Summary Measures and Determinants of Small-Area Population Health

M.F. Jonker

Summary Measures and Determinants of Small-Area Population Health

This thesis was printed with financial support of the Department of Public Health, Erasmus MC and the Erasmus University Rotterdam.

Summary Measures and Determinants of Small-Area Population Health

Indicatoren en determinanten van populatiegezondheid in kleine geografische gebieden

Proefschrift

ter verkrijging van de graad van doctor

aan de Erasmus Universiteit Rotterdam

op gezag van de rector magnificus

Prof.dr. H.A.P Pols

en volgens het besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 10 februari 2016 om 11.30 uur

door

Marcel Jonker

geboren te Voorburg

Erasmus University Rotterdam

Promotiecommissie

Promotoren:

Prof.dr. J.P. Mackenbach

Prof.dr. A. Burdorf

Overige leden:

Dr. D. Rizopoulos

Prof.dr. P.P. Groenewegen

Prof.dr. A.E. Kunst

Copromotor:

Dr. F.J. van Lenthe

Table of Contents

1

Introduction

1.1 BACKGROUND

Urban policy is a major route through which governments and health authorities attempt to deliver improvements in living conditions, economic opportunities, and public health. Relevant policies comprise, for example, large-scale urban regeneration and neighborhood renewal programs but also various local community-based initiatives. These policies are often area-based and typically involve investments in key socioeconomic determinants of health (such as income, employment, and housing conditions), combined with efforts to create a diverse socioeconomic mix of residents as well as initiatives to promote healthy life styles.[1]

The use of area-based urban programs is based on several factors. One important rationale is the spatial concentration of poor individuals, which implies that area-based targeting can be an effective way of reaching poor individuals in certain neighborhoods, thereby offering high levels of 'completeness' and 'efficiency'.[2] Secondly, concentrated deprivation is often thought to result in 'neighborhood effects'; i.e. negative effects on residents' life chances (e.g. in economic opportunities and health) over and above the effect of their individual characteristics.[3] This calls for additional investments directly into the neighborhoods of greatest need, both to support individuals and the general infrastructure in deprived neighborhoods.[4] Thirdly, area-based programs are often attractive from a political perspective. They are well-suited as a means to ration scarcely available resources, pilot innovative interventions, and signal political commitment to remedy larger socioeconomic and specific public health problems. Finally, area-based programs combine well with recent trends in 'new public management'; they usually allow for an integrated approach to tackling broad problems and also allow for a substantial role for community participation.[5]

Not all rationales for area-based programs are uncontested. For example, despite the usually substantial spatial concentration of poor individuals, spatial segregation is incomplete and hence some degree of inefficiency is inherently associated with area-based programs. That is, area-based programs inevitably target areas with inhabitants of mixed socioeconomic status,[2,6,7] which implies that a careful balancing of the advantages and disadvantages of area-based targeting is necessary. Moreover, despite the appealing simplicity of the neighborhood-effects hypothesis, there is surprisingly little evidence of spatial deprivation-amplification or contextual effects.[8,9] Thus far, most studies investigating neighborhood effects only identify correlations between individual outcomes and their environment and do not take into account that selection into neighborhoods is a non-random mechanism (see e.g. Oakes[10] and Van Ham and

Manley[11]). Also, without substantial neighborhood amplification effects, the benefits of large-scale area-based urban programs will be considerably smaller.

Irrespective of the absence of (scientific) consensus about the rationales for area-based urban programs, they have been widely implemented in countries such as the Netherlands, the United Kingdom, Canada, Germany, France, Denmark, Finland, and Sweden.[4,12] Accordingly, the lack of scientific consensus is, at least by policymakers, not considered as an important impediment to the implementation of area-based programs. The same conclusion holds for the inherent problems that are associated with reliably quantifying and summarizing population health at the small-area level. Due to the small populations and typically limited health information for individuals within each area, standard methodology to measure population health cannot be used for most area-based programs. This makes it difficult to measure, monitor, and reliably evaluate the impact of area-based programs in terms of population health.

Of course, for those programs not directly geared towards improving average population health, the absence of methodology to quantity population health may not be a problem. However, many area-based programs do include ambitions to affect and improve population health. For these programs, the availability of better methodology to measure average population health would be useful. It would, for example, allow for the initial selection of areas with most severe health problems and for the monitoring and evaluation of health impacts of area-based urban programs in a more reliable and consistent manner than with an ad-hoc selection and aggregation of indicator variables (e.g. Rothenberg et al.[13]). It would also allow for more comprehensive evaluations of area-based programs than, for example, by following trajectories of self-rated health outcomes or health-related behaviors.[14–16]

Additionally, the reliable quantification of population health at the small-area level would accommodate new investigations into the spatial pattern of small-area population health and its determinants, thereby opening-up new research possibilities and hopefully more efficient entry-points to improve population health and reduce health inequalities.[17–19] Such analyses would nicely fit into a rich history of epidemiological investigations into the determinants of spatial inequalities in health.[42] It could also make use of more sophisticated methodology based on the synthesis of ecological and individual-level data to further improve the explanation of spatial inequalities in health.[20,21] As such, it could extend upon existing evidence that shows that well-designed and well-resourced area-based interventions generally have scant impact on average population health[15,16,22] but can be effective at reducing, or at least

preventing the widening of, social inequalities in health outcomes (see e.g. Magnee et al.[23]).

Following these considerations and as part of an Erasmus MC Department of Public Health initiative to stimulate methodological improvements in research aimed at reducing urban health inequalities, conducted under the umbrella of CEPHIR (an acronym for Centre for Effective Public Health In the Rotterdam area), this thesis will address the following three research aims:

1. the first research aim is to develop and validate new methodology to estimate small-area indicators of population health;

2. the second research aim is to investigate the feasibility and merit of the selected indicators to map, rank, and select areas for policy purposes;

3. the third research aim is to explain geographic (small-area) differences in population health making use of the developed methodology.

Before turning to these research aims, the remainder of this chapter explains the concept of summary measures of population health (SMPH) together with the required data and their calculation in section 1.2. Subsequently, geographic variation in SMPH and the relationship to average income and socioeconomic status is covered in section 1.3, the difficulty of estimating SMPH for small geographical areas is described in section 1.4, and the benefits of Bayesian statistics explained in section 1.5. Finally, section 1.6 concludes with an overview of the remainder of this thesis.

1.2 SUMMARY MEASURES OF POPULATION HEALTH

Typology

Summary measures of population health (SMPH) combine information about the mortality and morbidity conditions of a population into a single numerical index, using a scale that remains comparable over time and using an aggregation method that provides adequate attention to both mortality and morbidity. SMPH can be divided into 2 broad categories: health expectancies and health gaps.[24] To explain

the difference between both categories, the bold curve in Figure 1.1 shows the survivorship curve of a hypothetical population. This curve indicates, for each age along the horizontal axis, the fraction of an initial birth cohort that is still alive at each age.

The life expectancy at birth (LE) of the population is then defined as the combined grey area (i.e. area A + area B) under the survivorship curve:

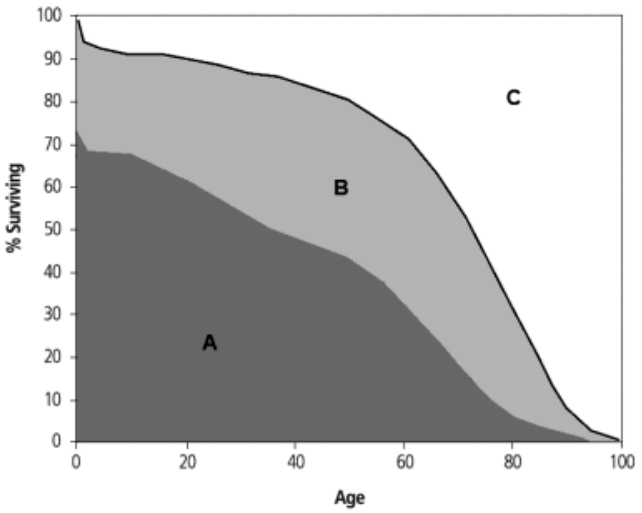
$$\text{life expectancy} = A + B \tag{1.1}$$

Health expectancies are measures of the same area; however, health expectancies assign full weight to years lived in perfect health (i.e. area A) and some lower weights to years lived in health states worse than perfect health (i.e. area B). More formally:

$$\text{health expectancy} = A + f(B), \tag{1.2}$$

where $f(\cdot)$ is a function that assigns weights to health states less than full health using a scale on which perfect health equals 1. Accordingly, there are as many health expectancies as there are definitions of perfect health and functions $f(\cdot)$ to assign weights to health states in less than perfect health.

Figure 1.1 The survivorship Curve of a hypothetical population]



Source: Murray et al.[24]

In contrast to health expectancies, health gaps quantify the difference between the actual health of a population and some benchmark or stated goal for population health.[25] As an example, in Figure 1.1 the implicit health goal is for everyone in the population to live (in perfect health) until the age of 100, corresponding to the vertical line enclosing area C on the right. Similar to health expectancies being an extension of life expectancy, health gaps are an extension of the mortality gap. The latter measures the years of life lost due to premature mortality (i.e. area C), whereas health gaps additionally account for years lived in health states worse than perfect health:

$$\text{health gap} = C + g(B). \tag{1.3}$$

Here $g(\cdot)$ is a function that assigns weights to health states less than full health, using a scale on which a weight of 1 implies that years lived in morbidity are equivalent to time lost due to premature mortality.

Health expectancies and health gaps thus provide similar information about a population's health, albeit from a different perspective. The choice for either category of SMPH usually depends on the need (or desire) to explicitly correct for the age structure of the underlying population. By construction, health expectancies are independent of the age structure of the population. In contrast, health gaps are usually calculated in absolute terms, reflecting the population's absolute number of healthy life years lost. This way, health gaps depend on the age structure of the underlying population, which is convenient for decompositions of life years lost by various causes of mortality and morbidity. Nonetheless, also health gaps can be directly age standardized, which implies that the decision whether or not to correct for the population age structure ultimately depends on the intended use of the SMPH and the information it has to convey. In this thesis, cause-specific decompositions and the quantification of health gaps relative to a stated goal are not of direct interest, which explains the subsequent focus on health expectancies – even though the methodology required to calculate health gaps is essentially the same.

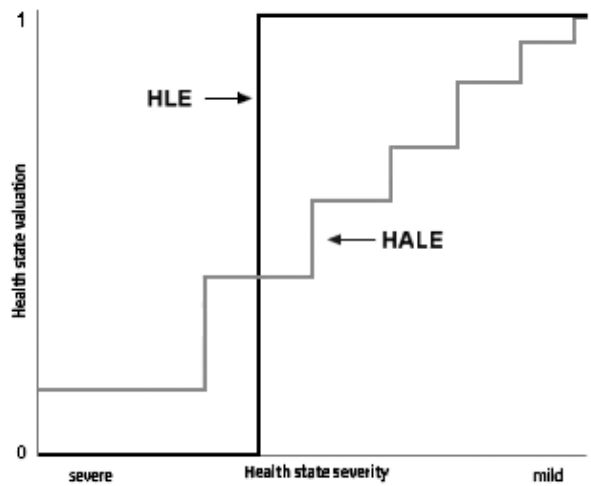
Dichotomous vs. polychotomous weighting

The definition and valuation of health states worse than perfect health (i.e. area B in Figure 1.1) defines the SMPH at hand. Many different options are available. The simplest option is a dichotomous definition and weighting scheme (see Figure 1.2), which constitutes the healthy life expectancy family of SMPH. With healthy life expectancy—in this thesis abbreviated as HLE—health states in less than perfect health are up to a certain threshold treated equal to perfect health (i.e. weighted with

a weight of 1) and beyond this threshold treated as the equivalent of being death (i.e. weighted with a weight of 0). As mentioned by Murray et al.,[24] the dichotomous valuation of health-states makes HLE measures sensitive to variation in and the correct interpretation of the threshold definition, which complicates comparisons between populations and assessments of changes in health over time.

In contrast, health-adjusted life expectancy measures—in this thesis abbreviated as HALE—are based on a polychotomous quantification and valuation scheme (see Figure 1.2). The latter implies that the entire health spectrum is divided into a large number of health states, each of which are weighted according to their health state severity using a comprehensive weighting function. For the calculation of HALE, there are 3 main generic health-state instruments available: the EQ-5D,[26,27] the HUI-3,[28,29] and the SF-6D.[30,31] Each of these instruments provides a descriptive profile of the health spectrum with an accompanying weighting function, which provides weights on a scale anchored at 0 (immediate death) and 1 (perfect health). The estimation of HALE thus requires a population based survey that includes one of the health state measurement instruments. These survey data are not as widely available as the required input data for a dichotomous weighting scheme, but HALE provides a more detailed and balanced measurement of average population health than HLE.[24,25]

Figure 1.2 Valuations of time spend in health states less than perfect health



Source: Murray et al.[24]

Dutch input data

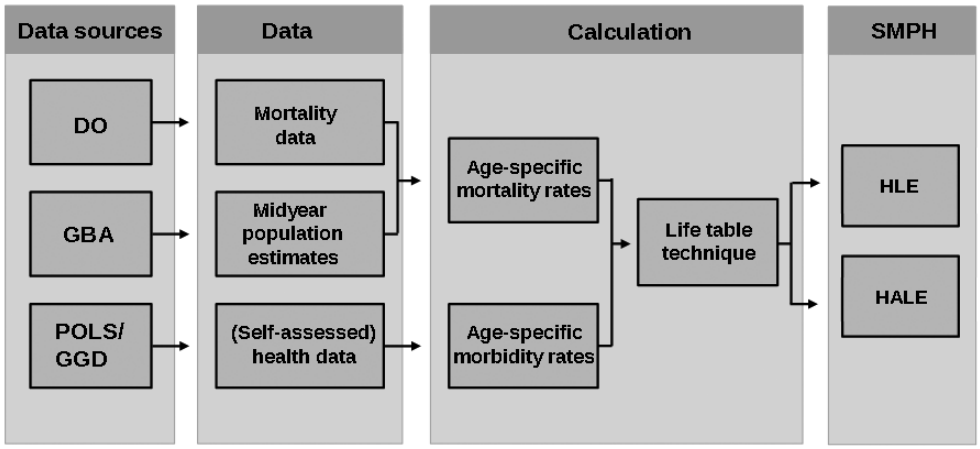
Dutch HLE and HALE measures can be calculated using the input data and standard life table technique as indicated in Figure 1.3. Starting with the mortality component, the required age-specific mortality rates need to be calculated using data obtained from 2 different sources: first, midyear population estimates as obtained from the Dutch population registry (GBA), and secondly, mortality data as obtained from the Dutch deaths registry (DO) or GBA. Both registries are maintained by Statistics Netherlands and are available for scientific research via the secure remote-access infrastructure of Statistics Netherlands.[32] In practice, this means that researchers access the data using a chipcard-reader and biometric fingerprint scan and process the data directly on the servers of Statistics Netherlands. Secondly, the required age-specific morbidity rates are usually derived from self-assessed health survey data. For HLE, required survey data can be obtained from the national health survey (POLS, or since 2010, GECON) as administered by Statistics Netherlands or from any other survey that contains one or more health questions that can constitute a dichotomous weighting scheme. For HALE, the POLS survey contains information for one generic health-state instrument, i.e. the SF-6D. The same SF-6D is also included in several health surveys from the local health authorities (GGD). Large-scale survey data that comprise the EQ-5D or HUI-3 are currently unavailable (in the Netherlands).

Life table calculations

As shown in Figure 1.3, the age-specific mortality rates and morbidity prevalences are combined in a single SMPH index using the life table technique. This technique provides a standardized approach to a) derive the survival curve of a population from the age-specific mortality rates, and b) calculate the area under the survival curve – which, depending on the weighting scheme used for years spend in less than perfect health, results in LE, HLE or HALE.

The first step in the life table technique is to derive the survivorship curve from the observed age-specific mortality rates. Various approaches have been proposed in the literature, which are all slightly different in their distributional assumptions regarding the time of death in the age intervals. Fortunately, the exact distributional assumption has little impact on the accuracy of the life table estimates (see e.g. Schoen[33] and Golbeck[34]) and in applied work, the Chiang life table approach[35] is considered the golden standard. This approach is summarized as follows. First, the

Figure 1.3 Schematic framework for the estimation of (Dutch) HLE and HALE*



** SMPH = Summary measure of population health*
DO = causes of death registry, GBA = Dutch population registry,
POLS = Dutch national health survey, GGD = Dutch local health authorities survey

conditional probabilities of death are derived from the age-specific mortality rates:

$$q_x = n_x \times M_x / [1 + (1 - a_x) \times n_x \times M_x] \tag{1.4}$$

with q_x denoting the conditional probability of death in age group x, and n_x, M_x, and a_x denoting the width, the observed mortality rate and the fraction of the age interval lived by persons who die in age interval x, respectively. Usually, a_x is set at 0.1 for the first age group and 0.5 for all other age groups (the latter corresponding to a uniform distribution of deaths in the age interval) and the width of the age intervals n_x is set depending on the population size. So-called "complete" life tables are based on one-year age intervals and require relatively large populations, whereas so-called "abridged" life tables are based on wider age categories that are more applicable to smaller populations. Abridged life tables usually comprise age groups 0, 1-4, 5-9, 10-14, 15-19, 20-24, etc.

The survivorship curve is then derived for a hypothetical cohort by exposing an initial cohort population (l_0) to the age-specific conditional probabilities q_x as calculated using equation 1.4. Starting with:

$$l_0 = 100,000 \text{ persons at birth (i.e. } 100\%), \tag{1.5}$$

the remaining survivors for the subsequent age groups in the life table are calculated as:

$$l_{x+1} = l_x \times (1 - q_x). \tag{1.6}$$

The second step of the life table technique is to calculate the area under the survivorship curve and apply a set of weights to obtain HLE or HALE. These calculations are relatively simple: for each age group in the life table, the area under the curve is calculated using:

$$L_x = n_x \times [l_{x+1} + a_x \times (l_x – l_{x+1})], \tag{1.7}$$

which is the average height of the age group multiplied with the width of the age group (i.e. n_x). Once the area under the survivorship curve for each age group is calculated, LE at birth can be determined as:

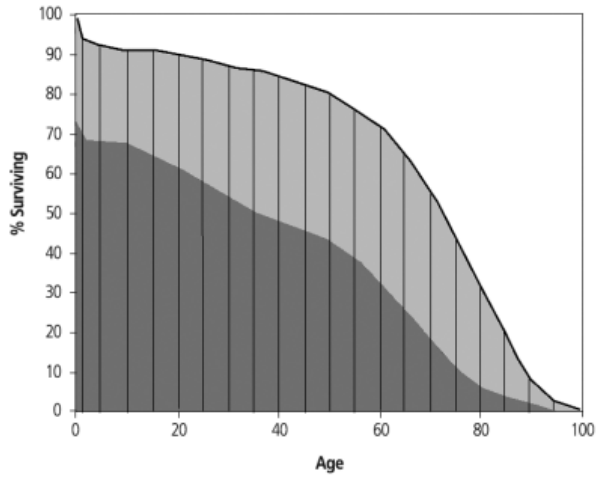
$$LE = \sum L_x / l_0, \tag{1.8}$$

and HLE and HALE at birth can be calculated as:

$$HLE \; or \; HALE = \sum [L_x \times w_x] / l_0, \tag{1.9}$$

where w_x denotes the set of age-specific weights to adjust for years lived in less than perfect health. Figure 1.4 provides a graphical representation of the calculations for a 5-year abridged life table with 95+ as the final age group.

Figure 1.4 Calculation of the area under the survivorship curve



1.3 GEOGRAPHIC VARIATION IN POPULATION HEALTH

International comparisons based on the previously described life table methodology highlight the existence of substantial global inequalities in population health. In 2010, for example, LE and HLE among 186 countries worldwide [1] were estimated to range from approximately 46 years LE and 40 years HLE in the Central African Republic to approximately 83 years LE and 73 years HLE in Japan.[43] This observed geographic variation is strongly correlated with average income levels. More than 50% of the variation in global LE and HLE is explained by countries' national income per capita as defined in 2010 purchasing power parities [44] (source: own calculations).

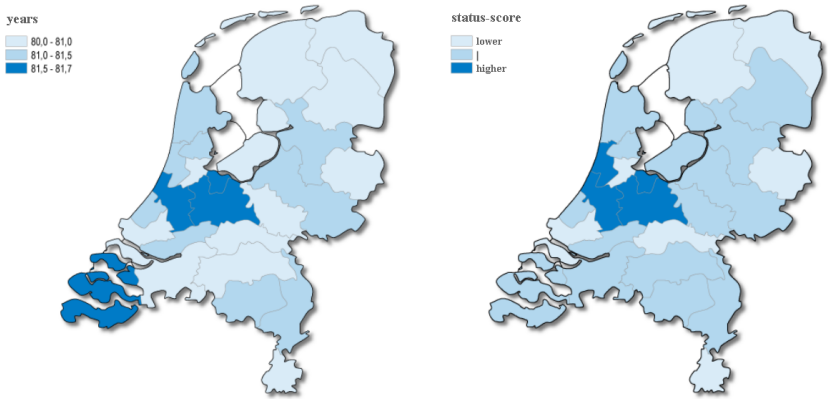
For the subset of European countries (excluding Russia), LE in 2010 was estimated to range from 70 to 82 years and HLE from 61 to 71 years.[43] Irrespective of the smaller range of LE and HLE, the observed geographic pattern is again strongly correlated with average income levels. In fact, at the European level, more than 70% of the variation in LE and HLE is explained by the European countries' average income per capita (source: own calculations).

Of course, many other determinants are correlated to average income, which is why more refined exploratory studies typically take additional determinants of population health into account. However, given the intra-country variation in and correlations between the explanatory variables, ecological studies are poorly suited to obtain causal estimates of the underlying determinants of population health. The solution, from a statistical perspective, is either to explicitly account for the (joint) within-area distribution of the explanatory variables or to reduce the spatial scale of the analysis—for example, to the regional, municipality or even neighborhood level—in order to obtain more homogeneous areas and hence to more reliably conduct ecological analyses.[45]

Starting with the reduction of the spatial scale of analysis, at the national level, the Dutch LE in 2010 was 81 years and HLE 69 years (i.e., exactly at the average of all Western European countries).[43] At the regional level, Dutch LE and HLE has a range of 2 and 7 years, respectively, when based on the 25 local health authority regions (GGD-regio's).[46] This is a considerably smaller range than observed in other countries

[1] Note: these estimates exclude Haiti, which experienced an earthquake in 2010 with major mortality consequences (resulting in a LE of 38 and HLE of 32 years).

Figure 1.5 Dutch regional life expectancy (left) and socioeconomic status (right) per local health authority region (in 2010)



Source: Public Health Status and Foresight Report 2014[46]

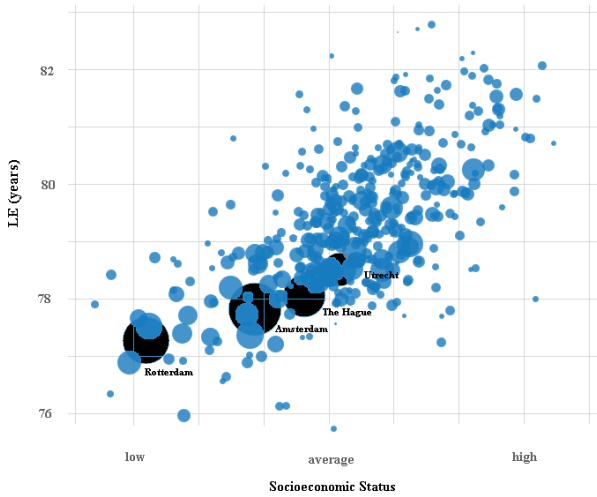
at the regional level (see e.g.[47,48,75,98]) although such comparisons may be confounded by differences in geographic scale. Also, as can be seen from Figure 1.5, the observed variation in LE (and HLE) is still strongly correlated to variation in average income and socioeconomic status.

At the municipality level, differences in LE (and HLE) are somewhat larger than at the regional level, with LE ranging roughly from 76 to 83 years for males and 80 to 88 years for females.[49] Figure 1.6 shows that inhabitants of the larger cities generally have lower LE than those of smaller municipalities, with inhabitants of the city of Rotterdam having by far the lowest LE of the "Big-4" Dutch cities (i.e. Amsterdam, Rotterdam, Utrecht, and The Hague). Furthermore, Figure 1.6 clearly shows that socioeconomic status is strongly correlated with the observed differences in LE at the municipal level.

Standard life table methodology cannot be used to obtain reliable estimates of LE and HLE at the neighborhood level (due to sparse data problems; see the subsequent section). Nonetheless, Figure 1.7 depicts the average socioeconomic status for neighborhoods in Rotterdam in 2010. As shown, at the smallest spatial level there are again substantial differences in socioeconomic status. Given the observed correlations between average income and socioeconomic status at larger spatial scales, this is at

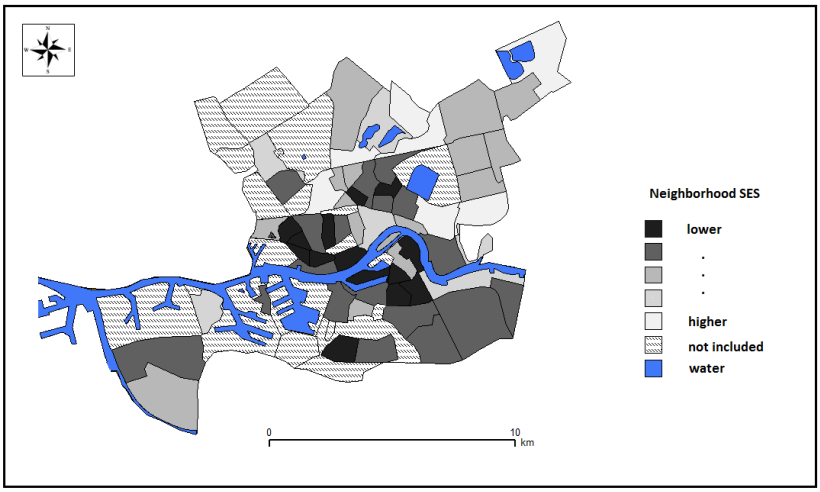
least indicative of substantial variation in small-area population health. Additionally, a broad range of multi-level studies (based on individual-level rather than area-based health outcomes) has shown that the composition of neighborhoods' populations in terms of socioeconomic status provides a substantial contribution towards explaining observed variation in urban health outcomes. Moreover, small-area disease mapping studies typically show substantial variation in small-area (standardized) mortality patterns. Accordingly, all existing evidence points towards substantive geographic variation in small-area population health, which also forms an important rationale for the area-based targeting of health policy as described in section 1.1. As such, the research aims of this thesis, which comprise both the estimation of summary measures of population health and the investigation of determinants of variation in small-area population health, nicely extends upon the existing evidence base regarding health inequalities and their determinants at the small-area level.

Figure 1.6 Association between male life expectancy and socioeconomic status in Dutch municipalities (2010)



Adapted from the Public Health Status and Foresight Report 2014 [49]

Figure 1.7 Neighborhood socioeconomic status in the municipality of Rotterdam (2009)



Source: own calculations

1.4 SPARSE DATA PROBLEMS AT THE SMALL-AREA LEVEL

As mentioned, the life table technique has originally been developed (and works well) for application in large populations, such as those of countries, regions, and municipalities. Without modifications, however, the standard life table technique is not particularly suited for the calculation of LE and health expectancies in smaller populations. The main reason is that the traditional life table technique calculates the required age-specific mortality and morbidity prevalences separately for each population, and within each population independently for all age groups, which requires a large number of deaths and survey respondents. Such large numbers are generally unavailable in small populations, resulting in considerable sampling variation.

Standard approaches to handle sparse life table data involve the aggregation of smaller geographical areas, the pooling of data over several years, the use of abridged instead

of complete life tables, and the closure of life tables at younger age groups. Each helps to reduce the sampling variation, but even combined they cannot fully accommodate the sparse data problems that are typically encountered in small-area analyses. First, for populations smaller than approximately 50,000 persons, the normality assumption that constitutes the calculation of standard errors and 95% confidence intervals no longer holds, resulting in an under-estimation of the true uncertainty in the estimate of the area under the survivorship curve.[36] Second, and more importantly, the life expectancy estimates themselves become unreliable: the smaller the population, the larger the bias of the LE estimates.[36,37]

Hence, to reliably estimate small-area LE and HLE/HALE, more advanced statistical methodologies are required that pool strength and recognize correlations between the various dimensions of the life table data –for example, between sexes, age groups, and contiguous areas– and thus no longer calculate the age-specific mortality and morbidity prevalences independently for each sex, area and age group. This stabilizes the estimates and reduces the required number of deaths and survey respondents for reliable estimations. However, such advanced statistical methodologies are generally not supported in standard software packages and often difficult to implement without making use of Bayesian statistics.

The same techniques that accommodate sparse data problems in the small-area life table estimations are also useful in the investigations into the determinants of difference in small-area population health. For example, the spatially structured components of the Bayesian life table methodology can be used to accommodate spatial dependency of the regression errors between geographically adjacent areas. Given the usually strong spatial autocorrelation in small-area regressions, this is an important consideration. Also, the same pooling of strength approach as in the LE and HLE estimations can be used to construct area and age specific risk factor profiles based on sparse survey data. Such profiles are essential for the calculation of the impact of behavioral risk factors (e.g. smoking, drinking, physical inactivity, etc.) on area-level outcomes. Accordingly, sparse data problems arise in many aspects of small-area estimations and the implemented solutions in this thesis provide a methodological consistency between the different chapters.

1.5 THE ADVANTAGE OF BAYESIAN STATISTICS IN SMALL-AREA ESTIMATIONS

There are two main approaches to statistics: frequentist and Bayesian. The frequentist approach is concerned with P(Data|Hypothesis), i.e. the probability of the data given a hypothesis. It is called frequentist because it deals with the frequency with which one would expect to observe the data, given some hypothesis about the world. Hence the frequentist approach treats the data as random (if the study was repeated, the data would be different) and the hypothesis as fixed (the hypothesis is either true or false, with a probability of either 0 or 1).

In contrast, the Bayesian statistical approach deals with P(Hypothesis|Data), i.e. the probability of the hypothesis given the data. This approach treats the data as fixed (these are the only data available to the researcher) and the hypothesis as random (the hypothesis might be true or false with some probability ranging from 0 to 1). This approach is called Bayesian because Bayes' theorem[38] is used to calculate the P(Hypothesis|Data), which involves updating the researcher's prior believes with the likelihood of the available data. The latter constitutes an updated probability, which is commonly referred to as the "posterior" probability.

In many situations, Bayesian and frequentist approaches arrive at similar conclusions. For example, with a sufficiently large dataset and flat priors (reflecting the absence of prior information), frequentist and maximum likelihood parameter estimates are mathematically (close to) identical.[39] However, Bayesian and frequentist approaches make fundamentally different assumptions about the notion of "probability". In frequentist statistics the chosen hypothesis is either true or false, which precludes claims about the probability that the (null) hypothesis is true. As a result, frequentist confidence intervals are somewhat counter-intuitive; a frequentist 95% confidence interval implies that when the experiment that generated the random sample would be repeated many times, 95% of the constructed intervals for those random samples would contain the true value of the parameter (and in 5% of the cases it would not). In contrast, the Bayesian 95% credible interval is more intuitive: it simply states the probability that the true value lies within the constructed interval, given the prior information and the available data.

Apart from their perspective on probabilities, there are other differences between the frequentist and Bayesian approach. First of all, making use of Markov Chain Monte Carlo (MCMC) estimation techniques, Bayesian statistics can be used to set up and estimate complicated models that are very difficult to implement using frequentist statistics. Secondly, Bayesian inference via MCMC has a theoretic guarantee that the MCMC algorithm will converge to the correct posterior, provided that the sampling is run long enough. In contrast, frequentist inference with Maximum Likelihood Estimation (MLE) has no guarantee of convergence and the latter can be problematic when dealing with complicated models. Thirdly, Bayesian statistical inference is unbiased with respect to the available sample size and can accommodate any sample size. In fact, given the prior and the data, the obtained results are exact. In contrast, frequentist statistical inference becomes increasingly biased as the available sample size decreases and is often wildly biased with small samples.[40] Fourthly, Bayesian statistics provides a natural way to include prior information in the analyses. This can improve Bayesian versus frequentist estimations, particularly when dealing with sparse data problems such as those encountered in small-area estimations.

There are also advantages of the frequentist approach. First and foremost, frequentist statistics does not depend on the validity of a chosen prior. In contrast, Bayesian statistics can produce misleading results with poorly chosen priors and the selection of sensible distributions and translation of prior believes into mathematically formulated priors requires considerable expertise.[41] In other words, the use of prior distributions in Bayesian statistics provides a powerful tool, but it also introduces a layer of subjectivity and complexity that is not encountered in frequentist statistics. Another advantage of frequentist models is the (often) considerably shorter run times compared to Bayesian models via MCMC. A few decades ago, this implied that Bayesian methods could realistically only be used for simple models. However, with the ongoing increase in computational power, Bayesian statistics has become feasible for a wide range of problems and, more specifically, has turned out to be particularly suitable for fitting the relatively complicated spatial random- and mixed-effects models as implemented in this thesis—both for the life table estimations aimed at measuring small-area population health as for the statistical analyses aimed at explaining the observed variation in small-area population health outcomes.

1.6 OVERVIEW OF THIS THESIS

The remainder of this thesis is organized as follows.

Chapters 2 and 3 validate an innovative (Bayesian random-effects) approach to estimating small-area LE and HLE. Given the inherent sparse data problems at the small-area level, there is no "golden" benchmark available by which the performance of different methodologies can be compared. Accordingly, chapters 2 & 3 are based on a novel benchmark: rather than creating fully synthetic simulation data, with artificially imposed correlations between the different dimensions of the data, the benchmarks are designed to mimic the exact male and female population age structures, age-specific mortality rates, and geographic locations of a large number of European benchmark countries. This benchmark is deliberately chosen to represent a worst-case scenario for the Bayesian random-effects approach and ensures a meaningful comparison between different small-area estimation methodologies.

