, J., Bloomfield, K. & Grittner, U., 2012. Do neighborhoods affect individual mortality? A systematic review and meta-analysis of multilevel studies. *Elsevier Social Science & Medicine 74,* pp. 1204-1212.

Molla, A., Biadgie, Y. & Sohn, K., 2014. *Network-Based Visualization of Opinion Mining and Sentiment Analysis on Twitter.* Beijing, IEEE International Conference on IT Convergence and Security.

Open Government Partnership, 2015. *Participating Countries.* [Online] Available at: http://www.opengovpartnership.org/countries [Geopend 07 August 2015].

OpenStreetMap, 2015. *OpenStreetMap.* [Online] Available at: http://www.openstreetmap.org/#map=14/51.4406/5.4444&layers=T [Geopend 07 August 2015].

Orange, 2015. *Challenge 4 Development.* [Online] Available at: http://www.d4d.orange.com/en/Accueil [Geopend 02 September 2015].

Ovadia, S., 2013. The Role of Big Data in the Social Sciences. *Behavioural & Social Sciences Librarian 32:2,* pp. 130-134.

Pickett, K. & Pearl, M., 2001. Multilevel Analysis of Neighbourhood Socioeconomic Context and Health Outcomes: A Critical Review. *Journal of Epidemiology and Community Health, Vol. 55, No. 2,* pp. 111-122.

Portes, A., 1998. Social capital: Its origins and application in modern sociology. *Annual Review of Sociology,* Volume 24, pp. 1-24.

Power, D., 2014. Using 'Big Data' for analytics and decision support. *Journal of Decision Systems,* pp. 222-228.

Raudenbush, S. & Sampson, R., 1999. Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology, 29,* pp. 1-41.

Ressler, S., 2006. Social Network Analysis as an Approach to Combat Terrorism: Past, Present, and Future Research. *Homeland Security Affairs,* 2(8).

Rijksoverheid, 2013. *Eerstelijnszorg.* [Online] Available at: http://www.rijksoverheid.nl/onderwerpen/eerstelijnszorg/grotere-rol-voor-eerstelijnszorg [Geopend 07 Augustus 2015].

Rijkswaterstaat Leefomgeving, 2013. *Lokaal Energie- en Klimaatbeleid.* [Online] Available at: http://www.rwsleefomgeving.nl/onderwerpen/lokaal_klimaatbeleid/publicaties/downloads/overzichten/ [Geopend 03 August 2015].

Riva, M., Gauvin, L. & Barnett, T., 2007. Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. *Journal of Epidemiology and Community Health, Vol. 61, No. 10,* pp. 853-861.

Russom, P., 2011. *TDWI Best Practices Report: Big Data Analytics,* sl: TDWI.

Sadilek, A., Kautz, H. & Silenzio, V., 2012. *Predicting Disease Transmission from Geo-Tagged Micro-Blog Data.* Rochester, NY, Association for the Advancement of Artificial Intelligence.

Shah, D., Cappella, J. & Neuman, W., 2015. Big Data, Digital Media and Computational Social Science: Possibilities and Perils. *The Annals of the American Academy of Political and Social Science 656,* pp. 6-13.

Shaw, C. & McKay, H., 1942. *Juvenile delinquency and urban areas; A study of rates of delinquents in relation to differential characteristics of local communities in American cities.* Chicago: University of Chicago Press.

SocialTech, 2014. *Peerby: A neighbourhood lending system.* [Online] Available at: http://socialtech.org.uk/projects/peerby/ [Geopend 07 August 2015].

Stock, C. & Ellaway, A., 2013. Neighbourhood Structure and Health Promotion: An Introduction. In: *Neighbourhood Structure and Health Promotion.* New York: Springer Science+Business Media, pp. 1-10.

Urist, J., 2015. *From Paint to Pixels.* [Online] Available at: http://www.theatlantic.com/entertainment/archive/2015/05/the-rise-of-the-data-artist/392399/ [Geopend 02 September 2015].

Vallée, J. & Chauvin, P., 2012. Investgating the effects of medical density on health-seeking behaviours using a multiscale approach to residential and activity spaces: Results frmo a prospective cohort study in the Paris metropolitan area, France. *International Journal of Health Geographics,* 11(1), pp. 54-65.

van der Gaag, M. & Snijders, T., 2004. Proposals for the measurement of individual social capital. In: H. Flap & B. Völker, red. *Creation and returns of social capital.* London: Routledge, pp. 199-217.

Voigtländer, S., Razum, O. & Berger, U., 2013. Methods to Measure Neighbourhoods and Analyse Their Impact on Health: An Overview. In: *Neighbourhood Structure and Health Promotion.* New York: Springer Science+Business Media, pp. 93-116.