Chapters 4 and 5 deal with the practical implementation of the selected indicators. Having established the validity and benefit of the Bayesian random-effects approach in chapters 2 and 3, chapter 4 investigates the impact of an important confounder, i.e. the migration of frail elderly to nursing homes. Elderly who move to a nursing home are known to have a substantially higher risk of mortality and, given the unequal spatial distribution of nursing homes at the small-area level, this migration can have a major impact on indicators of small-area population health. Chapter 4 investigates this impact and compares several possible corrections, either implemented as an integral part of the data collection or alternatively carried out as an integral part of the life table calculations. Both corrections remove the impact of nursing homes from the resulting summary measures of population health. Next, chapter 5 extends the Bayesian random-effects methodology to allow for the estimation of small-area health-adjusted life expectancy. The latter is an indicator of population health that provides balanced attention to fatal as well as non-fatal health outcomes and is particularly suited for policy purposes. Chapter 5 then highlights how HALE can be used to map, rank, and select areas for policy purposes while taking the full uncertainty of the estimates into account.

Chapters 6 and 7 subsequently use the selected indicators and Bayesian methodology to explain geographic differences in small-area population health. In chapter 6, associations between various indicators of urban green and small-area LE and HLE are investigated. On the one hand, this chapter provides an interesting example of the

use of the developed and validated indicators in small-area ecological analyses. On the other hand, chapter 6 also highlights the intrinsic difficulties of using an ecological outcome measure in a study that aims to recover causal determinants. As a result, chapter 7 pilots a novel approach that combines the benefits of an ecological analysis (in terms of data availability and statistical power) and an individual-level analysis (in terms of identification of the parameters and interpretation of the results). More specifically, by using a synthesis of ecological and individual-level outcome data, chapter 7 investigates the extent to which small-area variation in cardiovascular mortality in Dutch neighborhoods can be explained by several behavioral risk factors (i.e. smoking, drinking, overweight, and physical inactivity). As before, Bayesian techniques are used to accommodate the sparse data problems.

Chapter 8 concludes with a summary of important findings, strengths and weaknesses as well as a discussion of the scientific contributions and new insights obtained. Of course, relevant policy implications and recommendations for future research are also discussed.

2

Comparison of Bayesian random-effects and traditional life expectancy estimations in small-area applications

with F.J. van Lenthe, P. Congdon, B. Donkers, A. Burdorf, and J.P. Mackenbach.
American Journal of Epidemiology (2012)

ABSTRACT

There are several measures that summarize the mortality experience of a population. Of these measures, life expectancies are generally preferred based on their simpler interpretation and direct age standardization, which makes them directly comparable between different populations. However, traditional life expectancy estimations are highly inaccurate for smaller populations and consequently are seldom used in small-area applications. In this paper, the authors compare the relative performance of traditional life expectancy estimation with a Bayesian random-effects approach that uses correlations (i.e., borrows strength) between different age groups, geographic areas, and sexes to improve the small-area life expectancy estimations. In the presented Monte Carlo simulations, the Bayesian random-effects approach outperforms the traditional approach in terms of bias, root mean square error, and coverage of the 95% confidence intervals. Moreover, the Bayesian random-effects approach is found to be usable for populations as small as 2,000 person-years at risk, which is considerably smaller than the minimum of 5,000 person-years at risk recommended for the traditional approach. As such, the proposed Bayesian random-effects approach is well-suited for estimation of life expectancies in small areas.

2.1 INTRODUCTION

Public health researchers are frequently required to compare mortality in different geographic areas. The standardized mortality ratio (SMR), which measures the excess or deficit of mortality compared with a chosen standard population, is commonly used for this purpose. However, inter-SMR comparisons are valid only if the areas compared have identical population structures.[50–52] Accordingly, 2 or more areas can only be compared via their SMRs if they have (at least) very similar population age structures. This condition is often violated, particularly in small-area studies, where local differences are not averaged out as much as in larger geographic areas.

Life expectancies, in contrast, allow for comparisons between geographic areas to be made without having to assume a particular standard population. Even for areas with very diverse population structures, life expectancies are directly comparable, and, in contrast to SMRs, differences between areas are also more intuitive and straightforward to interpret.[52–54] As such, life expectancies are preferable over SMRs as the outcome measure in studies concerning geographic mortality differences.

One of the major problems with life expectancies, however, is that the standard life-table method cannot be used to calculate life expectancy estimates for populations smaller than approximately 5,000 person-years at risk; below this size, the bias in the estimates, as well as the size of the accompanying standard errors, becomes too large for meaningful analysis.[36,37,55] In virtually all small-area applications, a threshold of 5,000 person-years at risk implies that a substantial number of areas cannot be included in the analyses, even after pooling several years of data. Consequently, life expectancies are usually avoided in small-area analyses.

With a random-effects method, on the other hand, it should be possible to calculate accurate life expectancies with sufficiently small standard errors for significantly smaller populations at risk. This can be achieved using a modeling approach that recognizes correlations (i.e., borrows strength) between different age groups, geographic areas, and sexes to stabilize the life expectancy estimates. These models are conveniently fitted using a Bayesian estimation approach. A further advantage of a Bayesian estimation approach is that it facilitates the calculation of standard errors and confidence intervals without having to rely on asymptotic normality assumptions that generally do not hold in small-area applications. It thereby produces standard errors that remain accurate for much smaller populations.[56,57]

Thus far, however, there has been no clear evidence that a Bayesian random-effects approach indeed results in more accurate life expectancy estimates than the traditional method. Neither is there an indication of a minimum required population size for estimating small-area life expectancies using a Bayesian random-effects approach. Thus, our aims in this study were to compare the relative efficiency of both methods (in terms of accuracy of the estimates and precision of the standard errors and confidence intervals) and to provide evidence that a Bayesian random-effects approach can indeed be used to calculate reliable life expectancies for smaller populations than with the traditional approach.

2.2 METHODS

Benchmark data

In our analysis, we use Monte Carlo simulations to evaluate the performance of traditional and Bayesian life expectancy calculations for (hypothetical) populations of varying size. A total of 6 simulations are performed, for populations of 500, 1,000, 2,000, 5,000, 10,000, and 25,000 person-years at risk. Instead of creating fully synthetic benchmark life-table data, with artificially imposed correlations between the different age groups, geographic areas, and sexes, the simulations are designed to mimic the exact male and female population age structures, age-specific mortality rates, and geographic locations of 33 European benchmark countries. For these countries, complete life-table data for the 2005–2006 period are obtained from either the Human Mortality Database[58] or the Eurostat Populat database.[59] These data (based on millions of inhabitants) serve as the input for the simulations, implying that the simulations recreate Europe as if it were a city, with European countries as its neighborhoods.

The 33 European benchmark countries have very different life expectancies, ranging from close to 80 years in Scandinavia and the Mediterranean to approximately 60 years in Russia and Ukraine (see Figure 2.1). These differences are substantial and may be larger than typically encountered in small-area applications (e.g., Congdon[57] reports a range of 12 years between the highest and lowest life expectancies in wards in eastern England). On the other hand, these differences do allow for a meaningful comparison between the two methods without an inherent bias in favor of the Bayesian random-effects approach due to the inclusion of very similar areas in the simulations.

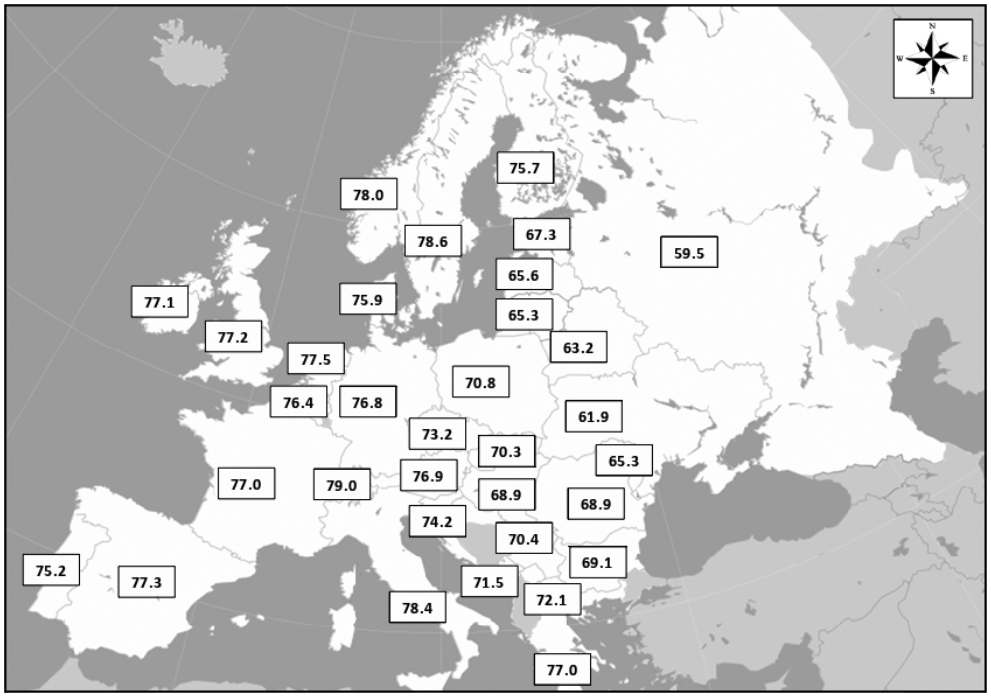
Creating the hypothetical small-area data sets

Each of the 6 simulations starts with the scaling down of the European benchmark populations to the required population size (i.e., 500, 1,000, 2,000, 5,000, 10,000, or 25,000 person-years at risk). This is programmed in MATLAB (The MathWorks, Inc.) and performed deterministically in order for the hypothetical small-area populations to mimic the exact age distribution of the benchmark countries. Subsequently, the numbers of deaths are simulated randomly—10,000 times per simulation—and independently for all age groups and both sexes using draws from a Poisson distribution with means set to the age- and sex-specific mortality rates of the benchmark countries. After the input data are generated, the data are automatically aggregated into abridged life tables that are ready for life expectancy estimations. Based on the recommendations of Toson and Baker[55] and Eayres and Williams,[37] standard 5-year abridged life tables with $\geq$85 years as the final age interval are used without any adjustment for age-specific death counts of zero within the life table. This results in 10,000 life-table data sets per simulation, each with a slight variation in the number of deaths but with average life expectancies exactly equal to those of the benchmark countries.

Life expectancy estimations

In the estimation phase, life expectancies and accompanying standard errors are calculated for all 10,000 life-table data sets per simulation using the traditional method (in MATLAB) and the Bayesian random-effects method (in OpenBUGS (www.openbugs.info)). Similar to Toson and Baker,[55] Eayres and Williams,[37] and Williams et al.,[60] the simulations focus on male life expectancies only (the difference in performance between the sexes is very small), and, following the recommendations in the same studies, all life expectancies are calculated using the Chiang life-table approach.[35] This holds for both the traditional method and the Bayesian method, but an issue that pertains only to the traditional life-table calculations is the occasional occurrence of zero deaths in the final age intervals. Whereas the Bayesian random-effects approach does not break down, the mortality rate of the final age interval will be exactly zero for the traditional approach, resulting in an infinite mean length of survival (1/mortality rate) and consequently an infinite life expectancy. To avoid these problems, zero death rates in the final age interval of the traditional life-table calculations are replaced by the average observed sex-specific mortality rate of the final age groups.

Figure 2.1 Male life expectancy (in years) in the 33 European benchmark countries, 2005–2006.



Another computational problem that only concerns the traditional method is the occasional occurrence of age-specific mortality rates that are equal to or higher than 0.40. In larger populations, such mortality rates generally do not occur, but in smaller populations it occasionally happens that a (nonfinal) age group has, for example, a population of 5 and a number of deaths of 2. Here the Chiang formula for the conditional probability that persons who enter the age interval will survive the age interval (P_x) is zero, implying that the life-table calculations no longer include subsequent age groups and that the standard life expectancy formulas break down. In these cases, the life expectancies of subsequent age groups are set to zero to exclude them from the life-table calculations.

Bayesian modeling approach

For the Bayesian method, a relatively basic model and a more advanced model are evaluated in the simulations. Both models pool strength over sexes, age groups, and

geographic areas using a random-effects method that includes multivariate (random) spatial effects that account for spatial clustering in mortality rates and multivariate (random) age effects that capture the usually high correlation between mortality rates for successive ages. The second model is more flexible than the first by allowing for age × area interactions, which means that the second model relaxes the assumption of a uniform age gradient per sex in the mortality patterns of the simulated areas. The model specifications are described in detail in the Appendix, and both models are realistic examples of Bayesian life expectancy models. These models can also be fitted using a frequentist estimation approach. An advantage of the Bayesian estimation approach, however, is that it is easily extendible to more complicated models and is more convenient for obtaining reliable standard errors and confidence intervals, which are directly available from the posterior life expectancy distributions that are already estimated for the life expectancy calculations.

The Bayesian models are fitted in OpenBUGS using iterative Markov chain Monte Carlo (MCMC) sampling techniques. The estimations for model 1 and model 2 both start with 25,000 burn-in MCMC iterations to allow the chains to converge, followed by 100,000 MCMC iterations with a thinning interval of 10 to reliably approximate the posterior life expectancy distributions. Note that we use relatively good starting values and also use a very conservative burn-in period in order to avoid having to inspect convergence for all 120,000 Bayesian estimations. Instead, convergence for the first 10 iterations of each simulation is evaluated using the Gelman-Rubin criteria based on 2 parallel chains.[61] Convergence was always obtained within 15,000 MCMC iterations, and, for all other estimations, convergence was assumed after 25,000 MCMC iterations. The Bayesian life expectancy estimations are distributed on the Dutch Life Science Grid[62] to considerably reduce the required computation time for the large number of regressions required for the simulations.

Comparison between methods

After the life expectancy estimations are completed, the point estimates and calculated standard errors of the traditional method and the means, standard deviations, and 95% credibility intervals of the Bayesian posterior distributions are aggregated and summarized in MATLAB. Together, these estimates form distributions of life expectancies and standard error estimates from which reliable inferences can be made. The method that produces 1) the most accurate life expectancies and 2) the most accurate estimated standard errors and confidence intervals is preferred.

Regarding the first criterion, the accuracy of the life expectancy estimates is characterized by two properties: 1) the bias of the life expectancy estimates and 2) the standard error of the life expectancy estimates. These measures capture the systematic error and random sampling variability of the estimates, respectively. The overall accuracy of the life expectancy estimates is summarized by the root mean square error (RMSE), a measure that combines the bias and standard error of the estimates into a single composite measure of absolute fit. Smaller values indicate a closer fit to the benchmark life expectancies; accordingly, the method with the smallest RMSE is preferred.

Regarding the second criterion, the accuracy of the reported standard errors and confidence intervals is evaluated by comparing the average estimated standard error with the simulated random error and a tally of how often the benchmark life expectancies are located within the reported 95% confidence intervals. The latter is referred to as the coverage of the 95% confidence intervals. The first criterion gives an indication of whether the reported standard errors reflect the true random sampling variability of the estimates, while the second criterion also takes the systematic error of the estimates into account and provides a more direct indication of how reliable the reported 95% confidence intervals actually are: values below 95% indicate that the confidence intervals are too optimistic, and values greater than 95% indicate that they are too conservative.

Comparison in a real-life example

The methods are also compared in a real-life example. Similarly to the simulations, standard 5-year abridged life tables with $\geq$85 years as the final age group are used for the life expectancy calculations. Unlike the simulations, there is no benchmark that can be used to formally compare the quality of the life expectancy estimates. Instead, the real-life example is used to more tentatively substantiate the simulation results. Firstly, the example shows how both approaches handle real-life data and associated problems that often occur in real-life analyses. Secondly, the sizes of the reported standard errors of both approaches are compared, which should be in line with the sizes of the reported standard errors in the simulations. Thirdly, the relative size of the bias of both approaches is investigated by examining the ratios of the estimated life expectancies per neighborhood; these should concur with the relative bias as reported in the simulations.

The selected geographic area for the example is the city of Rotterdam, the second-largest city in the Netherlands, with a population of approximately half a million people. The required population and mortality data for Rotterdam are obtained from

Statistics Netherlands and cover the 2008–2009 period. Rotterdam has 89 neighborhoods, of which several have little or no population (e.g., because they contain a public hospital, a city park, an airport, a zoo, or various industrial and harbor areas). Because it is not permitted by Statistics Netherlands to export life tables for populations smaller than 1,000 person-years at risk, 25 neighborhoods are excluded from the analysis. Together the excluded neighborhoods comprise 1% of the total population of Rotterdam.

2.3 RESULTS

Accuracy of the life expectancy estimates

Table 2.1 and Table 2.2 show the bias, standard error, and RMSE of the traditional and Bayesian life expectancy calculations. As can be seen, the life expectancy estimates of the traditional method are increasingly (upwards) biased for smaller populations. In contrast, the bias of Bayesian model 1 is considerably lower and the bias of Bayesian model 2 is close to zero.

All methods have standard errors that decrease with larger population sizes. The standard errors of the Bayesian approach, however, are approximately 40% smaller than those of the traditional approach for population sizes of 5,000 or less. For larger populations, Bayesian model 2 still performs better than the traditional approach, but Bayesian model 1 has a less flexible specification that becomes increasingly restrictive when more data become available. As a result, Bayesian model 1 performs even worse than the traditional approach for population sizes of 25,000.

This is also reflected in the RMSE, which indicates that Bayesian model 1 has a lower RMSE than the traditional approach for population sizes smaller than 10,000 person-years at risk but a higher RMSE for population sizes of 25,000. Bayesian model 2, on the other hand, has the lowest RMSE regardless of population size and clearly performs best in terms of the accuracy of the life expectancy estimations. Compared with the traditional approach, it has an approximately 40% lower RMSE for populations below 5,000 and still a 17% lower RMSE for populations as large as 25,000.

Table 2.1 Mean Error (Bias) and Standard Error (SE) for Male Life Expectancies (in Years), by Methodology and Population Size*

Population size	Bias			Standard error		
	Traditional	Bayesian 1	Bayesian 2	Traditional	Bayesian 1	Bayesian 2
500	1.1	0.8	0.1	6.2	3.8	3.8
1,000	0.7	0.5	0.0	4.4	3.0	2.8
2,000	0.6	0.2	0.0	3.1	2.2	2.0
5,000	0.3	0.1	0.0	2.1	1.8	1.5
10,000	0.2	0.1	0.0	1.5	1.5	1.1
25,000	0.1	0.0	0.0	0.9	1.2	0.8

** Results are based on 10,000 simulation iterations per population size*

Accuracy of the estimated standard errors and 95% confidence intervals

Table 2.3 shows the mean estimated standard errors and the coverage of the 95% confidence intervals for the traditional and Bayesian approaches. The estimated standard errors are highly correlated with the simulated standard errors in Table 2.1, with a correlation coefficient of 0.98. The traditional method, however, increasingly underpredicts the true variability of the life expectancy estimates for smaller populations. Bayesian model 1, on the other hand, over-predicts the true variability for smaller populations and under-predicts the true variability for larger populations, whereas Bayesian model 2 performs best with modest (but increasing) deviations for population sizes of 2,000 or less.

When one examines the coverage of the 95% confidence intervals, the results in Table 2.3 indicate that the traditional approach performs increasingly worse for smaller populations, which reflects the bias in the estimates and the under-prediction of the standard errors for smaller populations. In contrast, Bayesian model 1 performs well for smaller populations (where the small amount of bias mitigates the effect of the standard errors that are too large) but not for larger populations, where the size of the true standard errors is increasingly underpredicted. The more flexible Bayesian model 2 again performs best: it has a coverage of 94% for populations as small as 2,000 (note that the traditional approach requires populations of 25,000 to reach this level of accuracy), has a coverage of 95% for populations of 5,000, and even becomes somewhat conservative for populations of 10,000 and 25,000.

Table 2.2 Root Mean Square Error (RMSE) for Male Life Expectancies (in Years), by Methodology and Population Size*

Population size	Root Mean Square Error		
	Traditional	Bayesian 1	Bayesian 2
500	6.3	3.9	3.8
1,000	4.5	3.0	2.8
2,000	3.1	2.3	2.0
5,000	2.2	1.8	1.5
10,000	1.5	1.5	1.1
25,000	0.9	1.2	0.8

** Results are based on 10,000 simulation iterations per population size*

Performance in a real-life example

Figure 2.2 and Figure 2.3 summarize the results for the real-life example. Firstly, the ratios of estimated life expectancies in the city of Rotterdam, as depicted in Figure 2.2, substantiate the simulation results by showing that the traditional approach is slightly upwards biased relative to the Bayesian random-effects approach; this is indicated by the negatively sloped log-linear regression curve. Secondly, the ratios of estimated standard errors as shown in Figure 2.3 substantiate the simulation results by showing that the Bayesian approach produces smaller standard errors than the traditional approach, particularly for smaller populations.

Figure 2.2 and Figure 2.3 also depict several significant deviations between the Bayesian and traditional approaches. These occur in a limited number of situations, which involve small populations with 1) zero populations at risk in the life table, 2) zero deaths in the life table, 3) zero deaths in the final age group, or 4) 1 or 2 deaths in the entire life table but for very young persons. Here the traditional approach has computational difficulties and generates extreme life expectancies with unrealistically small standard errors (or infinite life expectancies with infinite standard errors), whereas the Bayesian random-effects approach simply borrows more information and produces larger standard errors. Overall, the results of the Bayesian random-effects approach are more stable and seem more reliable for smaller populations, in concurrence with the results of the simulations.

Table 2.3 Mean Estimated Standard Error (SE) and Coverage of the 95% Confidence Intervals for Male Life Expectancies (in Years), by Methodology and Population Size

Population size	Estimated Standard Error			Coverage of 95% Confidence Interval		
	Traditional	Bayesian 1	Bayesian 2	Traditional	Bayesian 1	Bayesian 2
500	5.0	4.3	3.4	86	95	89
1,000	4.0	3.2	2.5	90	94	92
2,000	2.7	2.2	1.9	91	93	94
5,000	1.9	1.7	1.5	92	91	95
10,000	1.4	1.3	1.2	93	89	96
25,000	0.9	0.9	0.8	94	84	96

** Results are based on 10,000 simulation iterations per population size*
** Coverage denotes the percentage of benchmark life expectancies contained in the estimated 95% confidence intervals. Values below 95% indicate too optimistic, and values greater than 95% too conservative confidence intervals.*

Finally, when comparing the estimated standard errors in the simulations (Table 2.3) with those in the real-life example (Table 2.4), the standard errors of the Bayesian random-effects approach are 10%–25% smaller than in the simulations, whereas the standard errors of the traditional approach are very similar in size. This is not entirely unexpected, since a random-effects approach performs better when more and more similar areas are included in the estimations, and the range of estimated life expectancies with Bayesian model 2 is only 9.2 years (74.3–83.5) for males and 8.2 years (78.5–86.7) for females, which is significantly smaller than in the simulations.

Table 2.4 Mean estimated standard error for life expectancies (in years) in neighborhoods in Rotterdam, the Netherlands, by methodological approach, 2008–2009

Population size	Men			Women		
	Estimated Standard Error			Estimated Standard Error		
	Traditional	Bayesian 2	Nr.Obs	Traditional	Bayesian 2	Nr.Obs
1,000—1,999	4.3	1.9	5	3.5	1.7	5
2,000—4,999	2.2	1.1	11	2.4	1.2	11
5,000—9,999	1.5	1.0	25	1.6	1.1	25
10,000—24,999	1.2	0.8	23	1.2	0.9	21
$\geq$ 25,000	n/a	n/a	0	0.8	0.6	2

Figure 2.2 Ratios (traditional approach/Bayesian approach) of life expectancy (LE) among men (left) and women (right) for neighborhoods in Rotterdam, the Netherlands, 2008–2009. The solid line represents the fitted log-linear regression curve.

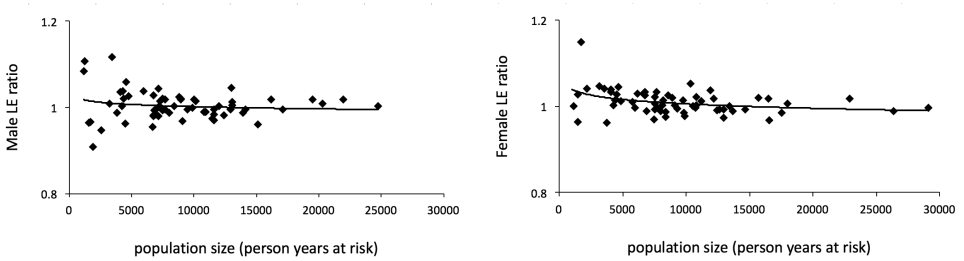
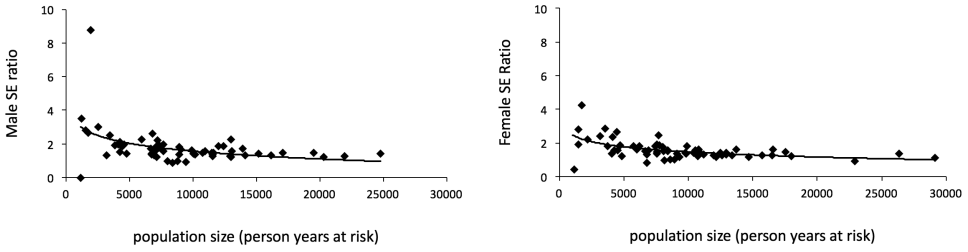


Figure 2.3 Ratios (traditional approach/Bayesian approach) of the estimated standard error (SE) of life expectancy among men (left) and women (right) for neighborhoods in Rotterdam, the Netherlands, 2008–2009. The solid line represents the fitted log-linear regression curve.



2.4 DISCUSSION

The presented simulations indicate that the Bayesian random-effects approach can improve significantly upon the traditional life expectancy calculations, in terms of both the RMSE and the accuracy of the standard errors and confidence intervals. In particular, a Bayesian random-effects model that allows for age × area interactions performs significantly better than the traditional approach at all simulated population sizes (i.e., 500-25,000 person-years at risk).

The simulation results also provide evidence that the often recommended minimum population size of 5,000 person-years at risk for the traditional approach is rather optimistic and that substantially larger populations are advisable to obtain sufficiently accurate standard errors and confidence intervals. As shown, the coverage of the confidence intervals was only 92% for populations of 5,000 and remained below 95% for populations as large as 25,000. Based on observed deviations from normally distributed standard errors, Scherbov and Ediev[36] arrived at similar conclusions and recommended caution for population sizes up to 50,000 person-years at risk. In contrast, a minimum required sample size of 2,000 person-years at risk for the Bayesian random-effects approach seems warranted. At this point, Bayesian model 2 has a lower RMSE and considerably more accurate standard errors than the traditional approach at its recommended minimum sample size of 5,000.

The simulations are even likely to provide a somewhat conservative estimate of the relative performance of the Bayesian random-effects approach vis-à-vis the traditional approach. The first reason is that the implemented simulation approach does not introduce random variation in the age distributions of the hypothetical small areas when scaling down the benchmark populations. This is similar to previous simulation studies and is the correct method for validating the accuracy of the reported standard errors; yet it also avoids computational problems with the traditional approach because it precludes zero populations at risk in the life tables. Whereas the Bayesian random-effects approach can adequately handle zero populations at risk in the life tables, the traditional approach breaks down and requires further aggregation or imputation of deaths to avoid computational problems. In real-life applications, where zero populations at risk frequently occur in small populations, the traditional approach is thus expected to perform somewhat worse than the simulations suggest.

The second reason for a conservative estimate of the relative performance of the Bayesian random-effects approach is that the simulations are based on benchmark

countries with substantial differences in population structures and life expectancies. These differences are deliberately used to represent something of a "worst-case scenario" for the Bayesian random-effects approach, which derives its strength from similarities between the areas under analysis. Many real-life small-area life expectancy estimations will involve smaller differences between the areas and also include substantially more areas; in these cases, the Bayesian random-effects method performs better than as suggested in the simulations (whereas the performance of the traditional approach remains the same). This point is also illustrated in the real-life example, where the Bayesian standard errors are 10%–25% smaller than as reported in the simulations.

Additionally, in a nonreported simulation (available from the first author upon request) where all geographic autocorrelation is removed by randomly shuffling the location of the benchmark countries in each individual regression while leaving the geographic structure of Europe intact, the results of the Bayesian random-effects approach remain unbiased and the RMSE remains 15%–25% lower than that in the traditional approach. Admittedly, such a scenario is quite unrealistic, but it does substantiate the robustness of the simulation results. Taking these considerations into account, a minimum sample size of 2,000 person-years at risk for the Bayesian random-effects approach seems prudent and is unlikely to be an artifact of the specific simulation approach.

Moreover, several improvements to the modeling approach can be envisaged that can further increase the performance of the Bayesian approach (in terms of bias and RMSE) and simultaneously have interesting public health applications. Firstly, due to computational constraints, the simulations in this paper do not incorporate more advanced spatial specifications and/or selection of random-effects models.[57,63] In real-life applications, where each model needs to be fitted only once, more flexible Bayesian models can be considered that are better suited to account for large differences in life expectancy between included areas. The estimation time for these models would increase from approximately 10–15 minutes for Bayesian model 2 to 20–30 minutes for more complicated models.

Secondly, it is also relatively straightforward to extend the Bayesian random-effects approach with a time dimension. By pooling strength over an additional dimension (in addition to sex, age group, and geographic area), the Bayesian results would further improve, and such a model could be especially useful for monitoring life expectancy in areas over time—for example, to detect atypical time trends or to evaluate the effect of interventions on the small-area level.

Thirdly, area-specific variables that are known to be correlated with geographic differences in life expectancy (e.g., the location of nursing homes and/or area deprivation scores) could be included to further improve the Bayesian life expectancy estimations. Such models are usually referred to as mixed models, and they can improve upon the results of purely random-effects models because the included predictors already explain part of the observed mortality differences between areas. Such an approach could, for example, be used to provide an elegant correction for the impact of nursing home deaths on small-area life expectancies.[64]

Finally, public health researchers and health authorities may also be interested in exploring the causes of variation in estimated life expectancies. For this purpose, the Bayesian random-effects approach can easily be extended to analyze differences in estimated life expectancy in followup regressions that directly take the precision of the life expectancy estimations as well as the spatial configuration of areas into account.

In conclusion, the Bayesian random-effects approach is versatile and well-suited for small-area life expectancy estimations. It performs better than the traditional approach for all simulated population sizes and allows for the estimation of accurate life expectancies and accompanying confidence intervals for populations as small as 2,000 person-years at risk.

Appendix A. Model 1

Let D_{six} and Pop_{six} denote deaths and populations at risk, classified by sex ($s = 1, 2$), area ($i = 1, \ldots, 33$), and age group ($x = 1, \ldots, 19$). Deaths are assumed to be Poisson-distributed:

$$\mathrm{D}_{six} \sim \ \mathrm{Poisson}(\mathrm{Pop}_{six} \times \mathrm{m}_{six}), \tag{2.1}$$

with m_{six} denoting mortality rates specified in the same dimensions. For larger populations, a binomial distribution could also be specified; however, given our focus on small populations with few observed deaths, the Poisson distribution is considered more appropriate. A standard log-link function is used, with the first model specified as

$$\log(m_{six}) = \alpha_s + \beta_{1\,sx} + \beta_{2\,si}. \tag{2.2}$$

This initial model contains:

1. Overall sex-specific mortality-level parameters α_s , which are assigned flat prior distributions from $-\infty$ to ∞ with the OpenBUGS "dflat()" distribution.

2. Parameters $\beta_{1\,sx}$ that represent the age-sex mortality rates for age group x and sex s; these are assigned a multivariate conditional first-order random walk prior that takes correlations between adjacent age groups and the correlation between the mortality experience of males and females into account.

3. Area effects $\beta_{2\,si}$ that represent spatially correlated mortality contrasts that are also allowed to be sex-differentiated; these are assigned a multivariate conditional autoregressive prior distribution that takes correlations between adjacent areas and the correlation between male and female mortality rates into account. Both β_1 and β_1 are estimated in OpenBUGS using the "MV.CAR" distribution with Wishart priors assigned to the precision matrices (both specified with 2 degrees of freedom and a 2-by-2 identity matrix as the inverse scale matrix).

Appendix B. Model 2

In model 1, the age effects $\beta_{1\,sx}$ are assumed to operate independently of the area effects $\beta_{2\,si}$. This assumption results in a parsimonious model, but the actual mortality pattern may not conform to the simplifying assumption of a (sex-differentiated) uniform age gradient in all areas. Accordingly, model 2 allows for age × area interactions:

$$\log(m_{six}) = \alpha_s + \beta_{1\,sx} + \beta_{2\,si} \times \beta_{3\,sx}. \tag{2.3}$$

The parameters α_s, $\beta_{1\,sx}$, and $\beta_{2\,si}$ are given exactly the same priors as in model 1, whereas the additional $\beta_{3\,sx}$ parameters are assigned gamma(1,1) priors. Together, the parameter combination $\beta_{2\,si} \times \beta_{3\,sx}$ provides a relatively parsimonious representation of age-sex-area mortality effects involving 104 parameters, thereby avoiding 1,254 (i.e. $2 \times 33 \times 19 = 1,254$) overall age-area interaction parameters $\beta_{4\,six}$.

In further extensions, this full set of parameters could still be introduced to correct for remaining discrepancies, but this would come at the cost of model parsimony if not combined with an automatic selection mechanism such as that described, for example, by Congdon.[57] Given the increase in computational time, these extensions are beyond the scope of the simulations presented in this paper. Finally, note that the $\beta_{1\,sx}$ and $\beta_{2\,si}$ parameters are constrained to sum to zero and that the $\beta_{3\,sx}$ are constrained to sum to 1 for identification of the parameters. The OpenBUGS code for models 1 and 2 is available upon request.

3

Estimating healthy life expectancy for small geographic areas

with B. Donkers, F.J. van Lenthe, F. Peters, A. Burdorf, and J.P. Mackenbach.

ABSTRACT

Healthy life expectancy (HLE) is a widely used summary measure of population health that is seldom used in small-area analyses because of sparse data problems. Standard approaches to alleviate sparse data problems comprise the aggregation or exclusion of smaller geographical areas, pooling of data over several years, use of abridged instead of complete life tables, and closure of life tables at younger ages. However, even when combined these approaches are insufficient to obtain reliable HLE estimates for small geographic areas. Accordingly, in this article we propose, validate, and demonstrate a Bayesian random-effects methodology that pools strength and recognizes correlations between the various dimensions of the sparse life table data (i.e. between sexes, age groups, and adjacent areas). Using Monte Carlo simulations, we show that the proposed random-effects methodology performs excellent with sparse life table data and does not perform worse than alternative methodologies with larger population and survey sample sizes. Additionally, we show that the proposed random-effects methodology provides accurate standard errors and 95% credibility intervals for even the smallest population and survey sample sizes considered in the simulations. This makes the Bayesian random-effects methodology ideally suited for the estimation of HLE in small-area applications.

3.1 INTRODUCTION

Healthy life expectancy (HLE), a concept introduced in the late 1960s and further developed in the 1970s and 1980s, provides an attractive addition to traditional life expectancies by partitioning total life expectancy into years spent in perfect health and years spent in less than perfect health. By being independent of the size of populations and of their age structures, HLE allows for direct comparisons between different populations and sub-groups within populations. It is an intuitive and meaningful summary measure of population health that has become a standard not only in demographic but also in public health research.[79]

Since its initial development, the use of HLE has grown substantially.[70] Health expectancies are used for the monitoring of differences between and changes in populations' health over time,[82] for the planning of health and social policies,[79] for assessments of the compression or expansion of population morbidity,[65] and for the targeting and evaluation of area-based initiatives aimed at improving health and reducing health inequalities.[1] For many of these applications it is informative or even necessary to calculate HLE at the sub-national level. Using traditional life table methodology, HLE can usually be calculated without any problems for areas with population sizes of at least 25,000 person years at risk. However, methodological problems make it increasingly difficult to calculate HLE for areas with smaller populations.

An important problem with the estimation of HLE for smaller geographic areas is that traditional life table methodology is based on analyses of each area independently, without any pooling of strength between sex and age groups. This methodology works well for large populations but not for smaller ones—particularly for populations smaller than approximately 5,000 person years at risk—because the bias in the estimates and size of the standard errors become too large for meaningful analysis due to sparse data problems.[37,55] Additionally, the estimated standard errors become increasingly less reliable for populations smaller than approximately 25,000 person years at risk.[36,72]

A second obstacle to the estimation of small-area HLE is the lack of adequate health status data for highly refined geographic areas. Because the required panel or event history data for multi-state life tables are usually unavailable, small-area HLE estimations are generally based on the less data-intensive Sullivan method.[80] Sullivan's method combines a period life table with age-specific disability prevalence rates obtained from cross-sectional survey data. These data are easier to obtain; however, at

the small-area level they are still difficult to acquire in sufficiently large sample sizes. As a result, small-area HLE estimations often suffer from insufficient statistical power to discriminate or differentiate between the areas under analysis.

Standard approaches to alleviate sparse data problems comprise the aggregation or exclusion of smaller geographical areas, pooling of data over several years, use of abridged instead of complete life tables, and closure of life tables at younger ages. Each helps to reduce sampling variation, but even combined they cannot fully accommodate the sparse data problems that are typically encountered in small-area analyses. To reliably estimate small-area HLE more advanced statistical methodologies are required that pool strength and recognize correlations between the various dimensions of the life table data (e.g. sexes, age groups, geographic areas, and health outcomes). This is performed using a random-effects methodology, which is essentially a regression without covariates and with a complicated structured error term. Sometimes area-level covariates are also included, which can further inform and stabilize the small-area estimates and results in a so-called mixed-effects instead of random-effects methodology.

For the estimation of small-area LE, several random and mixed effects specifications have been proposed and evaluated, e.g. random-effects models by Congdon[57] and Jonker et al.[72] and mixed-effects models by Jonker et al.,[73] Kulkarni et al.[75] and Wang et al.[81] For HLE, in contrast, only a few options are available. Besides the traditional HLE methodology, which is considered the reference methodology, several random-effects models have been proposed for the estimation of HLE in larger geographic areas,[67] for refined sub-groups within larger geographic areas,[76] and for the small-area estimation of health-adjusted life expectancy,[71] the latter being based on continuous (rather than discrete) health-state measurements. However, no small-area HLE estimation methodology has been developed or formally evaluated for smaller geographic areas.

In this paper, the authors introduce two different small-area HLE estimation approaches (i.e. a random and a mixed-effects methodology) and their statistical performance is formally compared with that of traditional life table methodology in a comprehensive simulation study. The simulation evidence provides guidance on the choice of appropriate methodology and an indication of the minimum required population and survey sample size for small-area HLE estimations. Furthermore, the proposed methodologies are applied in a real-life example, which further highlights the usefulness of the proposed methodologies in small-area HLE estimations and shows how well they perform relative to the traditional life table methodology.

3.2 METHODS

Monte Carlo simulations

Monte Carlo simulations are used to evaluate the performance of the traditional life table, random effects, and mixed effects methodologies with respect to a) the population size and b) the survey sample size of the life table data. Five different population sizes (i.e. 1,750, 2,000, 5,000, 10,000, and 25,000 person years at risk) and 6 different survey sample sizes (i.e. 7.5, 15, 45, 150, 375, and 1,125 respondents per sex) are evaluated. The lower values are anticipated to be problematic even for the random and mixed-effects methodologies whereas the highest values are considered large enough to suffice even for the traditional life table methodology.

Benchmark data

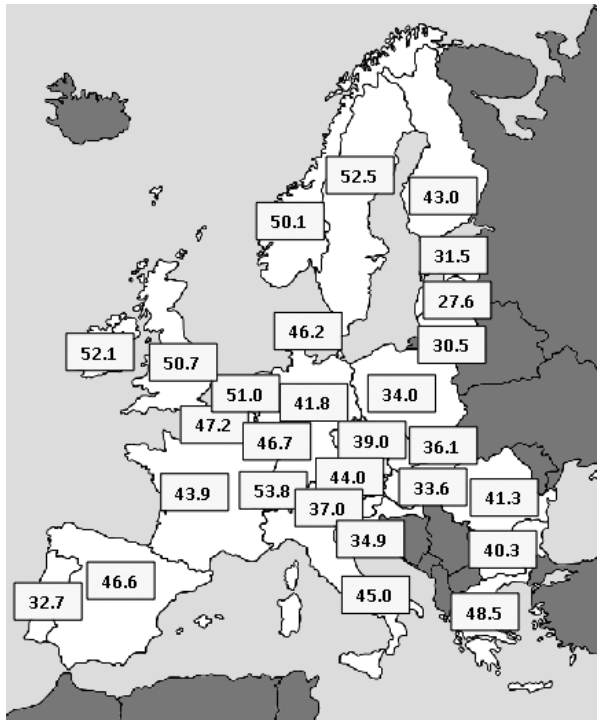
Instead of creating fully synthetic benchmark life-table data, with artificially imposed correlations between the different age groups, geographic areas, and sexes, the HLE simulations are designed—similar to Jonker et al.[72,73]—to mimic the exact male and female population age structures, age-specific mortality rates, age-specific health prevalence rates and geographic location of 28 European benchmark countries. At the small-area level, sufficiently reliable data to establish these benchmark inputs are inherently unavailable. Accordingly, the simulations recreate Europe as if it were a city, with European countries as its neighborhoods, based on reliable large-sample data.

For the 28 included benchmark countries, life table data and harmonized health surveys have been obtained from the Eurostat Populat database and EU-SILC surveys, respectively. The EU-SILC surveys contain information about respondents' self-assessed health ('How is your health in general?') on a 5-point Likert scale: 1) very good, 2) good, 3) fair, 4) bad, and 5) very bad. The benchmark age- and sex-specific health prevalences are calculated as the proportion of respondents that assess their health as either 'good' or 'very good'. Because the EU-SILC surveys only include people of 16 years and older, HLE at birth cannot be calculated without additional assumptions about or imputations for the prevalences of health in the youngest age groups. In the simulations, such assumptions are avoided by calculating HLE directly for age group 15-19.

Figure 3.1 shows the benchmark male HLE for age-group 15-19, which are based on life table data aggregated over the 2008-2010 period. As can be seen, the benchmark

countries have extremely different HLEs with benchmark values ranging from 27.6 years in Latvia to 53.8 years in Switzerland. This range is larger than with regular life expectancy and likely more extreme than actually observed at the small-area level (see e.g. Jonker et al.[71]). Accordingly, the European benchmark presents a 'worst-case scenario' for the Bayesian random-effects methodology, which pools strength from similarities between the areas under analysis. In contrast, the performance of the mixed-effects methodology is less affected by these differences (because the covariates are expected to already capture a significant proportion of the differences in mortality rates and morbidity prevalences), and the performance of the traditional methodology is even completely unaffected by similarities or differences between the benchmark countries because HLE is estimated independently for each area.

Figure 3.1 Benchmark male healthy life expectancies for age-group 15-19*



Based on 2008-2010 data (Source: Eurostat)

Creating the hypothetical small-area data sets

The simulation datasets are created in MATLAB (The MathWorks, Inc). First, all European benchmark populations are scaled-down to the required population size (i.e. 1,750, 2,000, 5,000, 10,000, or 25,0000 person-years at risk per gender). This is performed deterministically to ensure that the hypothetical small-area populations mimic the exact age-distributions of the benchmark countries. Subsequently, the number of deaths for all age groups, areas, and sexes are simulated randomly and independently—2,000 times for each population size—using draws from a Poisson distribution with means set to the age- and sex-specific mortality rates of the benchmark countries.

Next, the survey samples are created deterministically under the assumption that survey respondents are evenly distributed between sex and age-groups. For the survey sample sizes of 15, 45, 150, 375, and 1,125 respondents per sex this implies 1, 3, 10, 25, and 75 survey respondents for each of the 15 life table age groups, respectively. For the survey sample size of (on average) 7.5 respondents, only a single respondent per age group is available in each area. This respondent is specified as male in the odd age groups and as female in the even age groups in half of the areas, and vise versa in the other areas. Subsequently, the numbers of healthy respondents are simulated randomly and independently—2,000 times for each survey size—using draws from a binomial distribution with the number of draws set to the determined survey sample sizes and with the probability of success specified as the benchmark age and sex-specific proportions of healthy respondents.

After the input data are generated, the data are automatically aggregated into abridged life tables that are ready for the life expectancy estimations. Conform the recommendations made by Toson and Baker[55] and Eayres and Williams,[37] standard 5-year abridged life tables with $\geq$85 years as the final age interval are used without any adjustment for age-specific death counts of zero within the life table. This results in 60,000 datasets, each with slight variations but with average HLEs exactly equal to those of the benchmark countries.

Healthy life expectancy estimations

In the estimation phase, HLE and accompanying standard errors are calculated for all datasets using the Bayesian random effects and mixed effects methodologies (in OpenBUGS) and for all but the smallest survey sample size—due to missing values in the survey age groups—using the traditional methodology (in MATLAB). Based

on the recommendation of Toson and Baker[55] and Eayres and Williams,[37] the Chiang[66] life-table approach is used to derive the conditional probabilities of death (q_x), and Sullivan's method is used to calculate the HLE.[80] For the traditional approach, standard errors are calculated that take both the uncertainty in the mortality and prevalence rates into account.

Exactly the same life table approach is used for estimating HLE for the traditional, random effects and mixed effects methodologies There are two computational problems, however, that specifically pertain to the traditional HLE estimations. The first is the occasional occurrence of zero deaths in the final age intervals. These pose no problem for the random and mixed-effects methodologies, but the mortality rate of the final age interval will be exactly zero for the traditional methodology, which results in an infinite mean length of survival (1/mortality rate) and consequently in an infinite life expectancy. To avoid this problem, zero death rates in the final age interval of the traditional life table calculations are replaced by the average observed sex-specific mortality rate of the final age groups. The second computational problem is the occasional occurrence of age-specific mortality rates that are equal to or higher than 0.40. In larger populations, such mortality rates generally do not occur, but in smaller populations it occasionally happens that a (non-final) age group has, for example, a population of 5 and a number of deaths of 2. Here the Chiang formula for the conditional probability that persons who enter the age interval will survive the age interval (P_x) is zero, implying that the life-table calculations no longer include subsequent age groups and that the standard life expectancy formulas break down. Therefore, the age-specific mortality rates are truncated at 0.3999, which essentially removes the impact of subsequent age groups without causing numerical problems.

Random effects model specification

The proposed random-effects methodology has 3 components. The first component estimates the age-specific mortality rates, the second component estimates the age-specific proportions of people in good health, and the third summarizes these estimates into HLE using standard life table calculations. Starting with the first component, the age-specific mortality rates are calculated using the specification as proposed and validated by Jonker et al.[72,73] Accordingly, the observed mortality counts are assumed Poisson distributed:

$$\mathrm{D}_{six} \sim \text{Poisson}\ (\mathrm{Pop}_{six} \times \mathrm{m}_{six}) \tag{3.1}$$

with D_{six} and Pop_{six} denoting the number of deaths and the population at risk classified by sex ($s = 1, 2$), area ($i = 1, \ldots, N$), and age group ($x = 1, \ldots, X$),

respectively. The m_{six} denote the accompanying mortality rates specified in the same dimensions. A standard log-link function is combined with the following specification to predict the mortality rates:

$$\log(m_{six}) = \beta_{0\,s} + \beta_{1\,sx} + \beta_{2\,si} \times \beta_{3\,si} \tag{3.2a}$$

Here the β_0-parameters are fixed intercepts that capture level differences in the mortality rates for males and females, the β_1-parameters multivariate, gender-specific random age-effects that capture the generally high correlation between the mortality rates of successive age groups and genders, and the β_2-parameters multivariate, gender-specific random spatial-effects that account for the potential spatial clustering of mortality rates in adjacent areas while simultaneously taking the correlation between genders into account. The spatial effects are interacted with a set of β_3-parameters, which allows the spatial effects to be more pronounced for specific age-groups. The latter relaxes the assumption of an uniform (gender-specific) age-gradient for all areas while still retaining a relatively parsimonious model specification. For identification of the parameters, the β_1 and β_2 parameters are subject to sum-to-zero constraints and the β_3 parameters to a mean-of-one constraint.

The second component of the Bayesian random-effects methodology estimates the age-specific proportions of people in good health. Reflecting the structure of the discrete input data (i.e. survey respondents are either 'healthy' or 'unhealthy')[1], the observed number of healthy respondents is assumed binomial distributed:

$$HR_{six} \sim \text{binomial}(p_{six}, TR_{six}) \tag{3.3}$$

with HR_{six} and TR_{six} denoting the number of healthy and total survey respondents classified by sex ($s = 1, 2$), area ($i = 1, \ldots, N$), and age group ($x = 1, \ldots, X$), respectively, and p_{six} denoting the accompanying probabilities of being healthy specified in the same dimensions. Next, a standard logit-link function is combined with the following specification to predict the probabilities of being healthy:

$$\text{logit}(p_{six}) = \beta_{4\,s} + \beta_{5\,sx} + (\beta_{5\,sx})^2 \times \beta_{6\,si} + \beta_{7\;si} \tag{3.4a}$$

Here the β_4-parameters are fixed intercepts, the β_5-parameters multivariate, gender-specific random age-effects, and the β_7-parameters multivariate, gender-specific ran-

[1] Note that this is HLE-specific: with continuous health-state measurements an alternative small-area modelling strategy is required (see Jonker et al.[71]).

dom spatial-effects. The squared β_5-parameters are multiplied with sex- and area-specific β_6-parameters and for identification of the parameters, the β_5 and β_7 parameters are subject to sum-to-zero constraints.

Overall, specification 3.4a is sufficiently flexible to accommodate a wide range of morbidity patterns with only a limited number of parameters, while still being able to pool strength from similarities between the included areas. This makes specification 3.4a particularly suitable for the estimation of morbidity prevalences from (very) sparse survey data, without compromising its performance in case of larger survey sample sizes.

Mixed effects model specification

The mixed-effects methodology is based on the small-area LE methodology of Kulkarni et al.[75] and Wang et al.[81] combined with the small-area health estimation framework as proposed by Srebotnjak et al.[78] For simplicity, we refer to this model as the Murray et al. methodology – named after the author that is present on all papers. The Murray et al. methodology uses the same model components as the previously-described random-effects methodology but it relies heavily on the included covariates and uses a simpler random-effects structure. Accordingly, the proposed mixed effects model is a separate model and not simply the equivalent of the proposed random-effects model with additional covariates.

Conform Kulkarni et al.[75] and Wang et al.,[81] equation 3.1 in the mortality-component of the model remains identical, but the mortality rates are now predicted as:

$$\log(\mathrm{m}_{six}) = \beta_{8\,sx} + \beta_{9\,sx} \times \mathrm{income}_i + \beta_{10\,sx} \times \mathrm{educ}_i + \beta_{11\,sx} \times \mathrm{poverty}_i + \beta_{12\,six} \quad (3.2\mathrm{b})$$

Here the β_8-parameters are sex and age group specific (fixed) intercepts, the $\beta_9 - \beta_{11}$ sex and age group specific (fixed) coefficients, and the β_{12} (random) spatial effects. Based on data availability and on Kulkarni et al.,[75] the covariates included in equation 3.2b are GDP per capita, the percentage of upper secondary education attainment, and a specific social welfare indicator, respectively. Table 3.1 contains a short description and summary statistics of the included covariates, which are included to explain the bulk of observed differences in the age-specific mortality rates, with any remaining discrepancy to be captured by the random spatial (β_{12}) effects, which are sex and age-group specific and pool strength over dimension i.

Table 3.1 Description and summary statistics of the included covariates *

Covariate	Description	Mean	Std.Dev.	Min	Max
income	Gross domestic product per capita in purchasing power parities	103.1	46.5	44.5	265.0
education	Percentage of the population aged 18-64 having completed at least upper secondary education	76.2	13.6	30.0	91.4
poverty	Percentage of the population at risk of poverty or social exclusion (i.e. AROPE indicator)	23.9	8.37	14.8	47.3

** Based on 2008-2010 data (Source: Eurostat)*

Equation 3.3 in the morbidity-component of the model also remains identical to the previously described random-effects methodology (cf. Srebrotnjak et al.[78]). However, the morbidity prevalences are now predicted as:

$$\text{logit}(\text{p}_{six}) = \beta_{13\,sx} \times \beta_{14\,sx} \times \text{income}_i + \beta_{15\,sx} \times \text{educ}_i + \beta_{16\,sx} \times \text{poverty}_i + \beta_{17\,six} \quad (3.4b)$$

This equation mimics equation 3.2b but with a logit-link function and with a different set of parameters. The third and final component of the mixed-effects methodology is also similar to the one in the random effects approach; this component summarizes the estimated mortality rates and proportion of healthy respondents into estimated HLE.

Bayesian estimation and priors

Both the random and mixed-effects models are fitted using Bayesian methods. Bayesian parameter estimation involves selecting prior densities for the unknown model parameters and updating those densities via the likelihood of the data. The exact specification of the prior densities of the Bayesian random-effects and mixed-effects methodologies is included in Appendix A, which contains the full model codes.

All Bayesian models are fitted in OpenBUGS using Markov chain Monte Carlo (MCMC) sampling techniques. Each estimation starts with 25,000 burn-in MCMC iterations to allow the Markov chain to converge. Relatively good starting points and a very conservative burn-in period are used to ensure convergence, followed by 50,000 MCMC iterations to reliably approximate the posterior distributions.

Comparison between methodologies

After the healthy life expectancy estimations are completed, the point estimates and calculated standard errors of the traditional methodology and the medians, standard deviations and 95% credibility intervals of the Bayesian posterior distributions are aggregated and summarized in MATLAB. Together, these estimates form distributions of healthy life expectancies and standard error estimates from which reliable inferences can be made. Similar to Toson and Baker,[55] Eayres and Williams,[37] Williams et al.,[60] and Jonker at al.,[72,73] the simulations focus on male HLE—the difference in performance between the sexes is small—and the methodology that produces a) the most accurate healthy life expectancies, and b) the most accurate estimated standard errors and 95% confidence intervals is preferred.

Regarding the first criterion, the accuracy of the healthy life expectancy estimates is characterized by two properties: a) the bias of the life expectancy estimates and b) the standard error of the life expectancy estimates. These measures reflect the systematic error and random-sampling variability of the estimates, respectively. The overall accuracy of the life expectancy estimates is also summarized by the root mean square error (RMSE), which combines the bias and standard error of the estimates into a single composite measure of absolute fit. Smaller values indicate a closer fit to the benchmark life expectancies; accordingly, the methodology with the smallest RMSE is preferred.

Regarding the second criterion, the accuracy of the reported standard errors and confidence intervals is evaluated by a comparison of the average estimated standard error to the simulated random error and the coverage of the 95% confidence intervals. The latter is essentially a tally of how often the benchmark healthy life expectancies are located within the reported 95% confidence intervals. The first criterion gives an indication whether the reported standard errors reflect the true random-sampling variability of the estimates, while the second criterion also takes the systematic error of the estimates into account and provides a more direct indication of how reliable the reported 95% confidence intervals actually are: values below 95% indicate that the confidence intervals are too optimistic whereas values greater than 95% indicate that they are too conservative.

Real-life example

The methodologies are also compared in a real-life example. Unlike in the simulations, there is no benchmark available that allows for a formal comparison of the quality of

the estimates in terms of bias and RMSE. Therefore the real-life example is used for 2 purposes: a) to highlight the usefulness of the Bayesian random-effects methodology in small-area applications relative to the traditional life table and Murray et al. mixed-effects methodology, and b) to investigate the size of the standard errors under real-life conditions, which are likely to deviate from those in the simulations due to larger similarities between the areas under analysis.

The presented example is based on small-area data from all 22 Dutch metropolitan agglomerations aggregated over the 2006-2009 period. The included agglomerations comprise roughly 40% of the Dutch population and cover all major urban regions in the Netherlands (see Figure 3.2). Within each metropolitan agglomerations, life table data are collected at the neighborhood level. Neighborhoods are the smallest geographical level at which routinely collected data from Statistics Netherlands are available and the smallest geographical level at which microdata within the microdata-infrastructure of Statistics Netherlands can be aggregated. Neighborhoods are also a convenient unit for small-area analyses: their boundaries adhere to physical obstacles that break up urban landscapes (e.g. important traffic arteries, bodies of water, green spaces, etc.) and they are specifically designed to comprise relatively homogeneous areas in terms of type of housing and inhabitants.

In total, 1,157 neighborhoods within the 22 Dutch metropolitan agglomerations meet the required minimum population size of 1,750 person years at risk and minimum required survey sample size of 8 respondents. For each of these neighborhoods, population and mortality data are obtained from the Dutch population (GBA) and deaths registry (DO). Both databases are maintained by Statistics Netherlands and provide complete and continuous coverage of all Dutch inhabitants. Midyear population counts are extracted from the GBA at the first of July of 2006, 2007, 2008, and 2009 and the sum of these counts is taken as the population at risk for the 2006-2009 period. For the life table mortality data, all deaths in the same period are aggregated. Because real-life small-area data are confounded by the location of nursing homes, the life table data are corrected for the location of nursing homes using previous residential address information (cf. Jonker et al.[71]). This procedure reduces the variation in HLE between areas but results in more relevant and also more reliable HLE estimates.

For each of the 1,157 neighborhoods, the required health status data are obtained by combining responses to identical self-rated health questions from the WOON (2006 and 2009), POLS (2006-2009) and EU-SILC (2006-2009) surveys, all of which are

Figure 3.2 Geographic location of the 22 Dutch metropolitan agglomerations

administered by Statistics Netherlands. All surveys are nationally representative and contain the exact same self-rated health question as in the simulations. The respondents' health ratings are linked to the GBA using unique person identifiers. This allows for a correction for respondents that could theoretically be included more than once in the surveys. Similar as in the simulations, survey respondents are considered healthy when they rated their health as being "good" or "very good" and considered unhealthy otherwise. The required covariates for the mixed-effects methodology are also obtained from Statistics Netherlands. Based on data availability, the included covariates are not identical to those in the simulations, but they cover the same 3 dimensions (i.e. income, education, and social welfare). The included covariates are the average disposable household income, the percentage of inhabitants with a higher (tertiary) education, and the percentage of households with social welfare as its main source of income within each neighborhood.

Finally, HLE at age 15-19 is calculated for the neighborhoods within each metropolitan agglomerations using the exact same methodologies as in the simulations. The results of the real-life example estimations are summarized using standard descriptive statistics.

3.3 RESULTS

Accuracy of the life expectancy estimates

Table 3.2 contains the bias, standard error, and RMSE of the traditional life table and Bayesian random effects and mixed effects methodologies. Starting with the bias, all methodologies have relatively small bias insofar the populations are at least 5,000 person years at risk and the survey sample sizes at least 15 respondents. For smaller population and survey sample sizes, the Bayesian random-effects methodology performs best.

Turning to the standard errors, this is where the Bayesian random-effects methodology consistently outperforms the other methodologies. In comparison with the Murrey at al. mixed-effects methodology, the Bayesian random-effects methodology has almost consistently more than 1 year smaller standard errors. In comparison with the traditional life table methodology, the Bayesian random-effects methodology has 2.5 year smaller standard errors for the smallest survey sample sizes yet almost identical standard errors for the largest population and survey sample size. Accordingly, the traditional methodology performs well for larger population and survey sample sizes, but performs increasingly worse with sparse life table data.

The same conclusions can be drawn in terms of RMSE. The Bayesian random-effects methodology has the lowest RMSE in all simulations, the mixed-effects methodology performs worse regardless of population and survey sample size, and, compared to the traditional life table methodology, the difference increases from less than 1% for the combination of 25,000 person years at risk and 1,125 survey respondents to approximately 40% for the smallest population and/or survey sample sizes. Another way to summarize these results is that the Bayesian random-effects methodology performs excellent with sparse life table data without performing worse than the traditional life table methodology for even the largest included population and survey sample sizes.

Accuracy of the estimated SE and 95% CI

Table 3.3 contains the mean estimated standard error and the coverage of the 95% CI of the traditional and Bayesian methodology. As can be seen, the estimated standard errors of the traditional approach are increasingly underpredicted for smaller population and survey sample sizes and turn out to be very unreliable for the smallest survey sample sizes. The standard errors of the Murrey et al. mixed-effects methodology

Table 3.2 Mean Error (Bias), Standard Error (SE) and Root Mean Square Error (RMSE) for Male Healthy Life Expectancy Estimates (in Years), age-group 15-19, by Methodology, Population, and Survey Sample Size*

		Traditional life table methodology						Murray et al. mixed-effects methodology						Jonker et al. random-effects methodology					
		survey sample size (# males per area)						survey sample size (# males per area)						survey sample size (# males per area)					
measure	population size (#males per area)	7.5**	15	45	150	375	1125	7.5	15	45	150	375	1125	7.5	15	45	150	375	1125
Bias	1,750	n/a	0.32	0.29	0.28	0.28	0.28	-0.17	0.23	0.11	0.11	0.14	0.17	-0.17	-0.17	-0.21	-0.19	-0.12	-0.04
	2,000	n/a	0.29	0.27	0.26	0.26	0.25	-0.24	0.17	0.04	0.03	0.05	0.08	-0.16	-0.19	-0.21	-0.20	-0.12	-0.05
	5,000	n/a	0.16	0.13	0.12	0.12	0.12	-0.33	0.09	-0.04	-0.06	-0.05	-0.04	-0.13	-0.14	-0.18	-0.16	-0.11	-0.04
	10,000	n/a	0.10	0.07	0.06	0.06	0.05	-0.36	0.06	-0.07	-0.09	-0.08	-0.07	-0.12	-0.16	-0.16	-0.15	-0.09	-0.02
	25,000	n/a	0.06	0.03	0.02	0.02	0.02	-0.38	0.04	-0.09	-0.11	-0.09	-0.08	-0.11	-0.14	-0.15	-0.15	-0.09	-0.03
SE	1,750	n/a	6.67	4.09	2.64	2.10	1.82	6.16	5.12	4.66	4.48	3.31	2.28	4.94	3.99	2.81	1.95	1.50	1.17
	2,000	n/a	6.62	4.02	2.54	1.98	1.67	6.06	5.06	4.68	4.34	3.09	1.99	4.91	3.93	2.76	1.90	1.46	1.10
	5,000	n/a	6.44	3.81	2.21	1.57	1.18	5.92	4.96	4.59	4.27	3.00	1.79	4.83	3.85	2.69	1.82	1.35	0.96
	10,000	n/a	6.37	3.72	2.10	1.41	0.96	5.89	4.93	4.58	4.28	3.00	1.77	4.80	3.80	2.64	1.78	1.31	0.87
	25,000	n/a	6.34	3.69	2.04	1.33	0.83	5.82	4.83	4.48	4.19	2.92	1.69	4.79	3.78	2.62	1.75	1.28	0.83
RMSE	1,750	n/a	6.68	4.10	2.65	2.12	1.84	6.16	5.12	4.66	4.48	3.31	2.29	4.94	3.99	2.82	1.96	1.51	1.18
	2,000	n/a	6.62	4.03	2.55	1.99	1.69	6.06	5.06	4.68	4.34	3.09	1.99	4.91	3.94	2.77	1.91	1.46	1.10
	5,000	n/a	6.44	3.81	2.22	1.58	1.18	5.93	4.96	4.59	4.27	3.00	1.79	4.83	3.85	2.70	1.82	1.36	0.96
	10,000	n/a	6.37	3.73	2.10	1.41	0.96	5.90	4.93	4.58	4.28	3.00	1.78	4.80	3.81	2.65	1.79	1.31	0.87
	25,000	n/a	6.34	3.69	2.04	1.33	0.83	5.83	4.83	4.48	4.19	2.93	1.69	4.79	3.78	2.62	1.76	1.28	0.83

** Based on 2,000 simulation iterations per cell*
*** Unavailable because of missing values in the life tables*

Table 3.3 Mean Estimated Standard Error (Mean SE) and Coverage of the 95% CI for Male Healthy Life Expectancy (in Years), age-group 15-19, by Methodology, Population, and Survey Sample Size*

		Traditional life table methodology						Murray et al. mixed-effects methodology						Jonker et al. random-effects methodology					
		survey sample size (# males per area)						survey sample size (# males per area)						survey sample size (# males per area)					
measure	population size (#males per area)	7.5**	15	45	150	375	1125	7.5	15	45	150	375	1125	7.5	15	45	150	375	1125
Mean SE	1,750	n/a	1.42	3.39	2.43	1.92	1.59	3.00	2.46	1.75	1.32	1.23	1.06	5.13	4.17	2.97	2.07	1.61	1.24
	2,000	n/a	1.27	3.31	2.34	1.80	1.45	2.91	2.37	1.65	1.20	1.11	0.93	5.09	4.14	2.92	2.01	1.54	1.16
	5,000	n/a	0.82	3.11	2.09	1.50	1.09	2.81	2.26	1.52	1.03	0.93	0.74	5.03	4.08	2.85	1.91	1.33	0.98
	10,000	n/a	0.59	3.03	2.00	1.38	0.93	2.78	2.23	1.49	0.99	0.88	0.67	5.01	4.06	2.81	1.86	1.34	0.89
	25,000	n/a	0.37	2.99	1.94	1.30	0.82	2.77	2.22	1.47	0.97	0.85	0.64	4.99	4.03	2.79	1.83	1.30	0.82
Coverage	1,750	n/a	0.27	0.85	0.89	0.89	0.87	0.65	0.61	0.50	0.40	0.47	0.62	0.96	0.96	0.96	0.96	0.96	0.96
	2,000	n/a	0.25	0.85	0.89	0.89	0.88	0.64	0.61	0.49	0.38	0.44	0.60	0.96	0.96	0.96	0.96	0.96	0.96
	5,000	n/a	0.15	0.84	0.89	0.89	0.89	0.64	0.60	0.46	0.35	0.39	0.53	0.96	0.96	0.96	0.96	0.95	0.95
	10,000	n/a	0.10	0.84	0.89	0.90	0.90	0.63	0.60	0.46	0.34	0.38	0.51	0.96	0.96	0.96	0.96	0.95	0.95
	25,000	n/a	0.05	0.84	0.89	0.90	0.89	0.64	0.61	0.46	0.34	0.39	0.52	0.96	0.96	0.96	0.96	0.95	0.95

** Based on 2,000 simulation iterations per cell*
*** Unavailable because of missing values in the life tables*

are consistently underpredicted and thereby unreliable regardless of population and survey sample size. In contrast, the estimated standard errors of the Bayesian random-effects approach are accurate for all included population and survey sample sizes.

The same conclusion holds for the coverage of the 95% CI. The coverage of the 95% CI of the traditional life table approach never exceeds 0.90 and is extremely small for the smallest survey sample sizes. The coverage of the mixed-effects methodology does not even exceed 0.70, and the coverage of the Bayesian random-effects methodology is approximately 95% for all population and survey sample sizes. Accordingly, only the proposed Bayesian random-effects approach accurately predicts the true uncertainty of the HLE estimates.

Real-life example

Table 3.4 contains the results of the real-life example. As can be seen, neighborhoods with survey sample sizes of 375 or more respondents are not present in the dataset. Larger neighborhoods do have larger survey sample sizes because the health surveys are not stratified by neighborhood; accordingly, robust results are not available for all cells. When comparing the methodologies, the first major difference is the number of neighborhoods for which HLE can be estimated. Whereas the traditional methodology cannot handle missing values in the survey data (at least not without additional imputations), the Bayesian random-effects methodology has no problem with missing life table values and simply incorporates the additional uncertainty in the HLE estimates. Accordingly, HLE can be estimated for all 1,157 neighborhoods using the Bayesian random-effects methodology whereas only 68 neighborhoods are included for the traditional life table methodology. The performance of the mixed-effects methodology is somewhere in between. Because the mixed-effects approach does not pool strength between age groups, it also has trouble predicting HLE in several cases where the majority of neighborhoods within the metropolitan agglomerations has missing life table values. This particularly occurs in the oldest age groups and results in 788 neighborhoods being included in the example.