Wang, H. et al., 2012. *Data integration from open internet sources to combat sex trafficking of minors.* New York, ACM.

White, P. & Breckenridge, R., 2014. Trade-offs, Limitations, and Promises of Big Data in Social Science Research. *Review of Policy Research, Volume 31, Number 4,* pp. 331-338.

Withing, R., 2015. *The 10 coolest Big Data startups of 2015 (So Far).* [Online] Available at: http://www.crn.com/slide-shows/data-center/300077457/the-10-coolest-big-data-startups-of-2015-so-far.htm [Geopend 02 September 2015].

Yen, I., Michael, Y. & Perdue, L., 2009. Neighborhood Environment in Studies of Health of Older Adults: a systematic review. *American Journal of Preventive Medicine 37(5),* pp. 455-463.

Zikopoulos, P. et al., 2015. *Big Data Beyond the Hype: a guide to conversations for today's data center.* New York: McGraw-Hill Education.

Zimmerman, N. A. K. F., 2015. The internet as a data source for advancement in social sciences. *International Journal of Manpower, Vol. 36 Iss 1,* pp. 2-12.

Appendix A: Overview of the characteristics of the reviewed studies

Table 5. Overview of the characteristics of the reviewed studies

First author, publication year, country	outcome variable and study years	sample size	individual characteristics	area-level characteristics	conclusion on area-effect after adjustment for individual characteristics
Adams et al. (2009), Australia	obesity, smoking, quality of life,	200 dwellings??	socio-economic status, obesity, smoking, quality of life, CVD risk, diabetes, physical activity, at-risk alcohol use	neighbourhood socio-economic disadvantage	association between socio-economic disadvantage and obesity, smoking and health related quality of life; no association between IRSD and CVD risk, diabetes, physical activity or at risk alcohol use
Albor et al. (2014), UK	mental health, 2001	4871 individuals	socioeconomic status, neighbourhood friends, mental health,	neighbourhood status; socioeconomic incongruity	for high status mothers, living in mixed or high status neighbourhoods reduces the odds of having no friends in the neighbourhood; living in high status neighbourhoods reduced the odds of depression or anxiety for high status mothers; no association for emotional support or self-esteem amongst highest status mothers; no association for lowest status mothers
Antai et al. (2010), Nigeria	Under-5 all-cause mortality, 1990, 1999, 2003	2118 children from 1350 mothers, from 165 communities	demographic and socioeconomic variables	urbanization, socioeconomic disadvantage	Urban under-5 mortality is significantly associated with urban area disadvantage. Wang et al: no association between neighbourhood income inequality, median household income or household-level income and self-rated health
Becares et al. (2013), New Zealand	Self-rated & mental health, racial discrimination. 2006-7			area deprivation, ethnic density	Increase in Maori ethnic density was associated with decreased odds of reporting self-rated health, doctor-diagnosed common mental disorders, and experienced racial discrimination.
Behanova et al. (2013), Netherlands and Slovakia	mental health, 2001				Association between mental health and area unemployment was strong in NL, but absent in Slovakia, although there are more mental health problems in Slovakia.
Cramm et al. (2012), South Africa	self-rated health,		crime experience, health status, social capital, demographic variables	socioeconomic neighbourhood status	individual-level subjective well-being is influenced by neighbourhood-level socioeconomic status(crime experience, SC and demographic vars)
Cunningham-Myrie et al. (2015), Jamaica	physical activity	2848	obesity, diabetes	neighbourhood disorder, perceived neighborhood safety, availability of recreational facilities	Significant clustering in physical activity levels of 20 min at least once a week, diabetes, and obesity across neighbourhoods. Greater levels of neighborhood disorder, home disorder, and counterintuitively recreational space availability were associated with higher levels of low/no PA among women.
Halonen et al. (2012), Finland	risk behaviour, 2004-8	60694 individuals	age, sex, marital status	population density, household income, education attainment, unemployment rate	relationship between neighbourhood disadvantage and co-occurrence of risk factors within each level of individual socioeconomic status
Hull et al. (2008), USA	Mental health, 1994-6	7863 individuals	neighbourhood interaction, religious participation, extracurricular activities, ethnic background, sociodemographic variables	neighbourhood disadvantage	Neighbourhood interaction/religious participation was salient for both white & Hispanic teens; religious participation for whites was moderated by neighborhood disadvantage. Non-sport extracurricular activities and employment were salient factors for black teens, moderated by neighbourhood disadvantage.