The second difference is the size of the standard errors. Whereas the standard errors of the traditional methodology are in line with the results of the simulations, the standard errors for the Bayesian random-effects methodology are smaller than predicted. In fact, the observed differences are 10%-30% smaller than in the simulations. Based on the simulation results, which showed that the predicted standard errors of the Bayesian random-effects methodology accurately reflect the true uncertainty of the HLE estimates, the latter implies a substantial performance improvement under

Table 3.4 Number of Included Neighborhoods (N) and Mean Estimated Standard Error (MSE) of male Healthy Life Expectancy (in Years) of the included neighborhoods within the 22 Dutch CMAs, age-group 15-19, by Methodology, Population, and Survey Sample Size

		Traditional life table methodology *					Murray et al. mixed-effects methodology **					Jonker et al. random-effects methodology ***				
		survey sample size (# males per area)					survey sample size (# males per area)					survey sample size (# males per area)				
measure	population size (#males per area)	8-14	15-44	45-149	150-374	≥375	8-14	15-44	45-149	150-374	≥375	8-14	15-44	45-149	150-374	≥375
N	1,750 - 1,999	0	0	0	0	0	8	3	0	0	0	9	3	0	0	0
	2,000 - 4,999	0	1	3	0	0	131	103	14	0	0	210	143	16	0	0
	5,000 - 9,999	0	7	21	2	0	76	186	47	0	0	116	318	60	2	0
	10,000-25,000	0	0	22	12	0	2	119	98	9	0	2	148	115	15	0
Mean SE	1,750 - 1,999	n/a	n/a	n/a	n/a	n/a	1.20	1.07	n/a	n/a	n/a	3.20	2.93	n/a	n/a	n/a
	2,000 - 4,999	n/a	2.86	2.60	n/a	n/a	1.63	1.19	0.66	n/a	n/a	3.86	3.05	2.13	n/a	n/a
	5,000 - 9,999	n/a	2.46	2.31	1.67	n/a	1.53	1.15	0.79	n/a	n/a	3.68	3.10	2.20	1.90	n/a
	10,000-25,000	n/a	n/a	2.46	1.86	n/a	1.28	0.96	0.87	0.61	n/a	3.29	2.75	2.30	1.61	n/a

** = 94% of neighborhoods excluded, ** = 31% of neighborhoods excluded, and *** = 0% of neighborhoods excluded due to sparse data problemss*

real-life conditions. Finally, the mixed-effects methodology reports even smaller standard errors than in the simulations. Given the very sparse life table data and absence of pooling of strength between sexes and age groups in the Murray et al. specification, which also resulted in a substantial number of missing values, this likely implies a further underprediction of the true uncertainty of the HLE estimates under real-life conditions.

3.4 DISCUSSION

The presented Monte Carlo simulations confirm that the proposed Bayesian random-effects methodology performs very well with sparse life table data, much better than the Murray et al. mixed-effects and traditional life table methodology. Equally important, the Bayesian random-effects methodology does not perform worse than the traditional life table and Murray et al. mixed-effects methodology for larger population and larger survey sample sizes. The simulations also show that the Bayesian random-effects approach has superior performance when it comes to accurately predicting the true uncertainty of the HLE estimates. Based on these results, the proposed Bayesian random-effects approach is recommended for small-area HLE estimations, particularly for the population and survey sample sizes as included in the simulations but without any indication that larger areas cannot be included as well.

Regardless of the performance of the random-effects methodology in the simulations, the presented simulation evidence likely provides a somewhat conservative estimate of the true performance of the Bayesian random-effects methodology relative to the mixed and traditional life table methodologies. The first reason is the rather extreme difference in HLEs of the European benchmark countries. As mentioned, the large differences between European countries are deliberately used to represent something of a "worst-case scenario" for the Bayesian random-effects methodology, which derives an important part of its strength from similarities between the areas under analysis. At the small-area level, cultural differences in the perception of good health and/or institutional differences in terms of percentage of nursing home inhabitants, which affect the prevalence of good health in the general population, play a smaller role.[69] Accordingly, the results of the real-life example, with average standard errors of the Bayesian approach being 10-30% smaller than in the simulations, provides a more realistic indication of the true performance of the Bayesian random-effects approach.

The second reason for a conservative estimate of the relative performance of the Bayesian random effects relative to the mixed and traditional life table methodologies is the creation of simulation datasets without any missing values in the life tables (i.e. for all but the smallest survey sample sizes, which have less respondents than age groups). As shown in our real-life example, such missing values are very common in real-life small-area analyses and the Bayesian random-effects methodology can adequately handle extremely sparse life table data by pooling strength between sexes, age groups and adjacent areas. The traditional life table methodology, in contrast, breaks down and requires further aggregation or imputation to avoid numerical problems, whereas the Murray et al. mixed-effects methodology over-smooths towards the predicted HLE based on the observed values of the covariates. Consequently, the balanced construction of the simulation datasets -without missing values- places insufficient attention to the problems that the traditional life table and Murray et al. mixed-effects methodology encounter in real-life applications.

Another particular advantage of the Bayesian random-effects methodology is that it does not rely on covariates. Accordingly, in contrast to the Murray et al. mixed effects approach, the performance of the random-effects methodology does not depend on the identification and availability of covariate data. Moreover, the estimated HLE remain unaffected by the observed covariates in the areas, which is particularly beneficial when the estimated HLE are intended to be used in subsequent explanatory analyses. Needless to say, any random-effects approach can easily be extended to include covariates. Depending on the availability of covariates and purpose of the HLE estimates this could certainly be a prudent option; for example, to account for the location of nursing homes, which is known to be an important confounder in small-area life table analyses.[73,74]

Based on the presented simulation evidence and on the results of the real-life example, the available survey sample size seems to be the dominant factor in determining the standard error of small-area HLE estimates. This implies that the inclusion of population sizes of 1,750 person years at risk is still perfectly acceptable. For the minimum required survey sample size we consider a lower threshold of approximately 45 respondents per life table appropriate. This threshold results in slightly larger standard errors than recommended by Jonker et al.[72,73] for regular LE, but this seems acceptable for two reasons. First, the range of estimated small-area HLE is substantially larger than for LE, implying that slightly larger standard errors do not result in less statistical power to differentiate between areas. Second, in contrast to regular LE, the coverage of the 95% CI remains highly accurate for even the

smallest survey sample sizes. Therefore researchers may decide to include areas with even smaller survey sample sizes, as long as the 95% CI are appropriately taken into account.

A potential limitation of the proposed small-area HLE methodologies is that they are based on Sullivan's method to calculate HLE. Sullivan's method provides consistent and unbiased estimates of healthy life expectancy under the standard stationarity assumptions as imposed by regular period life tables and the assumption that the morbidity prevalences remain constant over time. The latter is less stringent than the assumption of stationarity of the transition probabilities between health states and usually not unreasonable once the stationarity assumptions of the regular period life table are invoked.[68] However, it does imply that real-life HLE estimates are only unbiased and consistent if the underlying age-specific mortality rates and morbidity prevalences evolve without sudden and substantial changes (see e.g. Mathers and Robine[77]). If these assumptions are likely to be violated, repeated cross-sections and explicit modeling of the age-specific mortality rates and morbidity prevalences may be required, as put forward by Imai and Soneji.[68] The latter can be easily incorporated in the Bayesian random-effects methodology, should it be necessary, but under normal circumstances the included model code in Appendix A does not require further modifications and provides a reliable method to estimate small-area HLE in real-life applications.

Appendix A. OpenBUGS codes

```
# Model #2. Murray et al. mixed-effects specification
# sex  s = 1 (males)
# area i = 1..N
# age groups x = 1..X
# survey age groups x = 1..M

model{
 for (i in 1:N){
  for (x in 5:X){
   gamma[i,x] <- pop[1,i,x] * mrate[i,x]
   deaths[1,i,x] ~ dpois(gamma[i,x])
   log(mrate[i,x]) <- b0[x] +b1[x]*gdp[i] +b2[x]*educ[i] +b3[x]*pov[i] +u[x,i]
  }
  for (x in 1:M){
   healthy[1,i,x] ~ dbin( prob[i,x] , survey[1,i,x])
   logit(prob[i,x]) <- min(25, b4[x] + b5[x]*gdp[i] + b6[x]*educ[i] + b7[x]*pov[i] + v[x,i])
 }}

 # b priors
 for (x in 5:X){
  b0[x] ~ dflat()
  b1[x] ~ dflat()
  b2[x] ~ dflat()
  b3[x] ~ dflat()}
 for (x in 1:M){
  b4[x] ~ dflat()
  b5[x] ~ dflat()
  b6[x] ~ dflat()
  b7[x] ~ dflat()}

 # u prior
 for (x in 5:X){
  u[x,1:N] ~ car.normal(adj_geo[],weight_geo[],numNeigh_geo[], tau_u[x])
  tau_u[x] ~ dgamma(1,0.001)
 }

 # v prior
 for (x in 1:M){
  v[x,1:N] ~ car.normal(adj_geo[],weight_geo[],numNeigh_geo[], tau_v[x])
  tau_v[x] ~ dgamma(1,0.001)}

 for (i in 1:sumNumNeigh_geo){
  weight_geo[i] <- 1
}}
```

```
# Model #3. Jonker et al. random-effects specification
# sex  s = 1,2
# area i = 1..N
# ages x = 1..X
# survey age groups = 1..M

model{
 for (s in 1:2){
  for (i in 1:N){
   for (x in 1:X){
    gamma[s,i,x] <- pop[s,i,x] * mrate[s,i,x]
    deaths[s,i,x] ~ dpois(gamma[s,i,x])
    log(mrate[s,i,x]) <- b0[s] + b1[s,x] + b2[s,i]*b3[s,x]
   }
   for (x in 1:M){
    healthy[s,i,x] ~ dbinomial( prob[s,i,x] , survey[s,i,x])
    logit(prob[s,i,x])  <- min(25, b4[s] + b5[s,x] + b5[s,x]*b5[s,x]*b6[s,i]+ b7[s,i] )
 }}}

 # b0 & b4 priors
 for (s in 1:2){
  b0[s]  ~ dflat()
  b4[s]  ~ dflat()
 }

 # b1 prior
 b1[1:2,1:X] ~ mv.car(adj_b1[], weight_b1[], numNeigh_b1[], tau_b1[,] )
 tau_b1[1:2 ,1:2] ~dwish(Q[,] ,2)

 for (q1 in 1:2){
  for (q2 in 1:2){
   Q[q1,q2] <- equals(q1,q2)
 }}

 adj_b1[1] <- 2
 weight_b1[1] <- 1
 numNeigh_b1[1] <- 1

 for(x in 2:(X-1)) {
  adj_b1[2+(x-2)*2] <- x-1
  adj_b1[3+(x-2)*2] <- x+1
  weight_b1[2+(x-2)*2] <- 1
  weight_b1[3+(x-2)*2] <- 1
  numNeigh_b1[x] <- 2
 }

 weight_b1[(X-2)*2 + 2] <- 1
```

```
 adj_b1[(X-2)*2 + 2] <- X-1
 numNeigh_b1[X] <- 1

 # b5 prior
 b5[1:2,1:M] ~ mv.car(adj_b5[], weight_b5[], numNeigh_b5[], tau_b5[,] )
 tau_b5[1:2 ,1:2] ~ dwish(Q[,] ,2)

 adj_b5[1] <- 2
 weight_b5[1] <- 1
 numNeigh_b5[1] <- 1

 for(x in 2:(M-1)) {
  adj_b5[2+(x-2)*2] <- x-1
  adj_b5[3+(x-2)*2] <- x+1
  weight_b5[2+(x-2)*2] <- 1
  weight_b5[3+(x-2)*2] <- 1
  numNeigh_b5[x] <- 2
 }

 weight_b5[(M-2)*2 + 2] <- 1
 adj_b5[(M-2)*2 + 2] <- M-1
 numNeigh_b5[M] <- 1

 #b2 prior -> adj_geo[], numNeigh_geo[], and sumNumNeigh_geo are loaded as data
 b2[1:2,1:N]  ~ mv.car(adj_geo[],weight_geo[],numNeigh_geo[],tau_b2[,] )
 tau_b2[1:2,1:2] ~ dwish(Q[,],2)
 for (i in 1:sumNumNeigh_geo) {
  weight_geo[i] <- 1
 }

 #b7 prior
 b7[1:2,1:N]  ~ mv.car(adj_geo[],weight_geo[],numNeigh_geo[],tau_b7[,] )
 tau_b7[1:2,1:2] ~dwish(Q[,],2)

 # b3 & b6  prior
 for (s in 1:2){
  for (x in 1:X){
   f3[s,x] ~ dgamma(1,1)
   b3[s,x] <- f3[s,x] / mean(f3[s,])
  }
  for (i in 1:N){
   b6[s,i] ~ dlnorm(mu_b6[s],tau_b6[s])
  }
  tau_b6[s] ~dgamma(1,0.01)
  mu_b6[s] ~ dflat()
}}
```

4

The impact of nursing homes on small-area life expectancies

with F.J. van Lenthe, B. Donkers, P. Congdon, A. Burdorf and J.P. Mackenbach.
Health & Place (2013)

ABSTRACT

The geographic distribution of nursing homes can significantly distort small-area life expectancy estimations. Consequently, uncorrected life expectancies should not be used for small-area life expectancy comparisons. Several nursing home corrections have been proposed. The practical use of these corrections, however, is severely limited by data availability. This paper introduces a new, model-based nursing home correction that requires considerably less detailed nursing home data. A formal comparison shows that the proposed model-based approach is the recommended correction for all small-area life expectancy estimations where detailed previous residential address information of the nursing home population is not available. This makes the approach relevant for a wide range of empirical applications.

4.1 INTRODUCTION

There are several measures that summarize the mortality experience of a population, such as standardized mortality ratios, comparative mortality figures, life expectancies, and directly or indirectly age standardized rates. All of these measures aim to express mortality in a single number that can be compared across populations and over time. Accordingly, these measures facilitate the exploration of geographical variations in health and the monitoring of populations' health and health inequalities over time.

Based on their simpler interpretation and direct age-standardization, life expectancies are usually preferred over alternative summary measures (e.g. Silcocks et al.[52]). Life expectancies, however, are thus far only occasionally used in small-area analyses. An important reason is that traditional life table methodology cannot be used to calculate reliable life expectancy estimates for populations smaller than approximately 5000 person years at risk. Below this threshold, the bias in the estimates as well as the size of the standard errors becomes too large for meaningful interpretation.[36,37] Using a Bayesian random effects approach, however, it is possible to calculate accurate life expectancies and accompanying confidence intervals for life table populations as small as 2000 person years at risk.[57,72] This is achieved using a modeling approach that recognizes correlations (i.e. borrows strength) between different age groups, geographic areas, and genders.

A major advantage of the Bayesian random effects approach, next to its ability to estimate accurate life expectancies for much smaller populations, is that it is easily extendible to incorporate corrections for area-specific confounders, such as, for example, the location of nursing homes. The latter affects life expectancy through the age-specific mortality rates of nursing home residents that are substantially higher than those of non-nursing home residents.[83,84] In fact, the asymmetric geographical distribution of nursing homes has been shown to have a large influence on small-area life expectancies: areas without nursing homes appear relatively healthy whereas areas with nursing homes appear relatively unhealthy, which can seriously distort small-area life expectancy comparisons.[64]

Hence several strategies have been proposed to correct for the impact of nursing home deaths. Ideally, the life table data that constitute the life expectancy calculations are corrected by making use of previous residential address information. Under this approach, all nursing home residents and nursing home deaths are re-assigned to their previous residential address prior to the creation of the life table data. This approach

is used by Veugelers and Hornibrook[85] but its application is difficult because the required individual data are often unavailable or of insufficient quality.

When previous address information is unavailable it might still be possible to exclude the entire nursing home population from the life table data. This approach has been suggested by Williams et al.[64] and only requires aggregated data about the size of the nursing home population and number of nursing home deaths. It has the disadvantage, however, that it results in small area life expectancies that only represent the health status of the healthier subset of people that does not reside in nursing homes. Consequently, this correction results in small-area life expectancies that are incomparable to those calculated at the regional or national level, where nursing home residents are not excluded from the life table data.

Accordingly, when previous address information is not available it might be preferable to correct the small area life expectancy estimates in a different way. A theoretically attractive approach is to take into account that nursing home deaths only affect the age-specific mortality rates of older age groups and hence to model the impact of nursing home deaths directly in the life table calculations for age groups of 65 years and older. Such a correction is easily incorporated in a Bayesian random effects approach, which has the additional advantage that all uncertainty regarding the correction is also accurately reflected in the estimated standard errors and 95% confidence intervals. Most importantly, however, is that this correction only requires information about the total percentage of nursing home deaths in each area. This is considerably less detailed and consequently easier to obtain than the data for the other two corrections.

The primary aim of this paper is hence to describe the impact of nursing homes on small-area life expectancies and to present evidence on the relative performance of nursing home corrections. Based on this evidence it becomes possible to provide guidance on which methodology to use and in which situation (e.g. based on the available data). The secondary aim of this paper is to provide a coherent framework for estimating reliable life expectancies at the small-area level that can be used in a wide range of small-area applications.

4.2 METHODS

Study setting and small-area level

Our study setting is the metropolitan agglomeration of Amsterdam, the capital city of the Netherlands. The metropolitan agglomeration of Amsterdam has a population

Figure 4.1 Geographic location of the metropolitan agglomeration of Amsterdam.

of approximately 3.1 million people, is located in the western part of the country (see Figure 4.1) and has stable geographical borders for the period under investigation (i.e. 2007–2009). The small-area definition used is the official neighborhood-level as established by Statistics Netherlands. Dutch neighborhoods adhere to physical obstacles that break up urban landscapes (important traffic arteries, bodies of water, green spaces, etc.) and are relatively homogeneous in terms of type of housing and inhabitants. Neighborhoods are also the smallest geographical unit in the Netherlands for which routinely gathered data are available. Note that a number of neighborhoods has been excluded from the analysis: either because they have populations that are too small for reliable life expectancy estimations (i.e. smaller than 2000 person years at risk, cf. Jonker et al.[72]) or because they comprise a public hospital, city park, zoo,

train station, or one of the various industrial and harbor areas and consequently have no population at all.

Population and mortality data

The population and mortality data cover the 2007–2009 period and are obtained from the Dutch population (GBA) and deaths (DO) registrations. Both databases are maintained by Statistics Netherlands and provide complete and continuous coverage of all Dutch inhabitants. The GBA includes current as well as previous address information (dating back until 1995), which implies that more than 98% of the nursing home residents can be re-assigned to their previous residential address. In contrast, Veugelers and Hornibrook[85] could retrieve previous address information for approximately 80% of the nursing home population. It should also be noted that the GBA only records address-changes when the intention is to stay at the new address for at least several months. Accordingly, people who are admitted to a nursing home in the last few weeks before their imminent death are not registered as being part of the nursing home population.

For the life table population data, midyear population counts were extracted from the GBA at the first of July of 2007, 2008, and 2009 and the sum of these counts was taken as the population at risk for the 2007–2009 period. For the life table mortality data, all deaths in the same 3-year period were aggregated. The population and mortality data were differentiated by sex, neighborhood, and age and converted into standard 5-year abridged life tables with $\geq$95 as the final age group. In total, three different life table datasets have been constructed: one standard (uncorrected), one in which all nursing home residents and nursing home deaths are fully excluded from the data, and, finally, one in which all nursing home residents and nursing home deaths are re-assigned to their previous residential addresses. The descriptive statistics of these life table data are provided in Table 4.1.

Nursing home data

To be able to exclude all nursing home residents and deaths from the life table data, assign all nursing home residents and deaths to their previous residential address, and calculate the percentage of individuals who were registered at a nursing home address in the GBA at the time of their death, a list of institutional addresses has been obtained from Statistics Netherlands. This list contains old age and nursing home locations—only available as a single category—at two different points in time: January 1st of 2008 and January 1st of 2009.

Table 4.1 Descriptive statistics of the life table data and the percentage of nursing home deaths, metropolitan agglomeration of Amsterdam, 2007-2009

Sex	Variable	Mean	Std.Dev.	Min	Max
Male	Population at risk (uncorrected)	9997	6768	2060	36,361
	Population at risk (nursing home excluded)	9958	6751	2060	36,261
	Population at risk (previous address adjusted)	9994	6771	2061	36,358
	Number of deaths (uncorrected)	74	60	5	297
	Number of deaths (nursing home excluded)	64	48	3	233
	Number of deaths (previous address adjusted)	74	55	3	268
	% of nursing home deaths*	8.9	15.9	0.0	77
Female	Population at risk (uncorrected)	10,358	6989	2077	36,890
	Population at risk (nursing home excluded)	10,260	6941	2077	36,647
	Population at risk (previous address adjusted)	10,352	6990	2092	36,854
	Number of deaths (uncorrected)	82	86	1	491
	Number of deaths (nursing home excluded)	58	50	1	261
	Number of deaths (previous address adjusted)	81	68	1	352
	% of nursing home deaths*	15.3	24.6	0.0	86.6

Unlike the GBA and DO registrations, the list of institutional addresses is solely based on information supplied by the Dutch municipalities without being audited or verified by Statistics Netherlands. Due to privacy regulations, only address-numbers (instead of verifiable addresses) were available to researchers. This makes it impossible to verify the list of institutional addresses using external information. To still establish the accuracy of the nursing home data, a separate regression model was implemented that predicts the most likely function of each address using available characteristics and the information of the two measurements per address, while allowing for misclassification errors (see Appendix A). Based on these results, addresses have been classified as a nursing home address when at least one of the measurements indicated that the address was a nursing home address, and classified as a residential (i.e. non-nursing home) address otherwise.

Life expectancy calculations

Following the recommendations and small-area life table methodology of Jonker et al.,[72] small-area life expectancies are calculated using a Bayesian random effects approach that is sufficiently flexible to allow for age–area interactions. In contrast to the traditional, fixed effects approach to estimating small-area life expectancies (e.g. Eayres and Williams[37] and Scherbov and Ediev[36]) this approach produces more

reliable life expectancy estimates and smaller and more reliable confidence intervals for populations as small as 2000 person years at risk.

The life expectancies for neighborhoods in the metropolitan agglomeration of Amsterdam were estimated in WinBUGS using iterative Markov Chain Monte Carlo (MCMC) sampling techniques. The estimations started with 25,000 burn-in MCMC iterations to allow the chains to converge, followed by 50,000 MCMC iterations with a thinning interval of 10 to reliably approximate the posterior life expectancy distributions. Convergence was evaluated using the Gelman–Rubin criteria based on two parallel chains.[61]

In total, four different sets of life expectancies have been calculated: one uncorrected and three corrected for the location of nursing homes. The first set of corrected life expectancies is based on the life table data where the nursing home population is reassigned using the previous residential address information. This is the most accurate nursing home correction. The second set of life expectancies is based on the life table data where the entire nursing home population has been removed from the data and the third is based on an extension of the Bayesian modeling approach of Jonker et al.[72] that uses the uncorrected life table data and the aggregated percentage of nursing home deaths per neighborhood to adjust the age-specific mortality rates of males and females prior to the life expectancy calculations.

The Bayesian model-based nursing home correction only adjusts the age-specific mortality rates for the age groups of 65 years and older and leaves younger age-groups unaffected (see equations C.1 and C.2 in Appendix C). Accordingly, the Bayesian nursing home correction only affects the life table calculations where it theoretically makes sense. A linear specification has been adopted in which the effect of nursing home deaths on the age-specific mortality rates of age groups of 65 years and older is allowed to be sex and age-group specific.

Comparison of nursing home corrections

The performance of the nursing home corrections is investigated using several goodness-of-fit measures. These are the mean error (ME), mean absolute error (MAE), root mean square error (RMSE), and the minimum and maximum error. All of these measures are calculated using the previous address corrected life expectancies as the benchmark; these represent the ideal nursing home correction and our aim is to see if whether this correction is adequately approached by the alternative corrections.

The ME indicates whether the corrected life expectancies are biased, i.e. whether they are disproportionately positive or negative compared to the benchmark. Ideally, the estimated life expectancies are on average right, which implies a ME of zero. Besides bias also the typical magnitude of the error is of interest. Both the MAE and RMSE are measures of the typical magnitude of the error but they are different in the way the errors are weighted. The mean absolute error weights all errors equally, whereas the RMSE gives increasingly larger weights to larger errors due to the squaring process. Accordingly, when both measures are equal, the variance of the errors is constant and the larger the difference, the larger the variance of the errors. It should be noted that the choice between MAE and RMSE only makes a real difference when the variance of the errors is large; in most cases, the MAE and RMSE vary in unison and the correction with both the lowest MAE and RMSE is preferred.

4.3 RESULTS

In the metropolitan agglomeration of Amsterdam, approximately one-third (34%) of the neighborhoods contains an old age or nursing home and the majority of nursing home residents (84%) comes from a different neighborhood. This implies that inter-area migration of elderly to nursing homes is substantial. Furthermore, Table 4.2 shows that the age-specific mortality rates of the nursing home population are indeed substantially higher than those of the non-nursing home population. Accordingly, nursing homes should theoretically have a strong impact on small-area life expectancy in the metropolitan agglomeration of Amsterdam.

This is confirmed in Figure 4.2, which provides a graphical presentation of the relationship between the estimated small-area life expectancies and the percentage of nursing home deaths. As can be seen, the uncorrected life expectancies are indeed strongly related to nursing home deaths. For every 10% points increase in nursing home deaths, estimated male and female life expectancies decrease with approximately 0.6 years. In contrast, the nursing home corrected life expectancies show no relationship with the percentage of nursing home deaths. Furthermore, the nursing home corrected life expectancies also seem to have a considerably smaller range than the uncorrected life expectancies. The latter is confirmed in Table 4.3, which provides the descriptive statistics of the estimated life expectancies.

Table 4.4 summarizes the statistical performance of the uncorrected, nursing home excluded, and Bayesian corrected life expectancies relative to the benchmark life expectancies that are corrected using the previous residential address information.

Figure 4.2 The relationship between the percentage of nursing home deaths and male and female life expectancy of neighborhoods in the metropolitan agglomeration of Amsterdam, 2007–2009

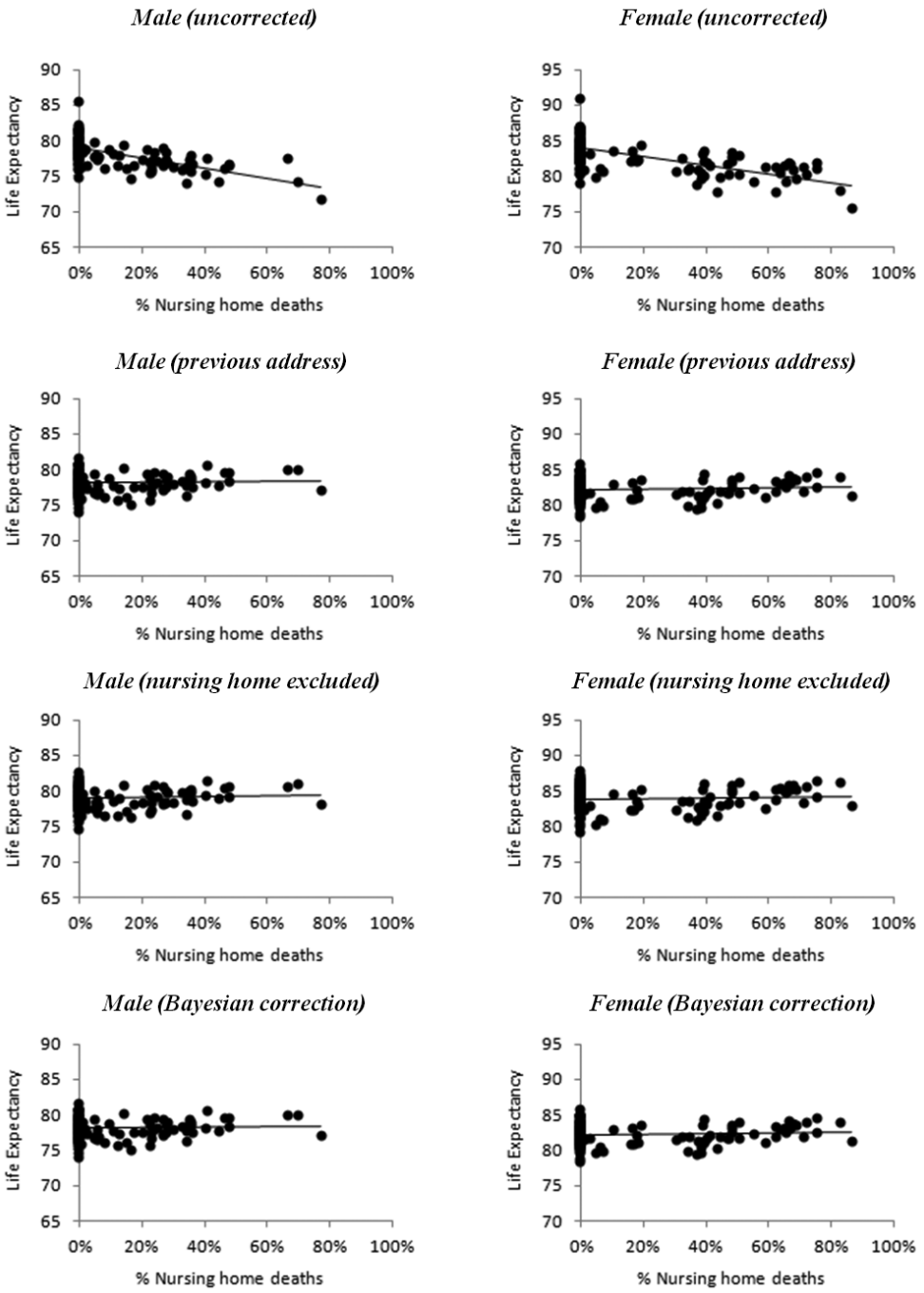


Table 4.2 Comparison of age-specific mortality rates in the metropolitan agglomeration of Amsterdam, 2007-2009

Sex	Age-group	Population	
		Non-nursing home	Nursing home
Male	65 - 74	0.02	0.19
	75 - 84	0.06	0.31
	85 - 94	0.13	0.37
	≥95	0.27	0.57
Female	65 - 74	0.01	0.14
	75 - 84	0.03	0.19
	85 - 94	0.09	0.28
	≥95	0.25	0.43

Starting with the uncorrected life expectancies, a mean error of 0.23 and 0.57 years for males and females seems modest. The average error, however, does not adequately reflect the differences between neighborhoods with and without nursing homes. These differences are better reflected by the MAE and RMSE because they do not allow positive and negative errors to average out. The MAE and RMSE show that uncorrected life expectancies are typically off by 1.11 and 1.44 years (MAE) and 1.44 and 1.77 years (RMSE) for males and females, respectively. Furthermore, the minimum and maximum errors are considerably larger than the average error, with maximum absolute errors of 5.62 years for males and 6.05 years for females. Compared with the

Table 4.3 Summary statistics of the estimated life expectancies (in years), metropolitan agglomeration of Amsterdam, 2007–2009

Sex	Variable	Mean	Std.Dev.	Min	Max	Range
Male	Life expectancy (uncorrected)	78.27	1.93	71.71	85.38	13.7
	Life expectancy (previous address corrected)	78.04	1.44	74.01	81.60	7.59
	Life expectancy (nursing home excluded)	78.95	1.50	74.61	82.69	8.08
	Life expectancy (Bayesian correction)	77.96	1.58	73.48	82.01	8.53
Female	Life expectancy (uncorrected)	82.79	2.21	75.35	90.86	15.5
	Life expectancy (previous address corrected)	82.22	1.44	78.3	85.64	7.34
	Life expectancy (nursing home excluded)	83.87	1.69	79.06	87.83	8.77
	Life expectancy (Bayesian correction)	82.33	1.63	77.87	86.42	8.55

range of nursing home corrected life expectancies, which is approximately 8 years (see Table 4.3), the impact of nursing homes on uncorrected small-area life expectancies is very substantial.

Turning to the life expectancies based on the fully excluded nursing home population, the mean error (or bias) is several times higher than mean error of the uncorrected life expectancies. The mean error is also equal to the mean absolute error, which implies that all neighborhoods (both with and without nursing homes) have higher life expectancies than the benchmark corrected life expectancies. Furthermore, the RMSE is only marginally larger than the MAE, which indicates that the variance in the errors is very small. Accordingly, the correction itself works very well in removing between-neighborhood variation due to nursing home locations, but also results in upwards biased life expectancy estimates.

Table 4.4 Comparison of the performance of uncorrected, fully nursing home excluded and Bayesian nursing home corrected life expectancies relative to previous address corrected life expectancies (in years) in the metropolitan agglomeration of Amsterdam, 2007–2009.

Methodology	Measure	Sex	
		Male	Female
Uncorrected	ME	0.23	0.57
	MAE	1.11	1.44
	RMSE	1.44	1.77
	Min Error	-5.62	-6.05
	Max Error	3.78	5.23
Fully excluded	ME	0.91	1.65
	MAE	0.91	1.65
	RMSE	0.97	1.71
	Min Error	0.17	0.55
	Max Error	3.13	3.90
Bayesian Correction	ME	-0.08	0.11
	MAE	0.37	0.42
	RMSE	0.56	0.59
	Min Error	-2.49	-2.36
	Max Error	2.08	2.29

The Bayesian nursing home correction, finally, performs much better than the uncorrected and nursing home excluded life expectancies: it produces almost unbiased estimates and results in typical errors that are approximately two to three times smaller than those of the alternative approaches. Moreover, the maximum absolute errors of the Bayesian correction are also considerably smaller. On the other hand, even though the Bayesian correction performs very well, it does have a typical error of approximately 0.4 years and a maximum absolute error of roughly 2.5 years compared to the benchmark correction. From this perspective, the Bayesian correction is certainly not a perfect substitute for the previous address correction.

4.4 DISCUSSION

This study confirms that the impact of nursing home locations on small-area life expectancies is substantial and cannot be neglected when interpreting life expectancies as a summary indicator of small-area population health. In the metropolitan agglomeration of Amsterdam, neighborhoods with nursing homes have uncorrected life expectancies that are up to six years lower due to the selective migration of frail individuals to nursing homes. Similarly, neighborhoods without nursing homes have uncorrected life expectancies that are up to four or five years too high. Compared to the range of nursing home corrected life expectancies, which is approximately eight years for both males and females, this is indeed substantial and implies that the location of nursing homes can seriously distort comparisons between small-area life expectancies.

The results also indicate that the fully nursing home excluded life expectancies do not provide a good alternative to the uncorrected life expectancies. By excluding the entire nursing home population from the data, the estimations are restricted to the healthier subset of inhabitants that do not reside in nursing home and this results in upwards biased estimates. The correction itself, however, does adequately control for the impact of nursing homes. Hence the nursing home excluded life expectancies could theoretically be used for small-area health comparisons, as long as comparisons between genders are avoided (because the magnitude of the bias is shown to be sex-specific).

With much less data required - only the average percentage of nursing home deaths in each area - the Bayesian correction performs much better than the nursing home excluded life expectancies. The Bayesian correction results in close to unbiased life expectancies and relatively small errors compared to the ideal correction based on the

previous residential address information. In typical applications, where the detailed individual data that are required for the previous residential address correction are unavailable, the Bayesian correction is thus the recommended approach. Still, it is certainly not a perfect substitute to the benchmark previous address correction, which implies that the Bayesian correction is only preferred when the required data for the previous residential address correction are unavailable.

Several refinements of the Bayesian modeling approach can be envisaged. First of all, with more detailed nursing home data (e.g. sex-specific instead of aggregate percentages of nursing home deaths per neighborhood) the Bayesian nursing home correction would be better able to differentiate between male and female mortality differences. Secondly, with more detailed data about the non-nursing home population in each neighborhood, the corrected mortality rates could be modeled to reflect the age-distributions of the non-nursing home populations in each neighborhood. This would allow the Bayesian correction to account for the smaller impact of nursing homes on neighborhoods with younger populations. Thirdly, in case the percentage of nursing home deaths is correlated with measures of social economic status (SES), the Bayesian modeling approach can take the impact of both nursing homes and SES into account by including additional explanatory variables into equation C.2 in Appendix C. This is not required for the metropolitan agglomeration of Amsterdam because SES and the percentage of nursing home deaths are insignificantly correlated (see Table 4.5). However, significant correlations between the location of nursing homes and SES have been reported by Nimmo et al.[84] and Gandarillas et al.[83] and in these situations such a refinement may be warranted.

In conclusion, the impact of nursing homes on small-area life expectancy estimations is substantial and uncorrected life expectancies should not be used for inter-area comparisons if the estimated life expectancies are to be interpreted as indicators of average population health. If the required individual data are available, the optimal correction is one in which all nursing home residents are re-assigned to their previous residential address prior to the life expectancy estimations. However, in many applications the required data will not be available and in these cases the Bayesian model-based correction provides a reliable alternative with relatively modest data requirements.

Table 4.5 Correlations between the percentage of nursing home deaths and several indicators of social economic status, metropolitan agglomeration of Amsterdam, 2007-2009

SES-indicator	Correlation*
Average income per capita	-0.04 (0.68)
Percentage on social welfare provision	-0.08 (0.33)
Average house price	-0.08 (0.33)

** Note: p-values in parenthesis*

Appendix A. Nursing home classification

To be able to assign all nursing home residents to their previous residential address and to calculate the percentage of individuals who are registered at a nursing home at the time of their death, a list of institutional addresses has been obtained from Statistics Netherlands. This list contains old age and nursing home locations—only available as a single category—at two different points in time: January 1st of 2008 and January 1st of 2009.

Unlike the GBA and DO registrations, the list of institutional addresses is solely based on information supplied by the Dutch municipalities without being audited or verified by Statistics Netherlands. Due to privacy regulations, only address-numbers (instead of verifiable addresses) are available to researchers. This makes it impossible to verify the list of institutional addresses using external information. Hence we allow statistical considerations to play an important role in establishing the accuracy of the old age and nursing home data.

When measurement errors are not taken into account, 99.9% of the addresses can already be correctly classified in each year based on the average number of inhabitants per address, their average age, an interaction between age and number of inhabitants and the number of deaths per address. This high classification score reflects the fact that old age and nursing home addresses differ fundamentally from (the vast majority of) regular residential addresses. The challenge, however, lies in correctly differentiating old age and nursing homes from other institutions, such as rehabilitation centers, mental institutions, penitentiaries, religious communities, etc., which are more difficult to distinguish based on the available characteristics per address. Consequently, the estimations are further improved by taking the information contained by the two separate measurements of each address into account.

Starting with the standard likelihood function for a logit-model (with Y denoting the vector of observed measurements, X denoting the matrix of explanatory variables and b the regression coefficients to be estimated),

$$\begin{aligned} f(Y, Xb) &= 1/(1 + \exp(\text{-}Xb)) && \text{if } Y = 1 \\ &= \exp(\text{-}Xb)/(1 + \exp(\text{-}Xb)) && \text{if } Y = 0 \end{aligned} \tag{A.1}$$

the notation is first simplified as:

$$\begin{aligned} f(Y, Xb) &= \text{invlogit}(Xb) && \text{if } Y = 1 \\ &= \text{invlogit}(\text{-}Xb) && \text{if } Y = 0 \end{aligned} \tag{A.2}$$

Now suppose that the actual Y is not directly observed but instead approximated by a measurement Y^* with measurement error P. In this case, Y is correctly observed with probability (1-P) and incorrectly observed with probability P. Accordingly, assuming a single measurement and a symmetrical misclassification error P the following likelihood can be constructed:

$$\begin{aligned} f(Y, Xb) &= \text{invlogit}(Xb) \times (1\text{-}P) + \text{invlogit}(\text{-}Xb) \times P && \text{if } Y^* = 1 \\ &= \text{invlogit}(Xb) \times P + \text{invlogit}(\text{-}Xb) \times (1\text{-}P) && \text{if } Y^* = 0 \end{aligned} \tag{A.3}$$

Because we have two measurements of Y, the likelihood becomes slightly more complicated:

$$\begin{aligned} f(Y, Xb) &= \text{invlogit}(Xb) \times (1\text{-}P) \times (1\text{-}P) + \text{invlogit}(\text{-}Xb) \times P \times P && \text{if } Y_1^* = Y_2^* = 1 \\ &= \text{invlogit}(Xb) \times P \times P + \text{invlogit}(\text{-}Xb) \times (1\text{-}P) \times (1\text{-}P) && \text{if } Y_1^* = Y_2^* = 0 \\ &= (\text{invlogit}(\text{-}Xb) \times P \times (1\text{-}P)) && \text{if } Y_1^* \neq Y_2^* \end{aligned} \tag{A.4}$$

Note that this structure resembles a single latent variable Y and two observed measurements (Y_1^* and Y_2^*) rather than two independent observations of Y; in the latter case the likelihood would be the multiplication of the single-observation likelihoods.

In the final step, we introduce the possibility of an asymmetrical measurement error with $P_{0\,1}$ denoting the probability that a zero is observed while Y equals one, and $P_{1\,0}$ denoting the probability that a one is observed while Y equals zero. In this case, the likelihood looks like:

$$\begin{aligned} & f(Y, Xb) \\ &= \text{invlogit}(Xb) \times (1\text{-}P_{0\,1}) \times (1\text{-}P_{0\,1}) + \text{invlogit}(\text{-}Xb) \times P_{1\,0} \times P_{1\,0} && \text{if } Y_1^* = Y_2^* = 1 \\ &= \text{invlogit}(\text{-}Xb) \times P_{0\,1} \times P_{0\,1} + \text{invlogit}(\text{-}Xb) \times (1\text{-}P_{1\,0}) \times (1\text{-}P_{1\,0}) && \text{if } Y_1^* = Y_2^* = 0 \\ &= 2 * (\text{invlogit}(Xb) \times P_{0\,1} \times (1\text{-}P_{0\,1}) + \text{invlogit}(\text{-}Xb) \times P_{1\,0} \times (1\text{-}P_{1\,0})) && \text{if } Y_1^* \neq Y_2^* \end{aligned} \tag{A.5}$$

This likelihood is programmed in STATA (v11) and maximized with a standard Newton–Raphson search algorithm. After estimation, the probability that the real Y equals one can be predicted using the Xb, i.e. $P(Y = 1|Xb)$ as the inverse logit of the linear prediction Xb. This prediction, however, does not take the information

of the two measurements into account. Hence, instead of $P(Y = 1|Xb)$, it would be more informative to calculate $P(Y = 1|Xb, Y_1^*, Y_2^*)$. The latter can be re-arranged as follows:

$$P(Y = 1|Xb, Y_1^*, Y_2^*) = p(Y = 1, Y_1^*, Y_2^*|Xb)/(Y_1^*, Y_2^*|Xb) \tag{A.6}$$

and because Y can only take two values, it holds that:

$$P(Y_1^*, Y_2^*|Xb) = P(Y = 1, Y_2^*, Y_2^*|Xb) + P(Y = 0, Y_1^*, Y_2^*|Xb) \tag{A.7}$$

Substituting this into the first equation results in:

$$\begin{aligned} P(Y = 1|Xb, Y_1^*, Y_2^*) &= \frac{P(Y = 1, Y_1^*, Y_2^*|Xb)}{P(Y_1^*, Y_2^*|Xb)} \\ &= \frac{P(Y = 1, Y_1^*, Y_2^*|Xb)}{(P(Y = 1, Y_1, Y_2|Xb) + P(Y = 0, Y_1^*, Y_2^*|Xb))} \end{aligned} \tag{A.8}$$

and using Bayes' rule it holds that:

$$P(Y = 1, Y_1^*, Y_2^*|Xb) = P(Y_1^*, Y_2^*|Y = 1, Xb) \times p(Y = 1|Xb) \tag{A.9}$$

From here, the first observation is that, conditional on $Y = 1$, the Xb are irrelevant since the asymmetrical errors are assumed to be independent of the observed characteristics of Y. The second observation is that, given Y, the realizations of Y_1^* and Y_2^* are independent, which implies that:

$$P(Y_1^*, Y_2^*|Y = 1) = P(Y_2^*|Y = 1) \times P(Y_1^*|Y = 1) \tag{A.10}$$

Combining these two observations results in:

$$\begin{aligned} P(Y = 1|Xb, Y_1^*, Y_2^*) = P(Y = 1|Xb) \times P(Y_1^*|Y) \times P(Y_2^*|Y)/(P(Y = 1|Xb) \times P(Y_1^*|Y) \times \\ P(Y_2^*|Y) + P(Y = 0|Xb) \times P(Y_1^*|Y) \times P(Y_2^*|Y)) \end{aligned} \tag{A.11}$$

in which the probability distributions $P(Y_1^*|Y)$ and $P(Y_2^*|Y)$ are identical and determined by the asymetrical measurement errors as follows:

$$\begin{aligned} P(Y^* = 1|Y = 1) &= (1\text{-}P_{0\,1}) \\ P(Y^* = 0|Y = 1) &= P_{0\,1} \\ P(Y^* = 0|Y = 0) &= (1\text{-}P_{1\,0}) \\ P(Y^* = 1|Y = 0) &= P_{1\,0} \end{aligned} \tag{A.12}$$

Table 4.6 contains the results of the regression including asymmetrical measurement errors. All parameter estimates are significant and as can be seen, the misclassification error $P_{1\,0}$ is more than thousand times smaller than $P_{0\,1}$. Based on these results it is easily verified that the measurement errors are so small that the function of each address can be solely determined by the two measurements:

a) If the two measurements both denote the same function (either a regular or a nursing home address), the probability that both measurements are wrong is extremely small. Hence, regardless of the characteristics of the address, the measurements determine the most likely function.

b) b) In case the two measurements do not agree, $P_{0\,1}$ is more than thousand times larger than $P_{1\,0}$ and hence the address can be classified as a nursing home regardless of the characteristics of the address.

In conclusion, the regression results indicate that the nursing home measurements are very reliable (i.e., have very small measurement errors) and provide strong evidence that the combination of two measurements per address results in a highly reliable classification of the actual function of each address within the metropolitan agglomeration of Amsterdam.

Table 4.6 Nursing home classification regression of all addresses in the metropolitan agglomeration of Amsterdam, 2007–2009, taking asymmetrical measurement errors into account*

	Model #1	Model #2
Constant	-13.895 (0.379) ***	-6.270 (0.105) ***
Count of number of inhabitants per address	-0.438 (0.103) ***	-5.181 (0.430) ***
Count of number of deaths per address	n/a	1.280 (0.124) ***
Average age of inhabitants per address	0.099 (0.005) ***	n/a
Interaction between the number of inhabitants per address and their average age	0.011 (0.002) ***	0.070 (0.005) ***
$P_{0\,1}$	0.024 (5.6e-3) ***	0.022 (5.6e-3) ***
$P_{1\,0}$	6.9e-6 (2.8e-6) **	1.6e-5 (5.0e-6) ***

** N = 572,459 addresses, standard errors in parenthesis, n/a = not included in model specification, ** p <0.05, *** p <0.01*

Appendix B. Bayesian life expectancy estimations

Let D_{six} and P_{six} denote deaths and populations at risk classified by sex ($s = 1, 2$), area ($i = 1, \ldots, N$), and age group ($x = 1, \ldots, 21$). Deaths are assumed to be Poisson distributed:

$$D_{six} \sim \text{Poisson}(\text{P}_{six} \times \text{m}_{six}) \tag{B.1}$$

with m_{six} denoting mortality rates specified in the same dimensions. For larger populations, a binomial distribution could also be specified; however, given our focus on small populations with few observed deaths, the Poisson distribution is considered more appropriate. Following Jonker et al.[72] a standard log-link function is used and the following specification adopted:

$$\log(m_{six}) = \alpha_s + \beta_{1\,sx} + \beta_{2\,si} \times \beta_{3\,sx} \tag{B.2}$$

This specification contains:

a) overall gender-specific mortality level parameters α_s, which are assigned flat prior distributions from $-\infty$ to ∞ with the WinBUGS "dflat()" distribution,

b) parameters $\beta_{1\,sx}$ that represent the age–sex mortality rates for age x and gender s; these are assigned a multivariate conditional first-order random walk prior that takes correlations between adjacent age groups and the correlation between the mortality experience of males and females into account, and

c) area effects $\beta_{2\,si}$ that represent spatially correlated mortality contrasts that are also allowed to be gender differentiated; these are assigned a multivariate conditional autoregressive prior distribution that takes correlations between adjacent areas and the correlation between male and female mortality rates into account, and

d) a set of $\beta_{3\,sx}$ parameters that are assigned Gamma(1,1) priors, such that the parameter combination $\beta_{2\,si} \times \beta_{3\,sx}$ provides a parsimonious representation of age–sex–area mortality effects with a significantly smaller number of parameters than the full set of $s \times i \times x$ parameters.

Both β_1 and β_2 are estimated in WinBUGS using the "MV.CAR" distribution with Wishart priors assigned to the precision matrices (specified with 2 degrees of freedom and a 2-by-2 identity scale matrix). For identification, $\beta_{1\,sx}$ and $\beta_{2\,si}$ are constrained to sum to zero and $\beta_{3\,sx}$ constrained to sum to 1 over dimension s.

Appendix C. Bayesian life expectancy estimations including a nursing home correction

In Model 2, the likelihood for age groups 1 to 14 (i.e. ages 0 to 64) is identical to that of model 1. The likelihood for age groups 15 to 21 (i.e. age groups 65 and older), however, additionally includes a correction for the impact of nursing home deaths. The following specification is adopted:

for $1 \leq x \leq 14$

$$\log(m_{six}) = \alpha_s + \beta_{1\,sx} + \beta_{2\,si} + \beta_{3\,sx} \tag{C.1}$$

for $15 \leq x \leq 21$

$$\log(m_{six}) = \alpha_s + \beta_{1\,sx} + \beta_{2\,si} \times \beta_{3\,sx} + \beta_{4\,sx} \times \text{perc\_NH-deaths}_i \tag{C.2}$$

in which the variable perc_NH-deaths$_i$ represents the standardized area-specific percentage of nursing home deaths and $\beta_{4\ sx}$ the sex and age-group specific parameters that capture the impact of nursing home deaths on the estimated mortality rates, which are assigned dflat() priors. Other specifications, such as non-linear terms and a random-effects prior on the $\beta_{4\ sx}$ parameters, did not improve the performance of the nursing home correction and therefore the simpler specification in equation C.2 is preferred.

Using this specification, nursing home adjusted mortality rates are predicted for age-groups 15 to 21 based on the average (or global) percentage of nursing home deaths (i.e. the total number of nursing home deaths divided by the total number of deaths in all areas) instead of the actual percentage nursing home deaths in each individual area. These adjusted mortality rates form the input for the subsequent life expectancy estimations and hence constitute the Bayesian model-based nursing home corrected life expectancies.

5

Small-area health comparisons using health-adjusted life expectancies: A Bayesian random-effects approach

with P. Congdon, F.J. van Lenthe, B. Donkers, A. Burdorf and J.P. Mackenbach.
Health & Place (2013)

ABSTRACT

Health-adjusted life expectancy (HALE) is one of the most attractive summary measures of population health. It provides balanced attention to fatal as well as non-fatal health outcomes, is sensitive to the severity of morbidity within the population, and can be readily compared between areas with very different population age structures. HALE, however, cannot be calculated at the small-area level using traditional life table methodology. Hence we propose a Bayesian random-effects modeling approach that recognizes correlations and pools strength between sexes, age-groups, geographical areas, and health outcomes. This approach allows for the calculation of HALE for areas as small as 2000 person years at risk and with relatively modest health state survey sample sizes. The feasibility of the Bayesian approach is illustrated in a real-life example, which also shows how differences in areas' health performances can be adequately quantified. Such information can be invaluable for the appropriate targeting and subsequent evaluation of urban regeneration, neighborhood renewal, and community-based initiatives aimed at improving health and reducing health inequalities.

5.1 INTRODUCTION

Geographic inequalities in life expectancy (LE) have been well documented and provide important information on the distribution of health across countries and regions.[87,88,99] However, differences in life expectancy only describe differences in the mortality experience of populations and do not explicitly capture differences in morbidity for those still being alive.[24] Health expectancy, a concept first introduced in the 1960s and further developed in the 1970s and 1980s, provides an attractive addition to traditional life expectancies by partitioning total life expectancy into years spent in perfect health and years spent in less than perfect health.

Similar to LE, health expectancy measures are independent of the size and composition of the population and provide an intuitive and reliable comparison between the health status of one population with that of another. Health-adjusted life expectancy (HALE), a specific type of health expectancy, measures the number of expected years of life equivalent to years lived in full health. It is one of the most attractive summary measures of population health because it takes various domains of health into account and uses pre-defined utility weights to combine these into a single measure of health-related quality of life. In contrast to other health expectancy measures, such as healthy life expectancy or disability-free life expectancy, HALE is more sensitive to changes in the severity of morbidity within the population because it uses polychotomous instead of dichotomous weights.[24,100] Consequently HALE is the only measure that can provide appropriate and balanced attention to the effects of fatal as well as non-fatal health outcomes on overall population health.[103]

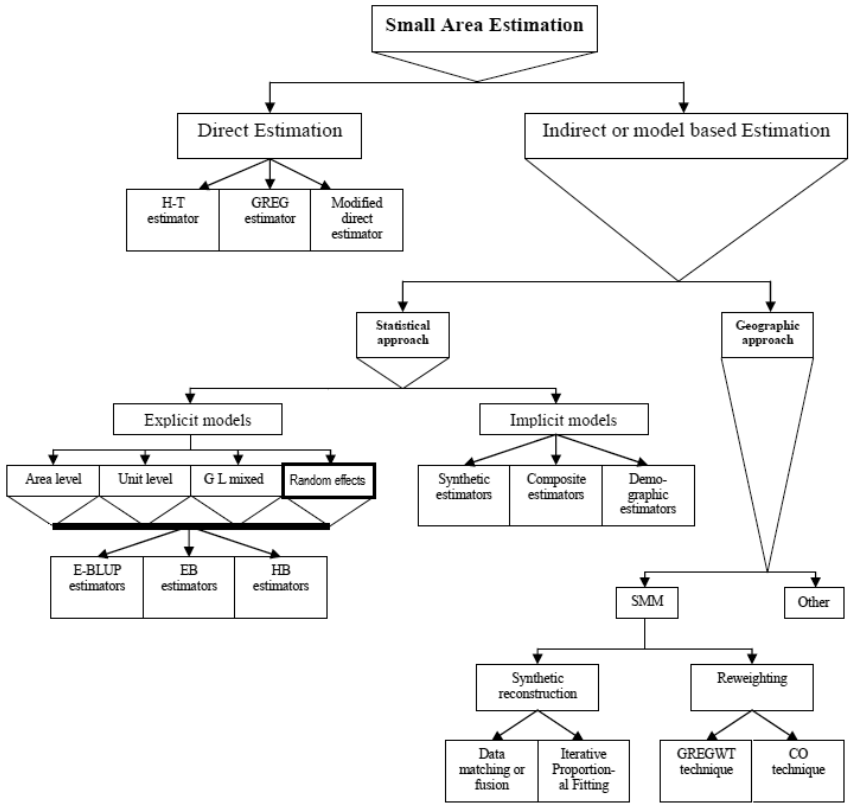
Thus far, HALE has only been calculated at the national or regional level.[97,98,100,102] Analyses for smaller areas are warranted because the calculation of HALE for more refined geographical areas allows for a more reliable approximation of the spatial distribution of health expectancy and for a study of its determinants on a smaller level than is currently possible.[94] It would also constitute an important input for the targeting and evaluation of area-based initiatives aimed at reducing health inequalities.[1] The calculation of HALE at the small-area level, however, is not possible using traditional life table methodology. Firstly, the traditional methodology is based on analysis of each area independently, with each life table having as many parameters as there are observations. This approach works well for large populations, but not for small areas—with populations smaller than 5000 person years at risk—because the bias in the estimates and size of the standard errors become too large for meaningful analysis due to sparse data problems.[36,37,55] The second barrier to the estimation

of small-area HALE is a lack of adequate health surveys. Adequate health surveys first of all comprise a health state instrument such as the EQ-5D, SF-12, or HUI3 (or any other health state classification instrument with an accompanying set of utility weights) but they also need to have a sufficiently large sample size to be representative at the small-area level. Often the latter is problematic, which results in biased HALE estimates and in confidence intervals that are too large to significantly differentiate between the areas under analysis (see e.g. Manuel et al.[100]). In fact, the inability to obtain adequate sample sizes is the most common reason for the inability to calculate HALE at the small-area level.[79]

The aim of this paper is hence to introduce a modeling approach that recognizes correlations between the mortality and morbidity information in various dimensions of the small-area data (e.g. between sexes, age groups, and adjacent geographic areas). This accommodates sparse data problems by pooling strength between the included dimensions and allows for the calculation of HALE at the small-area level using significantly smaller survey sample sizes. In contrast to existing small-area estimation techniques (see e.g. the methodological overviews of Rao[107] and Rahman[108]), our proposed random-effects modeling approach is solely based on the observed small-area data and does not include or rely on covariates. It also does not rely on assumptions about the transferability of correlations from aggregate data sources (e.g. large-scale surveys) to the small-area level. In fact, the proposed random-effects methodology belongs to a class of statistical approaches that has thus far not been recognized in the small-area estimation literature. Accordingly, Figure 5.1 contains an updated typology of existing small-area estimation techniques with the inclusion of our random-effects approach to small-area life table estimation.

The proposed modeling approach is illustrated in a real-life example, which shows the feasibility of the HALE estimations and demonstrates the sensitivity of HALE compared to regular LE estimates. Additionally, a selection of neighborhoods with the lowest estimated HALE, combined with estimates of the probability that each of the areas' HALE belongs to the worst five in the entire sample, is presented. These probabilities take the full uncertainty of the life table estimates into account and are hence more reliable than traditional rankings based on point estimates. This information is, for example, relevant for the appropriate targeting and subsequent evaluation of urban regeneration, neighborhood renewal, and community-based initiatives aimed at improving health and reducing health inequalities.[1] Finally, a small simulation study is included that supports the reliability and validity of the proposed modeling approach.

Figure 5.1 Overview of small-area estimation techniques



Abbreviations: Horvitz–Thompson (H–T), Generalized Regression (GREG), Generalized Regression and Weighting (GREGWT), Empirical Best Linear Unbiased Prediction (E-BLUP), Empirical Bayes (EB), Hierarchical Bayes (HB), Spatial Microsimulation Models (SMM) and Combinatorial Optimization (CO). Adapted from Rahman et al.[109]

5.2 METHODS

5.2.1 The HALE model

The calculation of HALE is essentially a three-step procedure. In the first step, the age-specific mortality rates for all included areas are calculated. In the second step, the age-specific health-adjustment rates are calculated. And in the third step, these two inputs are summarized into HALE via established life table calculations. The proposed HALE model reflects these three steps.

Mortality rate estimations

The first part of the HALE model estimates the mortality rates as required in the HALE life table calculations. Let D_{six} and P_{six} denote deaths and populations at risk classified by sex ($s = 1, 2$), area ($i = 1, \ldots, N$), and age group ($x = 1, \ldots, X$). Deaths are assumed Poisson distributed

$$\mathrm{D}_{six} \sim \mathrm{Poisson}(\mathrm{P}_{six} \times \mathrm{m}_{six}) \tag{5.1}$$

with m_{six} denoting the required mortality rates specified in the same dimensions. For larger populations, a binomial distribution could also be specified; however, given our focus on small populations with few observed deaths, the Poisson distribution is considered more appropriate. A standard log-link function is imposed, and, based on the simulation results of Jonker et al.,[72] the following specification is used:

$$\log(\mathrm{m}_{six}) = \beta_0 + \beta_{1\,sx} + \beta_{2\,si} + \beta_{3\,sx} \tag{5.2}$$

Here the β_0-parameters capture level differences in mortality rates for males and females, whereas the β_1-parameters represent multivariate (gender-specific) random age effects that capture the usually high correlation between the mortality rates of successive age groups and genders, and the β_2-parameters multivariate (gender-specific) random spatial effects that account for spatial clustering of mortality rates in adjacent areas while also taking the correlation between genders into account. The spatial effects are interacted with a set of β_3 parameters, which allows the spatial effects to be more pronounced for specific age-groups. The latter relaxes the assumption of an uniform (gender-specific) age-gradient for all areas while still retaining a relatively parsimonious model specification. For identification of the parameters, the β_1 and β_2

parameters are constrained to sum to zero and the β_3 parameters constrained to sum to one over dimension s.

Morbidity rate estimations

The second part of the HALE model estimates the morbidity component as required in the HALE life table calculations. Let HS_{six} denote average health state rates as measured on a continuous scale, ranging from 0 (death) to 1 (perfect health), classified by sex ($s = 1, 2$), area ($i = 1, \ldots, N$), and age group ($x = 1, \ldots, X$). Given these explicit boundaries, a beta distribution is used to model the morbidity component of the HALE calculations. As explained by Paolino,[106] Ferrari and Cribari-Neto,[92] and Smithson and Verkuilen,[110] this is statistically superior to alternative modeling approaches such as assuming normality on the log-transformed average health state rates. Hence the HS_{six} are assumed to be beta distributed

$$\text{HS}_{six} \sim \text{beta}(\alpha_{six}, \beta_{six}) \tag{5.3}$$

and based on Smithson and Verkuilen[110] we use the following re-parametrization to translate the beta shape parameters α and β into location and dispersion parameters:

$$\alpha_{six} = \text{location}_{six} \times \text{dispersion}_{six} \tag{5.4}$$

$$\beta_{six} = (\text{1-location}_{six}) \times \text{dispersion}_{six} \tag{5.5}$$

A logit-link function is used for the location sub-model, a log-link function for the dispersion sub-model (the dispersion has to be strictly positive), and for both sub-models the same random-effects structure as in the mortality model is used. The dispersion model additionally includes a count of the number of survey respondents that constitutes the corresponding average health state value. This reflects the standard relationship between sample size and variance.

$$\text{logit}(\text{location}_{six}) = \beta_{4\,s} + \beta_{5\,sx} + \beta_{2\,s+2,i} \times \beta_{6\,sx} \tag{5.6}$$

$$\log(\text{dispersion}_{six}) = \beta_{7\,s} + \beta_{8\,sx} + \beta_{9\,si} \times \beta_{10\,sx} + \beta_{11} \times \log(\text{count\_HS}_{six}) \tag{5.7}$$

As can be seen in 5.6, a shared spatial structure is specified that accommodates potential correlation between the mortality and morbidity part of the HALE model. This allows for a pooling of strength between the mortality and morbidity data for each neighborhood, which are exposed to the same risk-factors and therefore likely to be correlated.

Life table estimations

The third and final component of the HALE model summarizes the estimated age-specific mortality rates and estimated location values into life expectancies and health-adjusted life expectancies. Based on the recommendations made by Toson and Baker[55] and Eayres and Williams,[37] the life expectancies are calculated using the Chiang[35] life table approach. The health-adjusted life expectancies are subsequently calculated using Sullivan's methodology.[80] An excellent introduction to life table calculations using Sullivan's methodology, including several examples, can be found in the Sullivan's Guide (www.eurohex.eu).

Bayesian estimation and priors

Bayesian methods are used for the parameter estimations, which involves selecting prior densities for the unknown model parameters and updating those densities via the likelihood of the observed data. In the estimations, the β_0, β_4, β_7, and β_{11} parameters are assigned uninformative flat priors from $-\infty$ to ∞. The β_1, β_2, β_5, β_8, and β_9 parameters are assigned multivariate conditional auto-regressive priors with Wishart priors assigned to the precision matrices. The Wishart prior for β_2 is assigned a 4-by-4 identity scale matrix with 4 degrees of freedom, whereas the other Wishart priors are assigned a 2-by-2 identity scale matrix and 2 degrees of freedom. Finally, the β_3, β_6, and β_{10} parameters are specified with individual Gamma(1,1) priors subject to a sum-of-one constraint for identification of the parameters.

The model is fitted in OpenBUGS using iterative Markov chain Monte Carlo (MCMC) sampling techniques. The estimations start with 25,000 burn-in MCMC iterations to allow the Markov chains to converge, followed by a total of 75,000 MCMC iterations with a thinning interval of 10 to reliably approximate the posterior distributions. Convergence is evaluated using the Gelman et al.[61] criteria based on three parallel chains.

5.2.2 Real-life example

Study setting and small-area level

The proposed Bayesian modeling approach is applied to real-life data of the city of Rotterdam. Rotterdam has a population of approximately half a million people and is the second-largest city in the Netherlands. The small-area level used is the official neighborhood-definition as established by Statistics Netherlands. Neighborhoods adhere to physical obstacles that break up urban landscapes—important traffic arteries,

bodies of water, green spaces, etc—and are relatively homogeneous in terms of type of housing, inhabitants, and land use. Neighborhoods are the smallest geographical unit in the Netherlands for which routinely gathered data are available. Several neighborhoods in the city of Rotterdam (together comprising 2.9% of the total population) are excluded from the analysis: either because they have insufficient population for reliable life table estimates (i.e. smaller than 2000 person years at risk in the life tables, cf. Jonker et al.[72]) or because they comprise a public hospital, city park, zoo, airport, train station, or various industrial and harbor areas and have no population at all.

Population and mortality data

The required population and mortality data for the HALE calculations are obtained from the Dutch population (GBA) and deaths registry (DO). Both databases are maintained by Statistics Netherlands and provide complete and continuous coverage of all Dutch inhabitants. For the life table population data, midyear population counts are extracted from the GBA at the first of July of 2004, 2005, and 2006 and the sum of these counts is taken as the population at risk for the 2004–2006 period. For the life table mortality data, all deaths in the same 3-year period are aggregated. The population and mortality data are both summarized by sex, neighborhood, and age and subsequently converted into standard 5-year abridged life tables with $\geq$85 as the final age group.

Nursing home correction

The selective migration of frail individuals to nursing homes can have a substantial impact on small-area life expectancies and severely distort life expectancy comparisons between areas.[64] Accordingly, the life table data that constitute the calculations in this paper, with descriptive statistics presented in Table 5.1, are corrected using previous residential address information.[73,85] This approach requires detailed individual information about the nursing home population to re-assign all nursing home inhabitants and nursing home deaths to their last-known residential address. The required individual nursing home address data and previous address information are obtained from Statistics Netherlands.

Health-status data

The required health-status data for the HALE calculations are obtained from the 2005 Health Survey held by the local health authorities of Rotterdam (GGD). The

Table 5.1 Descriptive statistics of the life table data, by sex, for neighborhoods in Rotterdam, the Netherlands, 2004–2006

Sex	Variable	N	Mean	Std.Dev.	Min	Max
Male	Population at risk	61	13,715	7,545	2,586	36,268
	Number of deaths	61	124	110	10	597
Female	Population at risk	61	14,294	8,532	2,151	43,558
	Number of deaths	61	143	148	10	773

survey contains SF-12 health state measurements of approximately 5700 respondents aged 16 years and older, resulting in an average of 93 respondents per neighborhood. The SF-12 health state instrument is a well-validated and widely used instrument to assess respondents' health in various health domains. Even though the SF-12 is considerably shorter than the original SF-36 questionnaire, SF-12 summary scores have been found to be good predictors of the original SF-36 summary scores and have reproduced their psychometric performance.[95,96] Finally, health state preference weights, which are used to aggregate the various SF-12 health state dimensions into a single composite health index as required for the HALE estimations, are obtained from the SF-6D validation study by Brazier and Roberts.[89]

Table 5.2 presents the descriptive statistics of the SF-12 health status data. In the average SF-6D health status dataset there are 25 boundary observations (all values of exactly 1). These are handled by slightly re-scaling the health-status data, conform the recommendations made by Verkuilen and Smithson.[111] Because the GGD health state survey does not include respondents aged 15 years and younger, perfect health is assumed for the first four age groups in the HALE calculations. Alternatively, HALE at age 15–19 could be reported without making any additional assumptions.

5.2.3 Simulation study

The benchmark

To substantiate the reliability of the modeling approach and to confirm the validity of the HALE estimates in our real-life example, we have conducted a Monte Carlo simulation study. Similar to other Monte Carlo simulations aimed at validating statistical models (see e.g. Fiebig et al.[93] and Camacho et al.[90]), the benchmark or 'true' parameters are based on the estimated model parameters in our reference

Table 5.2 Descriptive statistics of the health status data, by sex and age group, for neighborhoods in Rotterdam, The Netherlands, 2004–2006.

Sex	Variable	Age group	N	Mean	Std.Dev.	Min	Max
Male	SF-12	15–24	254	0.89	0.08	0.53	1
	health status	25–44	867	0.86	0.11	0.38	1
		45–64	941	0.85	0.13	0.34	1
		65–84	547	0.84	0.14	0.34	1
		≥85	6	0.77	0.15	0.63	1
Female	SF-12	15–24	365	0.86	0.10	0.34	1
	health status	25–44	1018	0.85	0.13	0.34	1
		45–64	990	0.82	0.14	0.34	1
		65–84	677	0.8	0.14	0.40	1
		≥85	6	0.64	0.13	0.52	0.86

dataset, which is the dataset of neighborhoods in Rotterdam as described in the previous section. The aim of the simulation study is to verify whether the proposed HALE model is capable of accurately recovering the benchmark HALEs and 95% credibility intervals under the same conditions as encountered in the original dataset.