Table 5. Continued

First author, publication year, country	**outcome variable and study years**	**sample size**	**individual characteristics**	**area-level characteristics**	**conclusion on area-effect after adjustment for individual characteristics**
Ko et al. (2014), USA & Cuba	Self-rated health, 2004-5	1342 individuals	age, gender, marital status, education, financial strain, chronic conditions	proportion 65+; proportion of individuals below poverty, proportion of individuals with same racial/ethnic background	Those living in neighbourhoods with a higher proportion of residents below poverty were likely to report poorer health. Older age was only significant in Cuba, ethnic background equality only in Whites. Kwag: neighbourhood poverty predicted mental health perceptions.
Kwag et al. (2011), South Korea	self-rated health, mental health,	567 individuals in 233 census blocks	age, depressive symptoms, demographic variables	proportion 65+, proportion minorities, proportion below poverty level	neighbourhood poverty predicted mental health perceptions
Li et al. (2009), Taiwan	self-rated health, 1990, 1995, 2000	5784 individuals from 428 neighbourhoods	socio-economic and demographic variables	neighbourhood education, age structure, neighbourhood family structure and employment	indivs living in high concentration of youngsters, moderate education, moderate single-level families had significantly higher chances of having functional limitations and poor self-rated health than high education, medial single-parent families and media level of elderly people neighbourhoods
Maier et al. (2014), Germany	Diabetes & obesity, 2009-10	33,690	age, sex, BMI, smoking, sport, living with partner, education, diabetes, obesity	area level deprivation	For women, higher area level deprivation and lower educational level were both independently associated with higher Type 2 diabetes and obesity prevalence. For men, this was only found for obesity, not for diabetes.
Mohnen et al. (2014) Netherlands	self-rated health, 2005-8	1048 individuals from 259 neighbourhoods	individual social capital	neighbourhood social capital	both individual social capital and neighbourhood social capital at baseline were significantly associated with changes in self-rated health
Murayama et al. (2012), Japan	self-rated health, risk behaviour, 2009	4123 individuals from 72 city districts	sociodemographics, risk behaviour, individual social capital	aggregated social capital	higher district-level institutional mistrust was associated with self-rated poor health, but higher district-level mistrust in neighbours was inversely associated with it
Phillips et al. (2011), UK	Mental health,	4107 individuals	ability to manage income, employment, hope scale score, education level, demographic variables, neighbourhood satisfaction, interaction with friends	green space, land use, population density and turnover, crime, incivilities	Individuals who find it more difficult to manage their household income, and unemployed indivs, and those who speak less often with friends and neighbours and those who are dissatisfied with their neighbourhood have lower hope scale scores. Physical environment neighbourhood aspects were not associated with positive mental health
Riva et al. (2011), UK	Mental health, 2004	12962 individuals from 892 areas	demographic variables, employment	area deprivation, social cohesion	living in rural areas is associated with lower risk of reporting common mental health problems; the mental health advantage of employment is larger in rural areas; rural areas are associated with better mental health
Uthman & Kongnyuy. (2008), Nigeria	risk behaviour, 2003	6362 individuals	sexual activity, marital position, alcohol use, concurrent partners, household income	neighbourhood status	25y+ women less likely to have reported multiple concurrent sex partners in the last 12 years; currently/formerly married women less likely to have multiple concurrent sex partners; women who drank alcohol in the last three months were more likely, too; women from poorest and middle households as well.

Table 5. Continued

First author, publication year, country	outcome variable and study years	sample size	individual characteristics	area-level characteristics	conclusion on area-effect after adjustment for individual characteristics
Vallée et al. (2012), France	health-seeking behaviour, 2005-10	662 individuals from 50 census blocks	incidence of delayed screening, individual activity space	residential medical density	no sig association between residential medical density and incidence of delayed cervical screening; sig interaction between individual activity space and residential medical density; women living in low medical density neighbourhoods had a sig high er risk of delayed screening, but only if they reported that their daily activities were centred within their neighbourhood of residence
Voigtländer et al. (2010), Germany	Self-rated health, 2004		perceived distance to facilities, demographic variables	unemployment quota, average street section purchasing power, air pollution, noise	sign association between area deprivation and physical health; can be explained partly by specific physical features of neighbourhood environment
Wong et al. (2009), Hong Kong	self-rated health, 2002, 2005	25623, 24610 individuals from 287 areas	socio-economic variables, demographic variables	income inequality, median household income	no association between neighbourhood income inequality, median household income or household-level income and self-rated health
Wu et al. (2010), Canada	self-rated health, 2008	3421 individuals from 148 schools	demographic variables, family education	neighbourhood satisfaction providers, facilities	Children from families with higher educational attainment and in neighbourhoods that provide good satisfaction and facilities reported higher health-related quality of life