Creating the simulation datasets

The first step in our Monte Carlo study is to create a large number of new datasets that each represent a single draw from the data generating process that results from the proposed HALE model combined with the posterior means of the estimated model parameters. In total, 2000 different datasets are created. Each of these datasets has the exact same population at risk and the exact same number of survey respondents as in the reference dataset. However, conform the structure of the HALE model, the number of deaths are drawn randomly from a Poisson distribution with means set to the benchmark age-specific mortality rates and the average health-state scores drawn randomly from a beta distribution with shape parameters set equal to the benchmark values. As a result, each simulated dataset has slightly different numbers of deaths and slightly different average health-state scores.

Summarizing the model's performance

OpenBUGS is used to estimate the model parameters and HALE values for all 2000 simulated datasets. The estimated parameters and HALE values are subsequently

summarized in MATLAB. The estimated parameters and HALE values together form distributions from which reliable inferences can be made. Ideally, the HALE model is capable of exactly recovering the benchmark HALE values. This would imply that the bias, which is the average difference between the estimates and the benchmark values, is equal to zero. We also look at the coverage of the reported 95% intervals by counting how often the benchmark HALEs are located within the reported 95% credibility intervals. Values below 95% indicate that the credibility intervals are too optimistic and values greater than 95% that they are too conservative. Ideally, the coverage should be equal to 95%.

5.3 RESULTS

Table 5.3 shows the results of the estimations by summarizing the posterior LE and HALE estimates of the HALE model for the individual neighborhoods in Rotterdam. As can be seen, females have an average LE that is 4.8 years higher than that of males. In contrast, the average HALE of females is only 2 years higher. Hence when quality of life is taken into account, the health advantage of women reduces considerably. This is also reflected by the difference between LE and HALE, which represents the number of years spend in poor health. As can be seen, women spend on average 2.8 more years in poor health than men.

Also the range of estimated values differs between LE and HALE: for both males and females the range of LE is 6.2 years, whereas the range of HALE is 9.2 and 9.6 years for males and females, respectively. Health inequalities between neighborhoods are thus more pronounced when health-adjusted life expectancies are taken into account. Another important aspect of the estimations, which is not included in Table 5.3, is the average size of the standard errors. Compared to the range of 6.2 and 6.1 years in LE, the average standard errors are 0.76 for males and 0.79 for females. The average standard errors of the HALE are 0.96 and 1.06, which is, relative to the range of estimated HALE, slightly smaller.

Figure 5.3 visualizes the relationship between LE and HALE. Note that LE in this figure is calculated using only the mortality data (i.e. without taking correlations with morbidity into account), whereas HALE is calculated using the combination of mortality and morbidity data and the borrowing of strength approach as described in the previous section. Accordingly, Figure 5.3 nicely illustrates the added value of the HALE compared to LE calculations. As can be seen, LE and HALE are strongly

Table 5.3 Summary of life expectancy (LE) and health-adjusted life expectancy (HALE) estimates for neighborhoods in Rotterdam, The Netherlands, 2004-2006

Indicator	Measure	Sex	
		Male	Female
LE	Min	73.6	78.1
	Max	79.8	84.3
	Average	76.0	80.8
HALE	Min	63.3	64.9
	Max	72.5	74.5
	Average	67.0	69.0
(LE-HALE)	Min	7.2	9.8
	Max	10.4	13.7
	Average	9.0	11.8

correlated. This holds both for males and females and is to be expected because HALE is partially based on the same data. However, there is also substantial (i.e. up to 3 years) variation in HALE at any given level of LE. The latter implies that LE and HALE are similar but certainly not identical, and that the morbidity data in the HALE calculations are not overwhelmed or clearly dominated by the mortality data.

Figure 5.2 contains a map of the estimated HALE in neighborhoods in Rotterdam, separately for males and females. Several neighborhoods are not included in the HALE calculations; as mentioned, these are mostly industrial or harbor areas but also those neighborhoods that contain the public hospital, city park, zoo, airport, and central train station of Rotterdam. As can be seen, the geographic pattern of HALE is very similar for males and females. There are three distinct areas with neighborhoods that have relatively low HALEs, with the most problematic neighborhoods located south of the river (the Maas) and in the older northern neighborhoods of Rotterdam.

Table 5.4 contains the individual HALE estimates for several neighborhoods in Rotterdam. For both males and females, the 10 neighborhoods with the lowest estimated HALE are selected since these are arguably the most interesting for policy purposes—of course, other selections may be relevant in different situations. Instead of only looking at the means of the posterior HALE distributions, Table 5.4 also reports the (posterior) probability that each of the selected neighborhoods belongs to the top-5 of neighborhoods with the lowest estimated HALE. This measure takes the

Figure 5.2 Maps of estimated health-adjusted life expectancy (HALE), by sex, for neighborhoods in Rotterdam, The Netherlands, 2004–2006

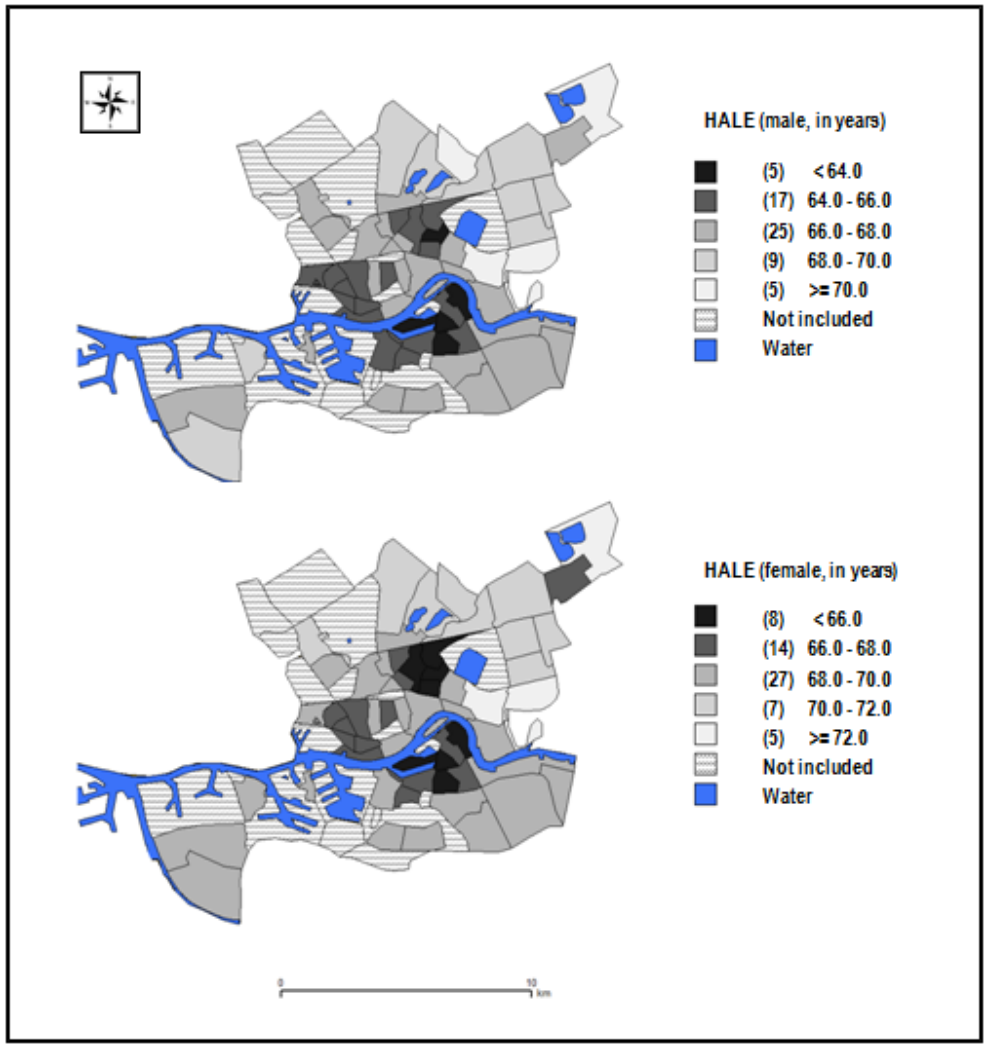


Table 5.4 Health-adjusted life expectancy (HALE) estimates for those 10 neighborhoods in Rotterdam, The Netherlands, that have the lowest estimated HALE in the 2004–2006 period, combined with the (posterior) probability that each neighborhood belongs to the top-5 of neighborhoods with the lowest HALE.

Sex	Neighborhood	Estimated HALE	Std.Dev	Rank	Prob(worst-5)
Male	Afrikaanderwijk	63.3	0.99	1	0.79
	Bloemhof	63.9	0.88	2	0.55
	Katendrecht	64.0	1.18	3	0.51
	Oud-Crooswijk	64.1	0.94	4	0.46
	Feijenoord	64.1	1.04	5	0.44
	Tussendijken	64.3	1.00	6	0.35
	Tarwewijk	64.5	0.91	7	0.25
	Schiemond	64.5	1.13	8	0.30
	Carnisse	64.5	1.00	9	0.27
	Oude Noorden	64.6	0.81	10	0.19
Female	Oud-Crooswijk	64.9	1.13	1	0.81
	Afrikaanderwijk	65.2	1.12	2	0.71
	Oude Noorden	65.6	0.94	3	0.57
	Bloemhof	65.9	0.96	4	0.42
	Feijenoord	66.0	1.13	5	0.40
	Nieuw-Crooswijk	66.1	1.32	6	0.36
	Rubroek	66.1	1.02	7	0.31
	Katendrecht	66.2	1.26	8	0.34
	Tussendijken	67.5	1.13	9	0.22
	Liskwartier	67.5	1.13	10	0.21

full uncertainty of the HALE estimates into account and provides a more reliable means of comparing areas with the worst health performance. As can be seen, the exact ranking of neighborhoods is uncertain and the five worst performing areas cannot be exactly identified. In fact, looking at simple rankings can be misleading; when the standard deviation of the neighborhoods' posterior HALE estimates are different, the estimated probabilities for neighborhoods with higher HALE estimates can be equally high or even higher than that for neighborhoods with lower HALE estimates. In other words, not only the point estimates but also the relative size of the standard deviations is important when comparing HALEs of different neighborhoods and the estimated (posterior) probabilities incorporate both of these aspects into a single measure.

Figure 5.3 Life expectancy versus health-adjusted life expectancy, by gender, for neighborhoods in Rotterdam, The Netherlands, 2004–2006.

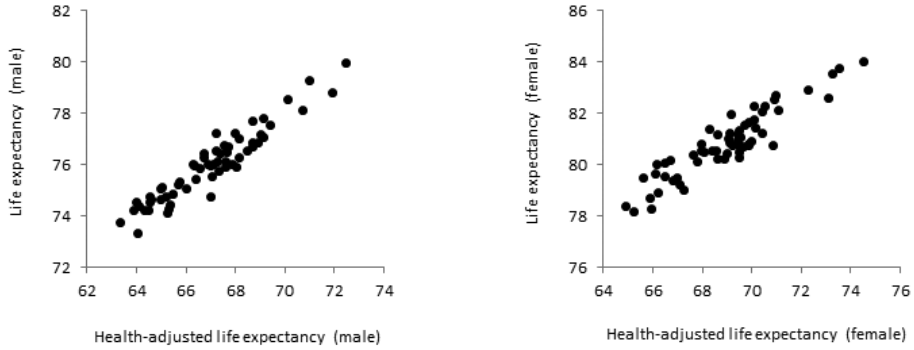


Table 5.5 contains the main results of the Monte Carlo simulation. The estimated bias of the HALE estimates is close to zero (-0.003) for males and slightly positive (0.067) for females. Compared to the average value and the range of male and female HALE (see Table 5.3), the bias is negligible. Interestingly, the coverage of the 95 credibility intervals is somewhat higher for females (94.2%) than for males (93.1%). This means that even though the reported HALE are slightly less accurate, the accompanying standard errors and 95% credibility intervals are slightly more accurate for females than for males.

To be able to explain the observed differences in bias between males and females, Table 5.6 contains a more detailed overview of the estimated bias for the age-specific mortality and the age-specific health-adjustment rates (which are the two inputs for the HALE calculations). As can be seen, the bias for the age-specific mortality rates is almost identical for males and females and essentially equal to zero for all age groups. The bias for the health-adjustment rates are slightly larger and also very similar for males and females. The only real difference is the larger bias of the health-adjustment rate for females in the final age group (i.e. $\geq$85 years). Apparently, women of age 85 years and older have a lower quality of life than would be predicted based on the observed male and female mortality rates and health-adjustment rates for younger age-groups. This pattern is also present in the descriptive statistics (see Table 5.2) and given the small sample size for this particular age group—only 6 observations in the entire dataset—the observed bias can be easily explained.

Table 5.5 Bias and coverage of the 95% credibility intervals, by sex, of the estimated health-adjusted life expectancies*

Measure	Sex	
	Male	Female
Bias	-0.003	0.067
Coverage	93.1	94.2

** Results are based on 2,000 simulation iterations*

5.4 DISCUSSION

The presented HALE modeling approach makes use of a Bayesian random-effects methodology that pools strength over age-groups, sexes, and geographical areas in order to calculate reliable life expectancies for areas as small as 2000 person years at risk.[72] The required health-status data for the HALE estimations are efficiently modeled using a beta-distribution and by combining information from the mortality and morbidity components of the model, the proposed methodology allows for the estimation of HALE at a much smaller level—and with substantially smaller survey sample sizes—than a traditional, fixed effects life table approach.

The presented results for neighborhoods in the city of Rotterdam confirm the feasibility of the calculations at the small area level and reveal major health inequalities between areas. The observed differences in small-area HALE are smaller than those observed at the international level (i.e. 8 years difference in HALE in Rotterdam compared to 14 years for the OECD[104] and up to 40 years globally[102] but roughly comparable to those observed at the regional level in the USA[98] and Canada.[100] The small-area results also confirm the general finding that differences between areas are more pronounced with HALE than regular LE. Additionally, a particularly attractive feature of the small-area example is that neighborhoods with the lowest HALE can be identified while taking the full uncertainty of the estimates and life table calculations into account. Such information can be valuable for governments and local health-authorities, who increasingly rely on small area-based approaches at reducing health inequalities but thus far have no reliable summary measure of (small-area) population health at their disposal.

Table 5.6 Bias of the age-specific mortality and age-specific health-adjustment rates, by sex*

Age-group	Mortality rate		Health-adjustment rate	
	Male	Female	Male	Female
1 (<1)	0.000	0.000	n/a	n/a
2 (1–4)	0.000	0.000	n/a	n/a
3 (5–9)	0.000	0.000	n/a	n/a
4 (10–14)	0.000	0.000	n/a	n/a
5 (15–19)	0.000	0.000	-0.002	-0.003
6 (20–24)	0.000	0.000	-0.001	0.000
7 (25–29)	0.000	0.000	0.000	0.000
8 (30–34)	0.000	0.000	0.000	-0.001
9 (35–39)	0.000	0.000	0.001	0.003
10 (40–44)	0.000	0.000	0.001	-0.001
11 (45–49)	0.000	0.000	0.001	0.001
12 (50–54)	0.000	0.000	0.001	0.000
13 (55–59)	0.000	0.000	0.001	0.000
14 (60–64)	0.000	0.000	-0.001	-0.001
15 (65–69)	0.000	0.000	0.000	-0.001
16 (70–74)	0.000	0.000	0.000	-0.002
17 (75–79)	0.000	0.000	-0.002	0.001
18 (80–84)	0.000	0.000	0.001	-0.002
19 (≥85)	0.000	0.000	0.003	0.030

** Results are based on 2,000 simulation iterations*

The presented Monte Carlo simulation evidence confirms that the results for Rotterdam are almost unbiased and that the coverage of the 95% credibility intervals is close to 95%. Accordingly, the uncertainty of the HALE estimates is accurately reflected in the reported standard errors and credibility intervals. In fact, the Monte Carlo results are very similar to those reported by Jonker et al.,[72] which further supports the validity of the small-area life table approach based on Bayesian random-effects estimations.

It should be noted that the main benefit of the proposed methodology is that it allows for an attractive and reliable summary measure of population health to be calculated at a (much) smaller geographic level than has thus far has been possible. HALE is indeed attractive: it is sensitive to the severity of morbidity within the population, provides balanced attention to fatal as well as non-fatal health outcomes, and can

be readily compared between areas with very different population age structures. Especially at the small-area level, where local differences in age structure are not averaged out as much, this is a clear advantage. Moreover, the HALE estimates can be easily corrected for the location of nursing homes and the interpretation of the results is relatively straight-forward, which makes HALE well-suited to communicate to a broad audience.

An important issue in the calculation of HALE, both in small and larger area analyses, is the reliance on an external set of health state preference weights to derive average health state values. One issue is that these value weights are treated as fixed coefficients in the HALE estimations, even though they are, in fact, estimates derived from a benchmark study themselves. Consequently, the variability of the HALE estimates is slightly under-predicted if the weights are treated as fixed coefficients.[105] A more important issue, however, is that health state value sets vary significantly between countries. In the presented example for Rotterdam, a UK value set for the SF-6D is used because there is no Dutch SF-6D health state tariff yet. This can have a major impact on the estimated HALE; evidence from Heijink et al.[97] shows that estimated HALE at the national level can differ between 2% and more than 20% depending on the value set used in the calculations. Admittedly, this evidence is based on the EQ-5D, which is an alternative health state instrument to the SF-6D, but it does emphasize the importance of using a health state value set that is as relevant as possible to the areas under investigation.

Finally, even though the proposed methodology already provides a comprehensive and coherent approach to estimating HALE at the small-area level, several improvements and extensions can still be envisaged. First of all, the required historical previous address information for the nursing home correction may not always be available. If so, an attractive alternative is to model the impact of nursing home deaths directly in the life table calculations, which allows for a correction of the age-specific mortality rates of the age groups that are directly affected (i.e. those of 65 years and older). As shown by Jonker et al.,[73] such an alternative correction can provide a close approximation of the ideal previous address correction, with the advantage that it only relies on the aggregate percentage of nursing home deaths in each area. The latter is considerably less detailed and consequently easier to obtain than historical address information for the entire nursing home population.

Secondly, the GGD health status survey that was used to obtain SF-12 health-state measurements was conducted in the non-institutional population. Accordingly, the

SF-12 health-state measurements that constitute the HALE calculations are restricted to the healthier non-nursing home population and this is likely to result in an upwards bias in the health-state measurements and in a reduction in the difference between the estimated LE's and HALE's. Alas, it is not possible to correct for this bias without additional data. Based on the results of Jonker et al.,[73] however, we know that the nursing home excluded measurements only affect the level of the life table measurements without invalidating comparisons between areas. And with additional SF-6D health-state measurements within the nursing home population, nursing home corrected HALEs could be calculated using the same approach as described by Jonker et al.[73]

Thirdly, for public health researchers and health authorities it would be interesting to monitor HALE over time. This can, of course, be accomplished by applying the proposed methodology to several independent cross-sectional datasets. However, it is also relatively straight-forward to extend the proposed modeling approach with a time dimension. This has the advantage that it allows for a slightly adapted version of Sullivan's methodology that is also robust to situations in which age-specific health-state prevalences change significantly over time.[68] Additionally, it allows for a pooling of strength between adjacent years in addition to the correlations between genders, age groups, areas, and mortality/morbidity that are included in the current specification. The latter not only improves the reliability of the estimations, but also allows for explicit and appropriate inferences about measurements of compression vs. expansion of morbidity, the existence of a-typical time trends, and the effects of interventions at the small area level.

Fourthly, after having estimated HALE at the small area level, it would also be interesting to further investigate the spatial variation in HALE in terms of associations with area socioeconomic characteristics and direct risk factors such as noise and air pollution. This can be done using follow-up regression that take the precision of the estimated HALE as well as the spatial configuration of the neighborhoods into account, as suggested by Arcaya et al.[86] Moreover, latent concepts that are important in the explanation of spatial variation in HALE, such as neighborhood deprivation and neighborhood social capital, can be summarized and included in the regressions using Bayesian factor analysis; see e.g. Congdon[91] and Mari-Dell'Olmo et al.[101]

In conclusion, the proposed methodology is versatile and provides a coherent framework for estimating HALE in small-area analyses. Compared to traditional, fixed effects life table estimations that can only be used for populations of at least 5000

person years at risk, the proposed Bayesian life table approach can produce accurate life table estimates for geographical areas with populations as small as 2000 person years at risk. Similarly, by pooling strength over geographical areas, sexes, age groups, and similarities between mortality and morbidity, significantly smaller sample sizes for the required health state surveys are required. This can reduce costs for future HALE surveys and makes HALE calculations using existing surveys increasingly more feasible.

6

The effect of urban green on small-area (healthy) life expectancy

with F.J. van Lenthe, B. Donkers, J.P. Mackenbach and A. Burdorf.
Journal of Epidemiology and Community Health (2014)

ABSTRACT

Background
Several epidemiological studies have investigated the effect of the quantity of green space on health outcomes such as self-rated health, morbidity and mortality ratios. These studies have consistently found positive associations between the quantity of green and health. However, the impact of other aspects, such as the perceived quality and average distance to public green, and the effect of urban green on population health are still largely unknown.

Methods
Linear regression models were used to investigate the impact of three different measures of urban green on small-area life expectancy (LE) and healthy life expectancy (HLE) in The Netherlands. All regressions corrected for average neighborhood household income, accommodated spatial autocorrelation, and took measurement uncertainty of LE, HLE as well as the quality of urban green into account.

Results
Both the quantity and the perceived quality of urban green are modestly related to small-area LE and HLE: an increase of 1 standard deviation in the percentage of urban green space is associated with a 0.1-year higher LE, and, in the case of quality of green, with an approximately 0.3-year higher LE and HLE. The average distance to the nearest public green is unrelated to population health.

Conclusions
The quantity and particularly quality of urban green are positively associated with small-area LE and HLE. This concurs with a growing body of evidence that urban green reduces stress, stimulates physical activity, improves the microclimate and reduces ambient air pollution. Accordingly, urban green development deserves a more prominent place in urban regeneration and neighborhood renewal programmes.

6.1 BACKGROUND

Urban regeneration and neighborhood renewal programmes often comprise urban green space developments. These are primarily aimed at creating more attractive neighborhoods, but increasingly also at affecting and improving population health. The latter is based on a growing number of studies that show that urban green spaces are positively associated with physical activity,[113] mental health and well-being,[114,115] self-reported health,[116–118] longevity of senior citizens,[119] and (all-cause) mortality.[118,120] The established evidence about the effect of green on population health, however, is still scarce and would benefit from further research using other health outcomes.[121] Particularly life expectancy (LE) measures are attractive in this respect.[24,72,73,103,122,123] Based on recently developed methodology to reliably estimate LE and healthy life expectancy (HLE) at the small-area level, we investigate the effect of urban green on small-area LE and HLE.

6.2 DATA AND METHODS

Population health data

Standard 5-year abridged life table data for the estimation of male and female LE and HLE for neighborhoods in all 22 metropolitan agglomerations in the Netherlands in the 2006–2009 period were obtained from Statistics Netherlands. Together, the included metropolitan agglomerations comprise roughly 40% of the Dutch population and cover all major urban regions in the Netherlands (Figure 6.1). Male and female LE and HLE at birth were subsequently calculated for all neighborhoods within the 22 agglomerations that met the minimum required population size of 1750 person-years at risk and minimum survey sample size of 10 respondents per sex using the small-area methodology as developed and validated by Jonker et al.[72] and Jonker et al. (Chapter 3; submitted). To avoid confounding from the migration of frail elderly to nursing homes prior to their death, the LE and HLE estimates were corrected for the location of nursing homes based on previous residential address information.[73]

Green space data

Three different urban green space measures were included in the analyses. The first is the percentage of green space per neighborhood, which was calculated using SAGA GIS based on the Dutch Land Use Database (BBG) for the year 2008. Our definition of green space included all types of green (excluding horticulture) and the percentage

of green was calculated using grid cells of 25m × 25m. Similar to Maas et al.[115,116] gardens and small-scale green spaces, such as road-side trees and grassy verges were therefore not classified as urban green. The second green measure is the average distance in kilometers, calculated over the road network, from all addresses in the neighborhood to the nearest public green (eg, a park, public garden or forest). This measure was obtained directly from Statistics Netherlands for the year 2008.[124] The third green measure is a subjective measure of the quality of the urban green in each neighborhood. Responses from two consecutive surveys (ie, Woon 2006 and Woon 2009) were combined to acquire a sufficiently large sample size. In both surveys, respondents were asked to evaluate the quality of urban green spaces in their own neighbourhood on a 5-point Likert scale. To be able to take the measurement uncertainty into account, the survey information was summarised by (1) the total number of respondents per neighborhood and (2) the number of respondents who were 'satisfied' or 'very satisfied' with the quality of green in their neighborhood, as opposed to being 'neutral', 'dissatisfied' or 'very dissatisfied'.

Neighbourhood income data

The average household income per neighborhood is a potentially important confounding factor in the relationship between urban green spaces and small-area population health. Hence, based on standardised disposable household income data as obtained from Statistics Netherlands, all neighborhoods in the 22 Dutch metropolitan agglomerations with a minimum population size of 1,750 person-years at risk were divided into average disposable household income quintiles. In the statistical analyses, the first quintile (i.e., the one with the lowest average household income) was used as the reference category.

Statistical analyses

The impact of the measures of urban green space on small-area LE and HLE was evaluated using linear regression models. To increase the efficiency of the estimations and alleviate problems with heteroskedasticity, the implemented specifications accommodate spatially autocorrelated errors as well as measurement uncertainty of the (healthy) life expectancy estimates. In the specifications that included the quality of urban green, the statistical uncertainty of the percentage of respondents who were 'satisfied' or 'very satisfied' with the quality of green in their neighborhood was also taken into account (using a binomial specification).

Figure 6.1 Geographic location of the 22 Dutch metropolitan agglomerations.

The models were programmed in the BUGS language and fitted in WinBUGS using Bayesian MCMC methods. This involves selecting prior densities for the unknown model parameters and updating those densities via the likelihood of the observed data. Appendix A contains the model code and the specification of the prior distributions. Proper priors were used that were much more diffuse than the posterior distributions. All estimations used 25,000 MCMC iterations to let the chains converge and 25,000 MCMC iterations with a thinning of 5 to reliably approximate the posterior distributions. Convergence was evaluated using the diagnostics implemented in the R CODA package.[125]

6.3 RESULTS

Table 6.1 contains the descriptive statistics of the included dependent and independent variables for the 1,190 included neighborhoods. As can be seen, there is substantial variation in the LE, HLE and urban green space measurements. There is also significant variation in the precision of the LE, HLE and the perceived quality of green space measurements, which was, as mentioned, taken into account in the regressions. Furthermore, the number of neighborhoods is smaller for the highest income quintiles. Neighbourhoods with less than 10 survey respondents per sex are predominantly affluent, which is explained by the smaller number of properties per square km, smaller population and consequently smaller survey sample sizes.

Table 6.2 contains the estimation results for the different model specifications, with each specification containing one of the three urban green space measures. All estimates are corrected for average neighborhood income, which has a strong impact on small-area population health. The differences between the lowest and highest income quintiles are 2.5–3 and 7.5–8.5 years for LE and HLE, respectively. Turning to the impact of urban green, the distance to public green is unrelated to small-area LE and HLE. The percentage of green space is unrelated to HLE but statistically significantly related to LE whereas the quality of the green space is significantly associated with LE and HLE. Overall, an increase of 1 standard deviation in the percentage of urban green is associated with a 0.10–0.14-year higher LE and for the quality of urban green with a 0.28–0.29-year higher LE and 0.26–0.33-year higher HLE.

6.4 DISCUSSION

Using three different measures of urban green, recently developed methodology to reliably estimate LE and HLE at the small-area level, and relatively sophisticated regression models, this short report has investigated the impact of urban green on small-area population health in the Netherlands. Having corrected for differences in average household income, we find evidence that the quantity and quality of urban green space are modestly related to small-area LE and HLE: an increase of 1 standard deviation in the percentage of urban green is associated with a 0.1-year higher LE, and, in the case of the perceived quality of green, with an approximately 0.3-year higher LE and HLE.

Table 6.1 Descriptive Statistics

Dimension	Variable	N	Mean	Std.Dev	Min	Max
Population health	Life expectancy at birth (male)	1,190	78.1	2.04	68.8	85.6
	Life expectancy at birth (female)	1,190	82.2	2.06	72.3	91.5
	Healthy life expectancy at birth (male)	1,190	65.1	5.25	46.1	81.2
	Healthy life expectancy at birth (female)	1,190	64.2	5.64	42.2	83.0
Measurement uncertainty	Life expectancy at birth (male—SD)	1,190	1.28	0.35	0.53	5.39
	Life expectancy at birth (female—SD)	1,190	1.37	0.45	0.55	8.12
	Healthy life expectancy at birth (male—SD)	1,190	3.33	1.06	1.38	8.56
	Healthy life expectancy at birth (female—SD)	1,190	3.78	1.12	1.47	9.20
	Total number of survey respondents for the perceived quality of green	1,190	46.8	58.4	5	646
Green	Percentage of green surface per neighbourhood	1,190	11.4	11.5	0.00	72.3
	Distance to nearest public green (in km)	1,178	0.45	0.27	0.10	2.60
	Percentage of survey respondents who consider the quality of the green spaces in their neighborhood as 'good'	1,190	74.8	16.6	12.1	100
Average Income*	Average standardised disposable household income (quintile 1)	270	17,035	1,148	11,194	18,451
	Average standardised disposable household income (quintile 2)	272	19,524	577	18,460	20,477
	Average standardised disposable household income (quintile 3)	278	21,588	642	20,483	22,766
	Average standardised disposable household income (quintile 4)	223	24,044	744	22,769	25,630
	Average standardised disposable household income (quintile 5)	147	29,743	5,948	25,652	75,940

** Note: the income quintiles are included as binary indicator (i.e. dummy) variables in the regressions*

Table 6.2 The effect of urban green on small-area LE and HLE, The Netherlands, 2006-2009*

	LE						HLE					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
Male												
Percentage Green**	0.14	(0.06—0.22)	n/a		n/a		0.01	(-0.19—0.21)	n/a		n/a	
Proximity to Public Green**	n/a		-0.04	(-0.12—0.05)	n/a		n/a		-0.12	(-0.32—0.08)	n/a	
Perceived Quality of Green**	n/a		n/a		0.28	(0.19—0.37)	n/a		n/a		0.33	(0.10—0.55)
Income_q1***	0		0		0		0		0		0	
Income_q2	0.83	(0.61—1.05)	0.82	(0.59—1.05)	0.75	(0.53—0.98)	2.55	(1.96—3.13)	2.53	(1.94—3.12)	2.47	(1.88—3.06)
Income_q3	1.66	(1.42—1.90)	1.64	(1.40—1.88)	1.6	(1.36—1.84)	4.53	(3.90—5.15)	4.51	(3.88—5.14)	4.48	(3.85—5.10)
Income_q4	2.38	(2.12—2.64)	2.38	(2.11—2.64)	2.27	(2.01—2.53)	5.98	(5.31—6.64)	5.95	(5.28—6.62)	5.83	(5.16—6.50)
Income_q5	3.12	(2.82—3.42)	3.13	(2.82—3.43)	2.97	(2.67—3.27)	7.71	(6.95—8.46)	7.69	(6.93—8.44)	7.5	(6.74—8.26)
R^2	0.67	(0.66—0.68)	0.68	(0.67—0.68)	0.67	(0.66—0.68)	0.69	(0.68—0.69)	0.69	(0.68—0.69)	0.69	(0.68—0.69)
Female												
Percentage Green**	0.1	(0.01—0.19)	n/a		n/a		-0.05	(-0.38—0.08)	n/a		n/a	
Proximity to Public Green**	n/a		-0.04	(-0.14—0.05)	n/a		n/a		-0.07	(-0.32—0.17)	n/a	
Perceived Quality of Green**	n/a		n/a		0.29	(0.19—0.39)	n/a		n/a		0.26	(0.00—0.53)
Income_q1***	0		0		0		0		0		0	
Income_q2	0.8	(0.55—1.04)	0.8	(0.56—1.05)	0.74	(0.50—0.98)	2.84	(2.16—3.50)	2.82	(2.13—3.50)	2.77	(2.09—3.46)
Income_q3	1.53	(1.27—1.79)	1.51	(1.25—1.77)	1.48	(1.22—1.74)	5.11	(4.39—5.82)	5.1	(4.37—5.83)	5.08	(4.34—5.80)
Income_q4	2.13	(1.84—2.40)	2.13	(1.85—2.41)	2.02	(1.74—2.30)	6.72	(5.95—7.50)	6.69	(5.91—7.47)	6.6	(5.81—7.39)
Income_q5	2.61	(2.29—2.94)	2.61	(2.28—2.94)	2.44	(2.12—2.77)	8.37	(7.47—9.26)	8.35	(7.45—9.24)	8.2	(7.29—9.10)
R^2	0.65	(0.63—0.65)	0.65	(0.63—0.65)	0.65	(0.63—0.65)	0.64	(0.64—0.65)	0.64	(0.63—0.64)	0.64	(0.64—0.65)

** 95% credibility intervals in parentheses ** Standardized coefficients, reflecting the impact of 1 standard deviation change (in years) *** The first income quintile is taken as the reference category*

The presented estimates are corrected for the impact of average household income. In regressions without such correction, the percentage of green and distance to public green estimates remain similar whereas the impact of quality of urban green approximately doubles. The latter implies that the perceived quality of green space is higher in more affluent neighborhoods (which may also reflect differences in accessibility and safety[121]) and suggests that the impact of the percentage of green space on LE is independent of neighborhood income. The degree of urbanity was not included as a potential confounder; as mentioned, only metropolitan neighborhoods were included in the analyses.

We are cautious to interpret our findings as causal effects. There are likely other confounders and selection effects that are insufficiently captured by the age standardisation and nursing home correction of the outcome measures and the inclusion of the income quintiles and spatial error term in the regression models. For example, the neighborhood microclimate (ie, temperature and humidity), air pollution and degree of social cohesion were not explicitly controlled for. However, these variables reflect some of the most important pathways from green to urban health.[126–128] The cross-sectional setting of the analysis, which implicitly assumes that current neighborhood exposures are indicative of cumulative exposures, is also a methodological limitation. Risk factor exposures may have changed over time and the effect of geographical mobility might not have been adequately taken into account. On the other hand, the implemented nursing home correction explicitly corrects for the selective migration to nursing home addresses and there is no clear evidence that migration between Dutch neighborhoods enlarged inequalities in health outcomes.[129]

The analysis also has several strengths. First, it is based on unique outcome measures (LE and HLE), which are directly age standardised and corrected for the selective migration of frail elderly to nursing homes. Furthermore, our analyses are based on a multiple city approach (ie, neighborhoods from all 22 Dutch metropolitan agglomerations are included), on robust statistical methodology and on three different measures of green, including the perceived quality of urban green. Additionally, the results are robust to changes in model specification, insensitive to the choice of priors and potential interactions between the quantity and quality of urban green are statistically insignificant.

Conclusion

Both the quantity and quality of urban green are positively associated with neighborhood (H)LE. This concurs with a growing body of evidence that urban green reduces

stress, stimulates physical activity, improves the microclimate and reduces ambient air pollution. Accordingly, we believe that urban green development deserves a more prominent place in urban planning, with the results indicating that not only the quantity but also the quality of urban green should be considered in future interventions.

What is already known on this subject

Contact with green can have substantial health benefits. Whereas the available evidence is relatively strong at the individual level, much less is known about the effect of the proximity, quantity and quality of green spaces on population health.

What this study adds

Based on recently developed methodology to estimate small-area life expectancy (LE) and healthy life expectancy (HLE), this paper evaluates the impact of three different measures of urban green on small-area population health. In contrast to previous epidemiological studies, our statistical models account for spatially correlated unobserved determinants, measurement uncertainty and the migration of frail elderly to nursing homes. Furthermore, the included measures capture the quantity of green as well as the perceived quality and average distance to public green. Particularly the perceived quality of urban green appears to be correlated with small-area population health, which suggests that not only the quantity but also the quality of urban green should be considered in future interventions.

Appendix A. OpenBUGS code

```
model {
 for (i in 1:N) {
  # LE and HLE are assumed normally distributed with mean mu and precission tau
  Y[i] ~ dnorm( mu[i],tau[i] )

  # mu is predicted with CMA-specific intercept, income-quintile, measure of
  green and spatial error term V
  #mu[i] <- alpha[CMA[i]] + beta[INC[i]] + beta[6]*GR_perc[i] + V[i]
  #mu[i] <- alpha[CMA[i]] + beta[INC[i]] + beta[6]*GR_dist[i]  + V[i]
  mu[i] <- alpha[CMA[i]] + beta[INC[i]] + beta[6]*GR_qual[i] + V[i]
  # tau takes the measurement uncertainty of the LE/HLE into account
  (Y_variance is loaded as data)
  tau[i] <- 1/ (Y_var[i] + sigma)
  # the uncertainty of the quality of green measurement is taken into
  account via a binomial specification
  GR_good[i] ~ dbin( GR_qual[i], GR_count[i] )}

 # priors
 for (a in 1:22){ alpha[a] ~ dnorm(mu_alpha,tau_alpha) }
 for (b in 2:6){ beta[b] ~ dnorm(0,0.001) }
 for (i in 1:N){ GR_qual[i] ~ dbeta(1,1) }
 beta[1] <- 0 # reference category
 mu_alpha ~ dnorm(0,0.001)
 tau_alpha ~ dgamma(1,0.001)
 gamma ~ dnorm(0,0.001)
 sigma ~ dgamma(1,0.001)

 # spatial error terms are assigned a (proper) CAR prior
 # num[], adj[], and adjIndex[] are loaded as data
 V[1:N] ~ car.proper( zeros[], C[], adj[], num[], M[], phi, rho)
 phi ~ dgamma(1,0.001)
 rho ~ dbeta(1,1)
 for (i in 1:N){
  zeros[i] <- 0
  M[i] <- 1/num[i]
  for (j in adjIndex[i]:adjIndex[i+1]-1){
   C[j] <- 1/num[i]}}

 # monitor R-squared
 Rsq <- 1 - (1/mean(prec[])) / pow(sd(Y[]),2);
 # Monitor standardized coefficients
 #green_std <- beta[6] * sd( GR_perc[] )
 #green_std <- beta[6] * sd( GR_dist [] )
 green_std <- beta[6] * sd( GR_qual[] )}
```

7

Estimating the impact of health-related behaviors on geographic variation in cardiovascular mortality

with B. Donkers, B. Chaix, F.J. van Lenthe, A. Burdorf and J.P. Mackenbach.
Epidemiology (2015)

ABSTRACT

Background
Incidence of and mortality from cardiovascular disease (CVD) exhibit a strong geographical pattern, with inhabitants of more affluent neighborhoods showing a substantially lower risk of CVD mortality than inhabitants of deprived neighborhoods. Thus far, there is insufficient evidence to which extent these differences can be attributed to differences in health-related behaviors.

Methods
Using a Hierarchical Related Regression approach, we combined individual and aggregate (ecological) data to investigate the extent to which small-area variation in CVD mortality in Dutch neighborhoods can be explained by several behavioral risk factors, i.e. smoking, drinking, overweight, and physical inactivity. The proposed approach combines the benefits of both an ecological analysis (in terms of data availability and statistical power) and an individual-level analysis (in terms of identification of the parameters and interpretation of the results).

Results
After correcting for differences in age and sex, accounting for differences in the behavioral risk factors reduces income-related inequalities in CVD mortality by approximately 30%.

Conclusion
Direct targeting of the excess prevalence of unhealthy behaviors in deprived neighborhoods is identified as a relevant strategy to reduce inequalities in CVD mortality. Our results also show that the proposed Hierarchical Related Regression approach provides a powerful method for the investigation of small-area variation in health outcomes.

7.1 INTRODUCTION

Some researchers have suggested that geographic variations in CVD mortality exist only because of variation in the numbers of disadvantaged individuals living in disadvantaged areas, and that individuals with similar social characteristics have similar risks regardless of their place of residence. Others have insisted that there are substantial area effects that are common to all individuals in an area even after correcting for relevant individual-level exposures.[130] In reviews of the available evidence, area-effects are usually found to be modest[131,132] and particularly challenging to measure using observational data.[10,133] From a policy perspective it may thus be more interesting to investigate how much geographic variation in incidence and mortality from CVD is explained by major behavioral risk factors, such as smoking, drinking, and physical activity. As mentioned by Morris et al.,[134] if behavioral risk factors explain much of the geographic variation in CVD mortality, research efforts should focus on the reasons why people in certain areas continue to adopt and maintain unhealthy behavioral patterns. Also, interventions aimed at tackling disparities between affluent and deprived neighborhoods could focus on these unhealthy behaviors, for example as an integral part of wider area-based initiatives to reduce the socioeconomic disadvantage of specifically targeted neighborhoods.[1]

Thus far, evaluations of the impact of behavioral risk factors on variation in all-cause or CVD mortality have been based on individual-level survey data and/or cohort studies.[135–137] Even though these studies are based on large sample sizes, the number of respondents available for each specific neighborhood is very small. Aggregate (or ecological) studies are based on exhaustive incidence information and often provide a more feasible alternative in terms of data availability, but suffer from well-known limitations in terms of interpreting associations found at the aggregate level as measures of individual effects.[138–141] In this paper, we therefore use a recently developed class of multilevel models, called Hierarchical Related Regressions, in which aggregate data and individual data are efficiently combined.[20,21] This allows for an investigation of the contribution of behavioral risk factors to geographic variation in CVD mortality that combines the benefits of both an ecological analysis (in terms of data availability and statistical power) and an individual-level analysis (in terms of identification of the parameters and interpretation of the results).

7.2 METHODS

The implemented modelling framework is based on the Hierarchical Related Regression methodology as developed by Jackson et al.[20,21] and incorporates individual-level as well as area-level outcomes within a single, coherent framework. By modelling individual and aggregate outcomes simultaneously, this methodology alleviates ecological bias caused by the model mis-specification and lack of identification in standard ecological analyses[21] and increases statistical power relative to standard multilevel models based on individual-level data.

The Hierarchical Related Regression framework consists of two related models; an individual-level and an aggregate-level model. The individual-level model is a standard multilevel model in which the observed individual outcome is explained in terms of individual and area-level predictors. The principle or "trick" of the Hierarchical Related Regression approach is then to specify the area-level model as the exact equivalent of the individual-level model aggregated over all (unobserved) individuals in each area. This way, both models contribute to the full likelihood, which is simply the product of the likelihoods of the individual-level and area-level model.

Individual-level model

In the individual-level model, the observed individual CVD mortality outcome y_{ij} (area i, individual j) is modelled in terms of a set of individual-level predictors X_{ij} and a set of area-level predictors Z_i via a logistic regression, with the individual risk p_{ij} defined as a logit-linear function of the included covariates:

$$y_{ij} \sim \text{Bernoulli}(p_{ij}) \tag{7.1}$$

$$\text{logit}(p_{ij}) = \mu + \beta X_{ij} + \gamma Z_i + u_i \tag{7.2}$$

Here μ denotes the intercept, β and γ regression parameters, and u_i an area-level random effect. Together, 7.1 and 7.2 describe a standard multilevel logistic regression that accounts for the correlation in the outcome among individuals who live in the same area.[130]

Area-level model

Because the available individual-level data alone provide insufficient power to obtain statistically significant parameter estimates, the individual-level model is combined

with an area-level model based on ecological data. In the area-level model, the observed outcome is the total number of CVD deaths (Y_i) in each area i, which is assumed binomial distributed:

$$Y_i \sim \text{binomial}(N_i, P_i) \tag{7.3}$$

with N_i denoting the population at risk and P_i denoting the average risk of CVD mortality in area i. Unlike in standard ecological studies, where P_i is explained directly in terms of area-level averaged covariates, the individual-level model as described by equations 7.1 and 7.2 is in the Hierarchical Related Regression approach averaged over all (unobserved) individuals within each area to obtain the theoretically correct ecological specification. The average risk P_i in equation 7.3 is thereby defined as the integral of the individual's conditional outcome probability $p_{ij}(x)$ over the included covariates with joint within-area distribution $\phi_i(x)$ in area i:

$$P_i = \int p_{ij}(X, Z_i)\ \phi_i(X) dX \tag{7.4}$$

With continuous explanatory variables this integral is difficult (and often impossible) to estimate from ecological data. However, the covariates in our analysis are all binary. The latter implies that ϕ_i in equation 7.4 represents the fraction of the population in each of the $k = 1, \ldots, K$ possible combinations of the included covariates and that equation 7.4 reduces to the weighted average risk as determined over the K possible combinations of the covariates:

$$P_i = \sum_{k=1}^{K} p_{ik} \times \phi_{ik} \tag{7.5}$$

As an example, suppose that there are only three covariates included in the Hierarchical Related Regression specification, i.e. smoking, drinking, and obesity. With these three covariates, there are $K = 2^3 = 8$ possible combinations of the risk factors (see table 7.1) and given the fraction of the population in each riskfactor combination (ϕ_k) and associated risk of CVD mortality for individuals within each riskfactor combination (p_k), the average risk of CVD mortality for the hypothetical area as a whole can be calculated using equation 7.5 and equals 0.05. With additional covariates included, the number of categories K increases exponentially but the principle application remains the same.

Table 7.1 Example of the calculation of the average risk of CVD mortality for an area with 3 covariates included in the Hierarchical Related Regression specification

Smoking	Drinking	Obesity	k	p_k*	ϕ_k**	Average risk $\sum p_k \times \phi_k$
0	0	0	1	0.55	0.02	
0	0	1	2	0.09	0.04	
0	1	0	3	0.16	0.06	
0	1	1	4	0.02	0.10	
1	0	0	5	0.09	0.08	
1	0	1	6	0.01	0.12	
1	1	0	7	0.07	0.14	
1	1	1	8	0.01	0.18	
						0.05

* p_k = *probability of CVD mortality in risk factor combination k*
** ϕ_k = *fraction of the area's population in risk factor combination k*

To complete the model specification, both ϕ_{ik} and p_{ik} in equation 7.5 need to be specified. Starting with p_{ik}, the risk of CVD mortality is determined using the exact same specification as in the individual-level model, i.e. on the log-odds scale and with identical coefficients:

$$\text{logit}(p_{ik}) = \mu + \beta X_{ik} + \gamma Z_i + u_i \tag{7.6}$$

Because ϕ_{ik} refers to behavioral risk factors, it is not directly available and needs to be estimated from survey data. For this, the count of the survey respondents in each of the 1 to K categories within each area i (i.e. $R_{i_{[1:K]}}$) is assumed multinomial distributed:

$$R_{i_{[1:K]}} \sim \text{multinomial}(\phi_{i_{[1:K]}}) \tag{7.7}$$

This provides a straight-forward way to estimate ϕ from area-specific behavioral survey data although the survey data may need to be weighted if a stratified sampling scheme was used. Regardless, the number of survey respondents per area is usually limited whereas the number of unique combinations of the covariates increases exponentially with the number of included covariates (i.e. with 2^n). Consequently, only a small number of behavioral covariates can be included in a Hierarchical Related Regression analysis.

Stratification by sex and age

In the presented analysis, the within-area classification of inhabitants and the within-area CVD mortality outcome are both available by sex and age. This implies that the area-level model can be stratified in the same dimensions. The latter is beneficial because it removes 2 covariates from ϕ and ϕ needs to be estimated from survey data. It also reduces uncertainty in the estimations; hence the area-level model is implemented as follows:

$$Y_{six} \sim \text{binomial}(N_{six}, P_{six}) \tag{7.8}$$

$$P_{six} = \sum_{k=1}^{K} p_{sixk} \times \phi_{sixk} \tag{7.9}$$

$$\text{logit}(p_{sixk}) = \mu + \beta X_{sixk} + \gamma Z_i + u_i \tag{7.10}$$

$$R_{six_{[1:K]}} \sim \text{multinomial}(\phi_{six_{[1:K]}}) \tag{7.11}$$

with Y_{six}, N_{six}, and P_{six} denoting the number of CVD deaths, population at risk, and average risk of CVD mortality by sex (s), area (i), and age group (x), respectively.

Mediating effect of the behavioral risk factors

Two separate Hierarchical Related Regression models are estimated: one with and one without behavioral risk factors, but with Z_i in both models containing a set of dummy variables that indicate the average household income quintile of each neighborhood. Assuming correctly specified models,[142, 143] an indication of the mediating effect of the behavioral risk factors on income-related inequalities in CVD mortality is obtained by comparing the parameters γ and γ' in the two Hierarchical Related Regression models for the lowest neighborhood income quintiles (both relative to the highest income quintiles). This provides an initial estimate of the mediating effect of the behavioral risk factors. A formal mediation test (with higher statistical power) is also conducted. This test is the Bayesian equivalent of the bootstrap test as proposed by Preacher and Hayes,[144, 145] albeit slightly modified to take the binary structure of the behavioral risk factor variables into account. Details about the implementation and limitations of the mediation tests are included in Appendix D.

Comparison with standard individual and ecological models

The results of the Hierarchical Related Regression models are also compared with results obtained from a standard individual-level and a standard ecological model. This comparison is intended to illustrate the usefulness of the HRR approach, which has higher statistical power than a standard individual-level model and, in constrast to standard ecological models, provides unbiased estimates of the indvidual-level effects. Note that the individual-level model is described by equations 7.1 and 7.2 and that the ecological model is defined by equation 7.3 with P_i explained in terms of area-level averaged covariates:

$$\text{logit}(P_i) = \mu + \gamma Z_i + u_i \tag{7.12}$$

Bayesian estimation and priors

Bayesian methods are used for all parameter estimations, which involves selecting prior densities for the unknown model parameters and updating those densities via the likelihood of the observed data. Apart from providing a convenient estimation framework, Bayesian methods also allow for the use of (weakly) informative priors for the purpose of regularization in sparse-data settings, which is particularly useful for the estimation of ϕ in Hierarchical Related Regression applications.

Compared to Jackson et al.,[20,21] slightly less informative priors are assigned to the β, γ and μ parameters: Normal(0,0.725) priors to β and γ, which represent the 95% prior believe that the true odds ratios are between 1/10 and 10, and improper flat priors to μ. To allow for spatial autocorrelation in the error term, the random area effect (u_i) is assigned a Leroux et al.[146] conditionally autoregressive prior, which is essentially a mixture of an unstructured and spatially structured error term with mixture parameter ρ – the latter being assigned an uniform(0,1) prior.

With a sufficiently large survey sample size $\phi_{six_{[1:K]}}$ in equation 7.11 could be estimated directly from the available survey data. However, to alleviate sparse data problems and to be able to increase the number of covariates in the analysis, a more informative prior is used instead. This informative prior is derived from additional survey data and based on the most important determinants of the (between-strata) variation in risk factor profiles. More specifically, the prior expectation of ϕ is calculated as:

$$\phi\text{-prior}_{sixk} = \sum_{e=1}^{E} \sum_{a=1}^{A} \phi^*_{sxeak} \times \pi_{sixea} \tag{7.13}$$

Table 7.2 Example of the calculation of ϕ-prior for an area with equal population shares π_e and 3 covariates in the HRR specification

				ϕ^* (City average)*			Individual Area	
Smoking	Drinking	Obesity	k	Low Education	Medium Education	High Education	π_e^{**}	ϕ-prior
0	0	0	1	0.38	0.56	0.62	0.33	0.52
0	0	1	2	0.12	0.07	0.05	0.33	0.08
0	1	0	3	0.13	0.17	0.18	0.33	0.16
0	1	1	4	0.04	0.03	0.03		0.03
1	0	0	5	0.14	0.08	0.06		0.09
1	0	1	6	0.03	0.02	0.01		0.02
1	1	0	7	0.09	0.05	0.04		0.06
1	1	1	8	0.07	0.03	0.01		0.04

* ϕ_k^* *= fraction of the city's population in risk factor combination k, stratified by education*
** π_e *= fraction of the individual area's population in each of the education strata*

with π denoting the strata-specific fractions of inhabitants per education (e) and ethnicity category (a) and with ϕ^* denoting the joint distribution of the included behavioral risk factors stratified by sex (s), age group (x), education (e), and ethnicity (a).

As an example, suppose that ϕ consists of $K = 2^3 = 8$ different combinations of the included behavioral risk factors and that ϕ^* is only stratified by education (low/med/high). Then, the prior expectation is formed as a weighted average of ϕ^* using the area's fraction of the population in each education group as weights (see table 7.2). Because this prior becomes increasingly more reliable with additional stratification, ϕ^* in equation 7.13 is additionally stratified by sex, age, and ethnicity. Also, note that ϕ-prior in equation 7.13 is only used as a prior and will be further refined using the area-specific survey information (via equation 7.11). Accordingly, areas with identical sex/age/education/ethnicity-profiles can still have divergent estimates of ϕ, with the degree of smoothing towards the prior expectation depending on the number of survey respondents in each area.

Based on the data availability in our application, π in equation 7.13 can be determined with high precision: ethnicity is directly observed from the population registry (i.e. based on exhaustive data) and the required education levels are derived using large-sample (non-survey) data from the national education registry. In contrast, ϕ^* in equation 7.13 cannot be directly estimated from the available survey data because they are too sparse. Rather than using a simpler stratification, which would result

in less reliable and potentially biased estimates, the estimation of ϕ^* is therefore performed using a Chow-Liu[147] approximation. This approach provides an efficient approximation using marginal and second-order conditional risk factor prevalences and is only slightly less accurate than a direct estimation of ϕ^*. Further detail about the implemented approximation and the estimation of ϕ-prior is included in Appendix A.

The Hierarchical Related Regression models are programmed in the BUGS language and OpenBUGS is used to estimate the parameters using iterative Markov chain Monte Carlo (MCMC) sampling techniques. Each estimation starts with 25,000 burn-in MCMC iterations to allow three separate Markov chains to converge, followed by 75,000 MCMC iterations to reliably approximate the posterior distributions. To speed-up the calculations, avoid numerical problems, and accommodate the stratified random sampling design that was used in one of the surveys, the standard multinomial distribution in OpenBUGS was replaced by a modified version that accommodates fractional counts and two user-written functions were added via the OpenBUGS Development Interface.[148]

Data

The data for the Hierarchical Related Regression implementation are obtained from two sources: from a large-sample anonymous health survey conducted in the 4 largest municipalities in the Netherlands by the local health authorities (GGD) in 2008, and from (exhaustive) registry data combined with small sample nationally representative health surveys (POLS) made available by Statistics Netherlands. In both sources, the behavioral survey data are derived from identical survey questions and, in both sources, the geographic unit of analysis is the Dutch neighborhood. The latter is a convenient spatial unit because neighborhoods are specifically designed to comprise relatively homogeneous areas in terms of type of housing, land-use, and socioeconomic status of their inhabitants.[149] In 2008, neighborhoods in the four largest Dutch cities comprised on average 5,400 inhabitants.

A full description of the data is provided in Appendix B. In total, 199 neighborhoods from the 4 largest Dutch municipalities are included in the area-level model and 1411 individuals from 3 consecutive POLS surveys in the individual-level model. To maximize the sample size, individuals from the 10 largest Dutch municipalities are included in the individual-level model; both models need not comprise the exact same neighborhoods (cf. Jackson et al.[21]). Also note that the GGD health survey is anonymous, which implies that individual respondents cannot be linked to individual

health outcomes and that the GGD survey thus cannot be used for the individual-level HRR model.

Based on the required coherence between the GGD and POLS survey questions and the limited number of covariates that can be included in a HRR analysis, the following behavioral risk factors are included: 1) no alcohol consumption 2) excessive alcohol consumption, 3) smoking, 4) obesity, and 5) physical inactivity. Table 7.3 contains the definition of the behavioral risk factors. As shown, excessive alcohol consumption is a combination of excessive drinking and binge drinking. This aggregation is used to reduce the number of behavioral risk factors in the Hierarchical Related Regression analysis.

Finally, three age groups are specified (i.e. 35-54, 55-74, and 75 years and older). Younger ages are excluded because CVD mortality is exceptionally rare below the age of 35 and more narrowly defined age groups are avoided because the latter would substantially reduce the available survey sample size per stratum and further increase the number of covariates in the Hierarchical Related Regression analysis.

Table 7.3 Definitions of the behavioral risk factors

1.	Alcohol abstinence	No alcohol consumption in the past 12 months
2a.*	Binge drinking	More than 6 standard alcoholic drinks per day—at least once per week in the past 6 months
2b.*	Excessive alcohol	On average 21 or more standard alcoholic drinks per week (14 or more for females)
3.	Smoking	Is a current smoker (regardless of smoking history)
4.	Obesity	Body mass index (BMI) based on self-reported height and weight equal to or greater than 30
5.	Physical inactivity	Does not meet the Dutch Standard for Healthy Exercise, i.e. 30 minutes of moderately intensive physical activity for at least 5 days a week[150]

** Binge drinking and excessive alcohol consumption are combined into a single indicator of 'excessive drinking'*

7.3 RESULTS

Table 7.4 provides the summary statistics of the POLS and GGD health survey datasets. As can be seen in the table, the survey data appear largely consistent, which is to be expected because all selections pertain to metropolitan populations and are based on the same survey questions. Small differences in the number of alcohol abstainers, obese, and physically inactive can be explained by differences in the ethnic composition of the selected populations (see Appendix B). Inhabitants with a non-western ethnicity, particularly from Muslim countries such as Morocco and Turkey, are more likely to abstain from alcohol consumption (men) and are more often obese and physically inactive (women), and the percentage of the population with a non-western ethnic background is higher in the four largest cities in the Netherlands than in other Dutch municipalities (see e.g. Kuipers et al.[151]).

Table 7.4 Summary of the behavioral survey data used in the HRR model*

		GGD** 2008	POLS*** 2005-2011	POLS*** 2006-2008
Sex	Used as/for: Riskfactor	Estimating ϕ (in equation 7.11)	Estimating ϕ^* (in equation 7.13)	individual-level data (in equation 7.1)
		N=5200	N=6797	N=685
Male	Is abstainer	0.20 (0.15)	0.12 (0.24)	0.15 (0.27)
	Is excessive	0.22 (0.11)	0.18 (0.28)	0.19 (0.29)
	Is smoker	0.30 (0.12)	0.31 (0.33)	0.32 (0.35)
	Is obese	0.15 (0.08)	0.12 (0.22)	0.13 (0.24)
	Is inactive	0.38 (0.12)	0.37 (0.36)	0.39 (0.35)
		N=6294	N=7265	N=726
Female	Is abstainer	0.29 (0.16)	0.24 (0.31)	0.25 (0.31)
	Is excessive	0.14 (0.09)	0.13 (0.25)	0.14 (0.24)
	Is smoker	0.24 (0.10)	0.26 (0.30)	0.25 (0.30)
	Is obese	0.19 (0.09)	0.14 (0.24)	0.15 (0.27)
	Is inactive	0.42 (0.13)	0.36 (0.35)	0.39 (0.37)

** Neighborhood averages (by sex) for respondents aged 35 and older (with standard deviations in parenthesis). ** Based on stratified random sampling design with over-weighting of non-Western ethnicities. *** Based on nationally representative samples of the Dutch population.*

Table 7.5 contains the Hierarchical Related Regression regression results for two specifications: one with and one without the behavioral risk factors included in the model specification. Starting with the first, the base-line risk (μ) is small, indicating that male and females in the age-group 35-54 that live in a neighborhood in the highest income quintile have a probability of CVD mortality that is 0.002 (for males) and 0.001 (for females). Males thus have a slightly higher risk than females, yet age is by far the most important determinant of CVD mortality. Neighborhood income also has an important effect on the risk of CVD mortality. There is a strong gradient in the risk of CVD mortality, with males and females in neighborhoods in the lowest income quintile showing a 66% and 34% higher risk of CVD mortality relative to those in the highest income quintile, respectively.

Turning to the specification with behavioral risk factors, the base-line risk has become considerably smaller, which reflects that the reference category now additionally experiences the most favorable behavioral risk-factor profile (in addition to belonging to age group 35-54 and living in the most affluent neighborhoods). Being male and particularly being part of an older age group increases the risk of CVD mortality. Most interesting in this specification, however, are the effects of the behavioral risk factors. Smoking has the strongest effect on CVD mortality; current smokers have a threefold higher probability of CVD mortality than non-smokers. Being obese or physically inactive results in an approximately twofold higher probability of CVD mortality. Alcohol abstinence and excessive drinking have a much smaller effect and increase the probability of CVD mortality with 17% (credible interval: -18% – 58%) and 22% (credible interval: -43% – 113%), respectively. Finally, with the inclusion of the behavioral risk factors the effect of neighborhood income on CVD mortality is reduced, particularly for the lowest income quintiles. This reduces income-related inequality in CVD mortality by approximately 30%. The latter is calculated as the (average) percentage difference in the parameter estimates of the lowest income quintile before and after inclusion of the behavioral risk factors and is further evaluated by the mediation test as described in Appendix D.

Table 7.6 and Table 7.7 contain the standard individual multilevel and standard ecological results, which allows for a direct comparison with the Hierarchical Related Regression results presented in Table 7.5. In the individual-level models, the sample size of 1,411 survey respondents (divided over approximately 300 neighborhoods) is too small to obtain reliable estimates for the degree of spatial autocorrelation in the data. Hence unstructured instead of spatially structured random effects are used. Regardless, the sample size is still too small to obtain meaningful results, which

Table 7.5 Hierarchical Related Regression (HRR) results*

		HRR specification #1		HRR specification #2	
		Male	Female	Male	Female
Neighborhood-level estimates	Income_q1	1.66 (1.50—1.84)	1.34 (1.21—1.47)	1.48 (1.32—1.66)	1.21 (1.09—1.35)
	Income_q2	1.53 (1.38—1.69)	1.24 (1.13—1.36)	1.43 (1.28—1.59)	1.15 (1.04—1.27)
	Income_q3	1.33 (1.20—1.48)	1.15 (1.04—1.27)	1.28 (1.14—1.42)	1.08 (0.97—1.21)
	Income_q4	1.14 (1.02—1.28)	1.08 (0.97—1.20)	1.08 (0.96—1.21)	1.03 (0.92—1.15)
	Income_q5**	1.00 n/a	1.00 n/a	1.00 n/a	1.00 n/a
Individual-level estimates	Baseline risk (μ)	0.0023 (0.0021—0.0026)	0.0013 (0.0012—0.0015)	0.0009 (0.0006—0.0012)	0.0005 (0.0004—0.0007)
	Age group 1 (35-54)**	1.00 n/a	1.00 n/a	1.00 n/a	1.00 n/a
	Age group 2 (55-74)	7.74 (7.18—8.34)	8.51 (7.64—9.46)	8.64 (7.84—9.55)	9.15 (8.09—10.3)
	Age group 3 (75+)	46.4 (43.0—49.9)	85.4 (77.2—94.5)	61.9 (54.4—71.6)	101 (87.5—118)
	Is Abstainer***	n/a		1.17 (0.82—1.58)	
	Is excessivedrinker	n/a		1.22 (0.57—2.13)	
	Is smoker	n/a		3.17 (2.19—4.52)	
	Is obese	n/a		2.07 (1.38—2.98)	
	Is physicallyinactive	n/a		1.91 (1.26—2.89)	
Spatial autocorrelation	rho (ρ)	0.73 (0.29—0.98)		0.74 (0.29—0.99)	
	precison (tau_u)	159 (64.6—363)		175 (68.2—408)	

** Mean posterior estimates on the odds-ratio scale with 95% credible intervals in parenthesis*
*** The fifth income quintile (i.e. the highest income group) and the first age group are taken as reference categories*
**** Estimates not differentiated by sex*

Table 7.6 Standard individual-level regression results*

		Individual-level specification #1		Individual-level specification #2	
		Male	Female	Male	Female
Neighborhood-level estimates	Income_q1	1.58 (0.30—4.87)	2.60 (0.58—7.58)	1.33 (0.25—4.66)	2.19 (0.47—6.47)
	Income_q2	1.92 (0.40—5.72)	1.77 (0.40—5.14)	1.87 (0.38—5.59)	1.65 (0.36—4.83)
	Income_q3	0.92 (0.13—3.09)	1.12 (0.20—3.56)	1.00 (0.13—3.43)	1.05 (0.18—3.32)
	Income_q4	1.39 (0.23—4.48)	2.28 (0.47—6.84)	1.50 (0.24—4.84)	2.57 (0.51—7.80)
	Income_q5**	1.00 n/a	1.00 n/a	1.00 n/a	1.00 n/a
Individual-level estimates	Baseline risk (μ)	0.0070 (0.0014—0.0200)	0.0063 (0.0013—0.0182)	0.0032 (0.0005—0.0100)	0.0030 (0.0005—0.0094)
	Age group 1 (35-54)**	1.00 n/a	1.00 n/a	1.00 n/a	1.00 n/a
	Age group 2 (55-74)	3.43 (0.86—9.54)	2.24 (0.55—6.15)	3.84(0.85—9.75)	2.38 (0.58—6.59)
	Age group 3 (75+)	6.38 (1.30—19.3)	13.0 (3.66—34.7)	8.03 (1.58—24.3)	13.8 (3.78—37.0)
	Is Abstainer***	n/a		1.56 (0.59—3.27)	
	Is excessivedrinker	n/a		1.00 (0.13—3.43)	
	Is smoker	n/a		2.37 (1.13—5.51)	
	Is obese	n/a		1.73 (1.53—3.94)	
	Is physicallyinactive	n/a		1.91 (0.81—3.90)	
Spatial autocorrelation	rho (ρ)	n/a		n/a	
	precison (tau_u)	n/a		n/a	

** Mean posterior estimates on the odds-ratio scale with 95% credible intervals in parenthesis*
*** The fifth income quintile (i.e. the highest income group) and the first age group are taken as reference categories*
**** Estimates not differentiated by sex*

Table 7.7 Standard ecological regression results*

		Ecological specification #1		Ecological specification #2	
		Male	Female	Male	Female
Neighborhood-level estimates	Income_q1	1.77 (1.58—1.97)	1.44 (1.30—1.60)	1.70 (1.50—1.92)	1.40 (1.24—1.57)
	Income_q2	1.59 (1.42—1.77)	1.33 (1.20—1.47)	1.55 (1.39—1.73)	1.30 (1.17—1.44)
	Income_q3	1.35 (1.20—1.51)	1.20 (1.08—1.34)	1.33 (1.18—1.49)	1.18 (1.06—1.32)
	Income_q4	1.33 (1.08—1.37)	1.12 (1.00—1.25)	1.20 (1.06—1.36)	1.12 (1.00—1.25)
	Income_q5**	1.00 n/a	1.00 n/a	1.00 n/a	1.00 n/a
Individual-level estimates	Baseline risk (μ)	0.0046 (0.0036—0.0057)	0.0048 (0.0037—0.0061)	0.0047 (0.0037—0.0059)	0.0048 (0.0037—0.0061)
	Age group 1 (35-54)**	1.00 n/a	1.00 n/a	1.00 n/a	1.00 n/a
	Age group 2 (55-74)	6.86 (3.39—12.4)	5.25 (2.63—9.74)	6.03 (2.94—11.0)	5.25 (2.55—9.67)
	Age group 3 (75+)	424 (217—755)	158 (97.1—218)	457 (230—818)	158 (102—234)
	Is Abstainer***	n/a		1.14 (0.93—1.38)	
	Is excessivedrinker	n/a		1.09 (0.87—1.34)	
	Is smoker	n/a		1.03 (0.83—1.25)	
	Is obese	n/a		1.19 (0.95—1.47)	
	Is physicallyinactive	n/a		0.88 (0.74—1.03)	
Spatial autocorrelation	rho (ρ)	0.58 (0.18—0.95)		0.60 (0.29—0.98)	
	precison (tau_u)	64.0 (34.0—116)		68.0 (35.0—131)	

** Mean posterior estimates on the odds-ratio scale with 95% credible intervals in parenthesis*
*** The fifth income quintile (i.e. the highest income group) and the first age group are taken as reference categories*
**** Estimates not differentiated by sex*

clearly indicates a lack of statistical power. Apart from the base-line risk, only the estimates for the oldest age groups and the effect of smoking have 95% credible intervals that lie strictly above 1. The standard ecological models perform much better in this respect. In the ecological specification without behavioral risk factors, all parameters have 95% credible intervals larger than 1 (except for the baseline risk, which is larger than 0). The estimates do deviate from those obtained using the Hierarchical Related Regression specification, which is indicative of the degree of ecological bias in the standard ecological specification. Particularly in the specification that includes the behavioral risk factors, the difference with the HRR specifications is substantial: none of the behavioral risk factors is found to have a meaningful effect on the probability of CVD mortality. This is not the result of problems with collinearity: the correlations between the risk factor exposures are somewhat higher at the ecological than at the individual level but not large enough to pose a problem (see Table 7.10 in Appendix B). Instead, the absence of meaningful estimates is the result of ecological bias that results from an inability of the ecological model to take the joint distribution of the risk factors into account, as is theoretically required to obtain unbiased estimates.

7.4 DISCUSSION

Using a Hierarchical Related Regression approach, this paper has investigated the extent to which income-related differences in CVD mortality in Dutch neighborhoods can be explained by several behavioral risk factors, i.e. excessive drinking, alcohol abstinence, smoking, obesity, and physical inactivity. Having corrected for differences in the sex and age structure of the populations, accounting for these five behavioral risk factors is shown to reduce income-related differences in CVD mortality between Dutch neighborhoods by approximately 30%.

The presented Hierarchical Related Regression approach combines the benefits of both an individual-level analysis (in terms of identification and interpretation of the parameter estimates) and an ecological analysis (in terms of data availability and statistical power). But the Hierarchical Related Regression models are also informative about the spatial structure of the outcome measure. For example, sex and age-group corrected probabilities of CVD mortality can easily be monitored for the included neighborhoods. Furthermore, stabilized behavioral risk-factor profiles are estimated as part of the model, which allow for the identification of neighborhoods with particularly unfavorable risk-factor profiles. Also, hypothetical changes to the area-specific risk-factor profiles (e.g. reflecting the result of specific interventions) can be evaluated,

both in terms of impact on the spatial distribution of CVD mortality as in terms of impact on the income-related differences in CVD mortality. Depending on the specific application at hand, these possibilities may provide additional reasons for using the proposed Hierarchical Related Regression approach.

Turning to the practical implications of the presented results, an important implication is that the direct targeting of excess prevalence of unhealthy behaviors in deprived neighborhoods appears to be a relevant strategy to reduce income-related inequalities in CVD mortality. Particularly smoking, physical activity, and overweight are interesting candidates for interventions, as they have the largest impact on CVD mortality and exhibit substantial prevalence differences across the income quintiles (see Table 7.9 in Appendix B). Additionally, given the geographic clustering of unhealthy risk factor behavior (see table 7.10 in Appendix B), area-based targeting of deprived neighborhoods seems to be a relevant and potentially effective approach to address income-related inequalities in CVD mortality.[152] The possibility of the Hierarchical Related Regression approach to identify specific neighborhoods and evaluate potential interventions before they are actually implemented would nicely complement such area-based strategies.

On the other hand, interventions aimed at reducing behavioral risk factor exposure in deprived neighborhoods without addressing the underlying or root causes of inequalities could also be less effective than implied by the presented results. As an example, smoking and drinking are often part of coping mechanisms related to the stress and adversity associated with living in deprived neighborhoods.[153] As a result, addressing existing behavioral risk factors without changing the underlying socioeconomic determinants could trigger alternative adverse risk factors to take their place, mitigating the impact of potential interventions.

Strengths and weaknesses

Our study has some important limitations. First, because the Hierarchical Related Regression model depends on a large scale GGD survey that was conducted in 2008, the study has a follow-up period that is limited to approximately 6 years. This is short compared to similar analyses based on individual-level datasets.[135–137] Second, we could not include several important risk factors for CVD mortality, such as ethnicity, diets, cholesterol/triglyceride levels, diabetes, hypertension, and access to and use of health care. Some of these factors are likely correlated to the included behavioral risk factors yet omitted due to various reasons, ranging from data availability and inconsistencies between the survey questions (e.g. dietary information and blood lipid

levels) to concerns about survey measurements error (e.g. in the case of hypertension prevalences). Most importantly, however, were limitations in the available survey sample size that have severely restricted the total number of covariates in the Hierarchical Related Regression analysis.

Indeed, with larger sample sizes it would have been possible to include more and more detailed risk factors, such as smoking history, ethnicity, binge drinking, and an individual socioeconomic variable in addition to the neighborhood-specific income quintiles. On the other hand, in real-life applications the required survey data are often limited, implying that an application of the Hierarchical Related Regression approach with survey data requires a careful balance between sparse data problems on the one hand and including sufficient covariates to achieve a meaningful analysis on the other. In this paper, with the available survey sample size from the GGD and POLS health surveys, we have tried to strike a proper balance.

Strengths of the presented analysis comprise the innovative application of the Hierarchical Related Regression approach using Bayesian methods, the availability of large-scale health surveys with identical survey questions, and the availability of all other (i.e. non-survey) data from Dutch national registries. Together, these inputs constitute the first application of the Hierarchical Related Regression approach in which behavioral risk factors are included. Additionally, the presented analysis is the first application of the Hierarchical Related Regression approach in which more than a handful of explanatory variables were included, which highlights the usefulness of the carefully constructed informative priors that were used to accommodate the sparse data problems that are typically encountered in Hierarchical Related Regression analyses. In this respect, we hope that the presented methodology enables further substantive applications of the Hierarchical Related Regression approach.

Appendix A. Informative prior

The Chow-Liu (1986) approximation

To alleviate sparse data problems and to be able to increase the number of included covariates in the HRR analysis, an informative prior is used to reliably estimate ϕ in the HRR model. With π_{sixea} denoting the fraction of inhabitants per sex (s), area (i), age group (x), education level (e), and ethnicity (a), an initial expectation of ϕ is calculated as:

$$\phi\text{-prior}_{sixk} = \sum_{e=1}^{E} \sum_{a=1}^{A} \phi^*_{sxeak} \times \pi_{sixea} \tag{A.1}$$

where ϕ^* denotes the average joint distribution of the behavioral risk factors stratified in the same dimensions. Instead of estimating ϕ^* directly, it is efficiently approximated using a Chow-Liu[147] dependence tree. Following the Chow-Liu approach, the selection of the branches in the optimal dependence tree is based on the mutual information (MI) criteria for all combinations of the behavioral risk factors. The MI criteria are calculated as:

$$MI(x_1, x_2) = \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} p(x_1, x_2) \log \left(\frac{p(x_1, x_2)}{p(x_1) \times p(x_2)} \right) \tag{A.2}$$

where $p(x1, x2)$ is the joint probability distribution function of X_1 and X_2, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X_1 and X_2 respectively. The MI measure the information that each combination of the risk factors has in common, or put differently, how much knowing about one risk factor reduces the uncertainty about another. Based on the ranking of the MI criteria (as calculated from the joint distribution of the 5 included risk factors aggregated for the 4 largest Dutch cities, see Table 7.8) while taking the mutual exclusiveness of alcohol abstinence and excessive drinking into account, ϕ^* is approximated as:

$$\phi^* = p(x_1|x_3, x_2) \times p(x_3|x_2) \times p(x_4|x_1) \times p(x_5|x_1) \tag{A.3}$$

with the strata indices omitted for readability. This approximation uses 9 rather than 23 parameters and is well-suited to accommodate a pooling of strength approach that further mitigates sparse data problems.

The pooling of strength approach

The required input for the Chow-Liu approximation are reliably estimated using 'Intelligent Bayesian smoothing' via Gaussian Random Markov Fields.[154] The implemented approach essentially searches for correlations between the behavioral risk factors in adjacent age groups, and uses these correlations (if present) to stabilize the estimates of the conditional risk factors.

The pooling of strength approach is implemented as follows. First, based on the counts obtained from the survey data, each risk factor is assumed binomial distributed:

$$\text{nrAbstainer}_{csxae} \sim \text{binomial}(\text{prob\_r1}_{csxae}\ ,\ \text{surveyAbstainer}_{csxae}), \tag{A.4}$$

$$\text{nrExcessive}_{sxae} \sim \text{binomial}(\text{prob\_r2}_{sxae}\ ,\ \text{surveyExcessive}_{sxae}), \tag{A.5}$$

$$\text{nrSmoker}_{csxae} \sim \text{binomial}(\text{prob\_r3}_{csxae}\ ,\ \text{surveySmoker}_{csxae}), \tag{A.6}$$

$$\text{nrObese}_{csxae} \sim \text{binomial}(\text{prob\_r4}_{csxae}\ ,\ \text{surveyObese}_{csxae}), \tag{A.7}$$

$$\text{nrInactive}_{csxae} \sim \text{binomial}(\text{prob\_r5}_{csxae}\ ,\ \text{surveyInactive}_{csxae}) \tag{A.8}$$

with all behavioral risk factors, except for excessive drinking, conditioned on 1 additional behavioral risk factor as indicated by subscript c. Based on the ranking of the MI criteria, the probability of abstinence is determined separately for smokers and non-smokers, the probability of smoking separately for excessive drinkers and non-excessive drinkers, and the probability of obesity and physical inactivity separately for alcohol abstainers and alcohol drinkers. This provides the required input for the Chow-Liu approximation in equation A.3.

The risk factor probabilities are subsequently predicted on the logit scale using an intercept and an age-group specific term that takes correlations between behavioral risk factor exposures in adjacent age groups into account:

$$\text{logit}(\text{prob\_r1}_{csxae}) = b_{0\_}r1_{csae} + b_{1\_}r1_{csxae}, \tag{A.9}$$

$$\text{logit}(\text{prob\_r2}_{sxae}\ ) = b_{0\_}r2_{sae}\ + b_{1\_}r2_{sxae}, \tag{A.10}$$

$$\text{logit}(\text{prob\_r3}_{csxae}) = b_{0\_}\text{r3}_{csae} + b_{1\_}r3_{csxae}, \tag{A.11}$$

$$\text{logit}(\text{prob\_r4}_{csxae}) = b_{0\_}\text{r4}_{csae} + b_{1\_}r4_{csxae}, \tag{A.12}$$

$$\text{logit}(\text{prob\_r5}_{csxae}) = b_{0\_}\text{r5}_{csae} + b_{1\_}r5_{csxae}. \tag{A.13}$$

The b_0 terms are assigned weakly informative Normal(0,0.01) priors—as opposed to improper flat priors, which are avoided to increase numerical stability—and the

Table 7.8 *(Approximation of) the joint distribution of the included behavioral risk factors*

Abstainer	Excessive drinker	Smoker	Obese	Physically inactive	k	True joint distribution*	Chow-Liu approximation	Approximation assuming independence
0	0	0	0	0	1	0.261	0.252	0.227
0	0	0	0	1	2	0.142	0.144	0.152
0	0	0	1	0	3	0.030	0.032	0.036
0	0	0	1	1	4	0.022	0.018	0.024
0	0	1	0	0	5	0.075	0.081	0.079
0	0	1	0	1	6	0.049	0.046	0.053
0	0	1	1	0	7	0.008	0.010	0.012
0	0	1	1	1	8	0.006	0.006	0.008
0	1	0	0	0	9	0.061	0.057	0.067
0	1	0	0	1	10	0.027	0.033	0.045
0	1	0	1	0	11	0.008	0.007	0.011
0	1	0	1	1	12	0.006	0.004	0.007
0	1	1	0	0	13	0.042	0.041	0.023
0	1	1	0	1	14	0.023	0.023	0.016
0	1	1	1	0	15	0.004	0.005	0.004
0	1	1	1	1	16	0.003	0.003	0.002
1	0	0	0	0	17	0.070	0.073	0.090
1	0	0	0	1	18	0.074	0.080	0.060
1	0	0	1	0	19	0.019	0.020	0.014
1	0	0	1	1	20	0.023	0.022	0.009
1	0	1	0	0	21	0.020	0.016	0.031
1	0	1	0	1	22	0.021	0.018	0.021
1	0	1	1	0	23	0.004	0.004	0.005
1	0	1	1	1	24	0.005	0.005	0.003

** Based on aggregated GGD survey data of the 'Big-4' cities (i.e. Amsterdam, Rotterdam, The Hague, and Utrecht)*

b_1 terms are assigned an intrinsic conditional autoregressive (CAR) random-walk(1) prior. The latter is a standard distribution in the OpenBUGS software package and in this application specified with an adjacency matrix based on age-group contiguity and with Gamma(1,0.01) priors placed on the precision parameters. Finally, an additional age group (i.e. 20-34 years) is included in the estimations because the data are readily available and the risk factor exposures in this age group are likely informative about the risk factor exposures in the adjacent age group of 35-54 years.

From ϕ-prior to a prior distribution

Having estimated ϕ-prior, the final step is to translate the posterior distribution of ϕ-prior into an informative prior that can be used by OpenBUGS to estimate ϕ in the HRR model. A standard and often suitable approach would be to express ϕ-prior by a Dirichlet distribution. This distribution is used in the sensitivity analyses that were conducted (see Appendix E) but, given the large number of risk factor combinations in the full HRR model, a more flexible prior distribution is required for the full HRR models that were included in the main text.

Several generalized, mixed, or nested Dirichlet distributions have been developed that are more flexible and do not impose negative correlations between the risk factor exposures (see e.g. O'Hagan and Forster[155]). However, a distribution of scaled normal distributions is easier to implement and used instead. Starting with K-1 independent normal distributions per stratum (Z_{sixk}):

$$Z_{sixk} \sim \text{normal}(\mu_{sixk}, \tau_{sixk}) \qquad , \qquad k = 1, \ldots, K\text{-}1 \tag{A.14}$$

with mean μ and precision (i.e. inverse variance) parameters τ, the implied distribution on ϕ-prior is defined as:

$$\phi\text{-prior}_{sixk} = \frac{exp(Z_{sixk})}{1 + \sum_{k=1}^{K\text{-}1} exp(Z_{sixk})} \; , \quad k = 1, \ldots, K\text{-}1 \tag{A.15}$$

$$\phi\text{-prior}_{sixK} = \frac{1}{1 + \sum_{k=1}^{K\text{-}1} exp(Z_{sixk})} \, . \tag{A.16}$$

Note that A.14 – A.16 essentially describes a logistic-normal distribution (cf. Aitchison[156]) with independent rather than multivariate normal distributions. To derive the input parameters of the normal distributions a 2-step approximation approach was used. In the first step, the inverse of the transformation as described in equation A.15

was monitored during the estimating of ϕ-prior:

$$Z_{sixk} = \ln\left(\frac{\phi\text{-prior}_{sixk}}{1-\sum_{k=1}^{K\text{-}1}\phi\text{-prior}_{sixk}}\right) \quad , \quad k = 1,\ldots,K\text{-}1. \tag{A.17}$$

Subsequently, 9 strategically chosen quantiles (i.e. 2.5, 5, 10, 25, 50, 75, 90, 95, 97.5) of $Z_{six\ k}$ were used to estimate $\mu_{six\ k}$ and $\tau_{six\ k}$ in a separate regression, with weakly-informative Normal(0,0.01) and Gamma(1,0.01) priors placed on μ and τ, respectively. The estimated input parameters are subsequently loaded as data in the HRR estimations.

Appendix B. Data

HRR area-level data

The population at risk for neighborhoods in Amsterdam, Rotterdam, The Hague, and Utrecht is obtained from the national population registry (GBA). The selected date (1st of April, 2008) matches the field period of the 2008 GGD health survey as closely as possible and, because the health surveys exclude the institutionalized population, a similar selection is made in the GBA.

Information about sex, age, and ethnicity are directly included in the GBA. To obtain information about the education profile of the neighborhoods, the selected population at risk is linked to the national education database as maintained by Statistics Netherlands. Following standard convention, 3 education levels are distinguished: 'low', denoting primary or first-stage secondary education (SOI level 1-3), 'medium', denoting second-stage secondary education (SOI level 4), and 'high' denoting a tertiary education (SOI levels 5-7).

To obtain information about the average household income in each neighborhood, the selected population at risk is also linked to the national household income database. This database contains detailed income information based on social welfare transfers and income information from the tax authorities. The disposable household incomes are first standardized using equivalence factors as established by Statistics Netherlands (to take the economies of scale associated with household size and composition into account) before averages per neighborhood are calculated and each neighborhood is assigned into a quintile of the average neighborhood income distribution.

To obtain the CVD mortality outcome, the selected population at risk is linked to the national deaths registry. Cause-specific mortality data are available until the 1st of January 2014, resulting in a follow-up period of 5 years and 9 months to establish the mortality outcome. Mortality cases with primary ICD-10 codes in chapter IX (i.e. all codes starting with 'I') are considered CVD deaths whereas all other mortality is coded as 0. Note that CVD mortality is a relatively rare event due to the limited follow-up period: the observed mortality rates are on average 1.6% and 1.7% in the 2 highest income quintiles and 2.2% and 2.5% in the lowest income quintiles.

Except for the average household income data, all area-level variables are stratified by sex and age. In total, 3 age groups are specified (i.e. 35-54, 55-74, and 75 years

and older). As mentioned in the main text, younger age groups are excluded because CVD mortality is exceptionally rare below the age of 35 and more narrowly defined age groups are avoided because the latter would significantly reduce the available survey sample size per stratum.

Individual-level data

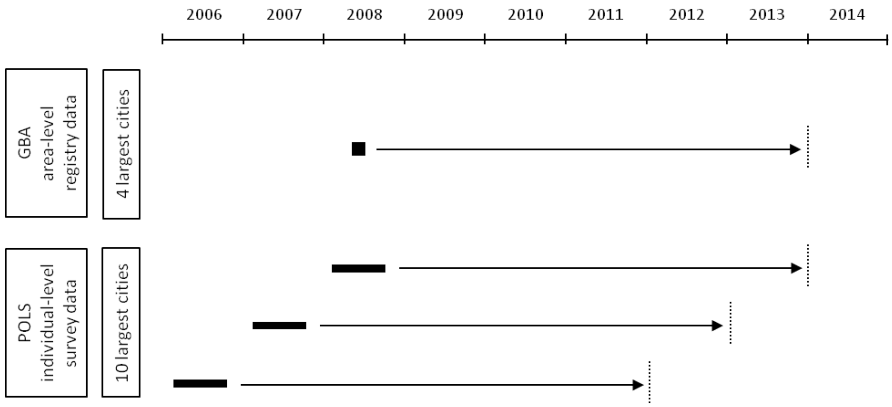
The individual-level data are obtained from 3 consecutive nationally representative health surveys (POLS) as conducted by Statistics Netherlands in the 2006-2008 period. The selected period is considered a reasonable trade-off between sample size and bias: several POLS surveys are combined to increase the number of respondents in the individual-level dataset, but older POLS surveys are avoided because the population composition, behavioral risk factor distribution, and average disposable income level of specific neighborhoods could have changed over time. Based on similar considerations, POLS survey respondents from the 10 largest Dutch cities are included in the individual-level dataset (instead of respondents only from the 4 largest cities). The latter increases the number of respondents from 1,139 to 1,411 (i.e. +24%) when taking a minimum of 4 respondents per neighborhood into account for areas not included in the HRR model.

For the 1,411 individuals included in the individual-level dataset, CVD mortality outcomes are obtained by linking the POLS surveys to the national deaths registry—taking identical follow-up periods as for the area-level data into account, see Figure 7.1. The respondents' sex, age, and neighborhood are obtained from the GBA and the following binary risk factors are obtained from the surveys: 1) no alcohol consumption, 2) excessive drinking (i.e. either binge drinking or excessive alcohol consumption), 3) smoking, 4) obesity, and 5) physical inactivity. Table 3 in the main text contains the definition of the behavioral risk factors.

Survey data used for the estimation of ϕ

Whereas the POLS surveys directly comprise the required behavioral risk factor exposure in the individual-level data, the GBA registry data does not contain any behavioral risk factor information. As explained in the main text, information about the latter is obtained from a large-sample anonymous health survey conducted in the 4 largest municipalities in the Netherlands in 2008 by the local health authorities (GGD).

Figure 7.1 *Base-line measurement and CVD mortality follow-up period* *,**



**Base-line measurement indicated by a black square (area-level data) or rectangle (individual-level data), indicating the measurement period. **The area-level data has a mortality follow-up period until January 1st of 2014 and the 3 consecutive POLS surveys until January 1st of 2012, 2013, and 2014, respectively.*

The GGD health survey contains (for many risk factors) the same survey questions as the POLS survey but has a much larger sample size with approximately 60 respondents per neighborhood. Unfortunately, the GGD survey is anonymous, which implies that individual-level health outcomes cannot be linked to the survey and that the GGD data cannot be used in the individual-level model. Nonetheless, the GGD survey is well-suited for the estimation of ϕ in the area-level model. With a minimum of 25 survey respondents in the included age groups (i.e. age 35 and older), a total of 199 neighborhoods and 11,494 survey respondents are included for the estimation of ϕ. As can be seen from Table 7.9, average risk factor prevalences vary considerably between sex and neighborhoods' income quintiles.

A complicating factor with the estimation of ϕ is the stratified random sampling design of the GGD survey, which involves stratification by sex, age, neighborhood, but also by ethnicity. Given that ethnicity is available from the Dutch population registry (GBA), a desirable solution would have been to further stratify the area-level model by ethnicity. Unfortunately, the GGDsurvey sample size is not large enough to accommodate additional stratification in the HRR analysis. Accordingly, survey weights are used instead to account for the higher probability of inclusion of non-Western ethnicities. These survey weights are calculated conform the guidelines of the Dutch health authorities[157] and based on the within-strata fractions of 4 ethnic

groups, i.e. Western, Turkish, Moroccan, and other non-Western ethnicities. Based on the GGD guidelines, the sum of the weighted respondents within each stratum equals the actual number of survey respondents; hence individual respondents are weighted to ensure a more representative survey sample but no additional weight is assigned to the survey information.

Table 7.9 *Average (unweighted) behavioral risk factor prevalences, by sex and income quintile, for GGD health survey respondents aged 35 and older*

Sex	Neighborhood income quintile	% Abstainer	% Excessive drinker	% Smoker	% Obese	% Physically inactive
Male	1 (lowest)	0.32	0.20	0.33	0.17	0.42
	2	0.19	0.24	0.32	0.15	0.37
	3	0.14	0.22	0.27	0.13	0.34
	4	0.13	0.19	0.22	0.13	0.35
	5 (highest)	0.10	0.21	0.23	0.11	0.41
Female	1 (lowest)	0.41	0.10	0.24	0.22	0.49
	2	0.31	0.13	0.24	0.20	0.40
	3	0.22	0.16	0.23	0.18	0.38
	4	0.18	0.17	0.21	0.15	0.40
	5 (highest)	0.16	0.21	0.17	0.08	0.35

Survey data used for the estimation of α in the mediation model

The GGD health survey data is also used to estimate the marginal effect of income on the prevalence of the behavioral risk factors, which is possible because the estimation of α in the mediation model (as described in Appendix D) does not require information about CVD mortality. To estimate α, the same geographical selection of 199 neighborhoods in the 4 largest Dutch cities is used but respondents for an additional age group (20-34 years old) are also included, resulting in a total of 15,270 rather than 11,526 survey respondents. As before, survey weights are calculated conform the guidelines of the Dutch health authorities[157] and based on the within-strata fractions of the 4 ethnic groups.

Survey data used for the estimation of the informative prior ϕ^*

The Chow-Liu approximation of ϕ^* is based on the marginal and second-order conditional probabilities of the behavioral risk factors that are stratified by sex, age, education and ethnicity (see Appendix A). This information is obtained from POLS health surveys aggregated over the period 2005-2011 and with a selection of respondents that comprises all 22 Dutch metropolitan agglomerations. Both the geographical area and the selected period are larger than used in the individual-level dataset, which is possible first because a follow-up period to establish CVD mortality is not required (hence the use of more recent survey data) and second, because the required data only pertain to marginal and second-order conditional probabilities in the larger population and not to individual neighborhoods. The larger geographical selection does rely on the assumption that sex, age, education, and ethnicity are the only important determinants of differences in risk-factor exposure; if other structural determinants exist they should either be included in the stratification of ϕ^* or, alternatively, the geographical selection should be shrunk to comprise a more similar selection as in the largest Dutch municipalities.

Standard ecological data

Standard ecological data (i.e. aggregate population and CVD mortality counts per neighborhood) are also extracted from the GBA, with the required neighborhood-specific prevalences of the risk factors obtained from the GGD health survey. Because only marginal prevalences are used in the standard ecological model (i.e. the fraction of the neighborhoods' populations that smoke, drink, etc.), the available sample size per neighborhood is large enough to calculate these fractions without making use of informative priors. As before, survey weights are used to ensure representative samples and only respondents of 35 years and older are included in the calculations.

As for the other covariates, the area-level income quintiles are the same as in the HRR model but the male and female population at risk and accompanying CVD mortality are aggregated over the 3 included age groups, with the fraction of males and females in the older age groups used to control for the population age structure in each area. As can be seen in Table 7.10, correlations between the behavioral risk factors at the ecological level are higher than those observed in the individual level data. This is not the result of the use of tetrachoric rather than standard (phi-) correlations, which intuitively suggests a geographical clustering of unhealthy behavior. The latter was confirmed by a simulation study in which individuals constituting the ecological

data were repeatedly shuffled and thereby randomly assigned to one of the included neighborhoods prior to calculating new ecological data and estimating correlation coefficients for further inference. Based on 2,000 draws, 8 out of 10 correlation parameters in the observed ecological data were outside of the constructed 95% confidence intervals.

Overview of the behavioral risk factor data

As explained, there are only 2 sources of behavioral risk factor data, but both sources are used in different capacities and in different estimations. Accordingly, to summarize the usage of the behavioral risk factor data, Figure 7.2 provides an overview of the behavioral survey data that includes the source, geographical selection, temporal selection, and usage of the data.

Table 7.10 *Correlations between behavioral risk factors at the individual and ecological level*

Individual level*		Is abstainer	Is excessive drinker	Is smoker	Is obese	Is physically inactive
	Is abstainer	1				
	Is excessive drinker	-1.00***	1			
	Is smoker	0.01	0.24	1		
	Is obese	0.19	-0.01	-0.10	1	
	Is physically inactive	0.19	-0.11	0.06	0.13	1
Ecological level**		% abstainers	% excessive drinkers	% smokers	% obese	% physically inactive
	% abstainers	1				
	% excessive	-0.4	1			
	% smokers	0.03	0.24	1		
	% obese	0.35	-0.19	0.11	1	
	% physically inactive	0.34	-0.26	-0.08	0.28	1

** Tetrachoric correlations (based on individual-level data)*
*** Pierson correlations (based on neighborhood-averaged weighted GGD survey data)*
**** Alcohol abstinence and excessive drinking are mutually exclusive in individual-level data*

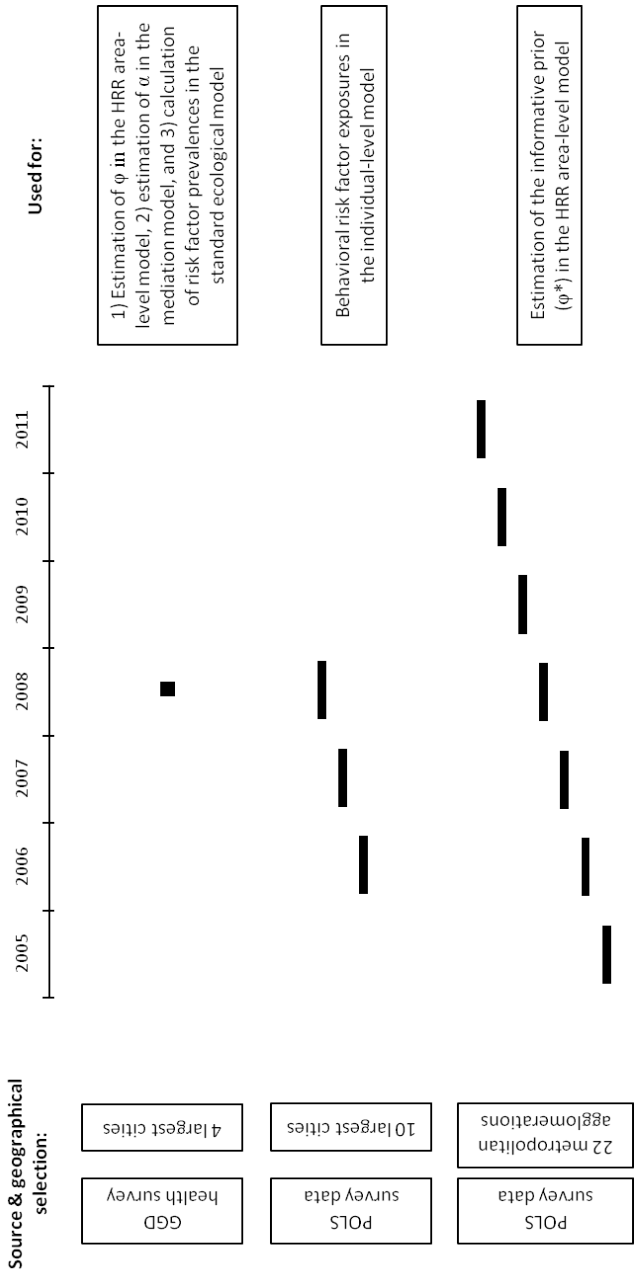


Figure 7.2 *Overview of the behavioral risk factor data (source, selection, and usage)*

Appendix C. OpenBUGS HRR model code

```
model{
 # PART A. Aggregate model
 for (s in 1:2) {
  for (i in 1:I_aggr) {
   for (x in 1:X) {
    cvd_deaths[s,i,x] ~ dbin( aggr_prob[s,i,x] , pop[s,i,x] )
    # aggr_prob[s,i,x] <-  inprod( phi[s,i,x,1:K] , q[s,i,x,1:K])
    # faster than OpenBUGS inprod() function:
    for (k in 1:K) { Prod_aggr[s,i,x,k] <-  phi[s,i,x,k] * q[s,i,x,k] }
     aggr_prob[s,i,x] <- sum( Prod_aggr[s,i,x,] )
     logit( q[s,i,x,1] ) <- mu[s] + alpha[s,x]                                                                  + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,2] ) <- mu[s] + alpha[s,x]                                                 + beta[5]  + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,3] ) <- mu[s] + alpha[s,x]                                    + beta[4]              + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,4] ) <- mu[s] + alpha[s,x]                                    + beta[4] + beta[5]    + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,5] ) <- mu[s] + alpha[s,x]                       + beta[3]                           + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,6] ) <- mu[s] + alpha[s,x]                       + beta[3]              + beta[5]    + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,7] ) <- mu[s] + alpha[s,x]                       + beta[3] + beta[4]                 + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,8] ) <- mu[s] + alpha[s,x]                       + beta[3] + beta[4] + beta[5]       + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,9] ) <- mu[s] + alpha[s,x]             + beta[2]                                     + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,10]) <- mu[s] + alpha[s,x]             + beta[2]                     + beta[5]       + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,11]) <- mu[s] + alpha[s,x]             + beta[2]          + beta[4]                  + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,12]) <- mu[s] + alpha[s,x]             + beta[2]          + beta[4] + beta[5]        + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,13]) <- mu[s] + alpha[s,x]             + beta[2] + beta[3]                           + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,14]) <- mu[s] + alpha[s,x]             + beta[2] + beta[3]          + beta[5]        + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,15]) <- mu[s] + alpha[s,x]             + beta[2] + beta[3] + beta[4]                 + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,16]) <- mu[s] + alpha[s,x]             + beta[2] + beta[3] + beta[4] + beta[5]       + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,17]) <- mu[s] + alpha[s,x] + beta[1]                                                  + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,18]) <- mu[s] + alpha[s,x] + beta[1]                                       + beta[5]  + theta[s, HHinc[i]] + u[i]
     logit( q[s,i,x,19]) <- mu[s] + alpha[s,x] + beta[1]                          + beta[4]              + theta[s, HHinc[i]] + u[i]
```

```
    logit( q[s,i,x,20]) <- mu[s] + alpha[s,x] + beta[1]                    + beta[4] + beta[5]  + theta[s, HHinc[i]] + u[i]
    logit( q[s,i,x,21]) <- mu[s] + alpha[s,x] + beta[1]          + beta[3]                      + theta[s, HHinc[i]] + u[i]
    logit( q[s,i,x,22]) <- mu[s] + alpha[s,x] + beta[1]          + beta[3]           + beta[5]  + theta[s, HHinc[i]] + u[i]
    logit( q[s,i,x,23]) <- mu[s] + alpha[s,x] + beta[1]          + beta[3] + beta[4]            + theta[s, HHinc[i]] + u[i]
    logit( q[s,i,x,24]) <- mu[s] + alpha[s,x] + beta[1]          + beta[3] + beta[4] + beta[5]  + theta[s, HHinc[i]] + u[i]
}}}

# PART B. Estimate phi using GGD survey data
for (s in 1:2) {
 for (i in 1:I_aggr) {
  for (x in 1:X) {
       # update phi using user-defined multinomial distribution that accepts fractional counts
       sum_Naggr[s,i,x] <- sum( Naggr[s,i,x,1:K] )
       Naggr[s,i,x,1:K] ~ dmulti( phi[s,i,x,1:K] , sum_Naggr[s,i,x] )
       # phi is assigned an informative prior (inputs loaded as data)
       for (k in 1:K-1) { exp_approx[s,i,x,k] ~ dlnorm( mu_approx[s,i,x,k] , tau_approx[s,i,x,k]) }
       phi[s,i,x,1:K] <- informativeNormal( exp_approx[s,i,x,1:K-1] )
}}}

# PART C. Individual model
for (n in 1:N){
 cvd_deaths_ind[n] ~ dbern( ind_prob[n] )
 logit( ind_prob[n] ) <- mu[sex[n]] + sum(Prod_ind[n,] ) + theta[ sex[n] , HHinc[ area[n]] ] + u[ area[n] ]
 # faster than OpenBUGS inprod() function
 Prod_ind[n,1] <- is_age2[n]       * alpha[ sex[n], 2]
 Prod_ind[n,2] <- is_age3[n]       * alpha[ sex[n], 3]
 Prod_ind[n,3] <- is_abstainer[n]  * beta[1]
 Prod_ind[n,4] <- is_excessive[n]  * beta[2]
 Prod_ind[n,5] <- is_smoker[n]     * beta[3]
 Prod_ind[n,6] <- is_obese[n]      * beta[4]
 Prod_ind[n,7] <- is_inactive[n]   * beta[5]
}
```

```
 # PART D. Priors
 # 95% prior believe that the true odds ratio is between 1/10 and 10
 for (b in 1:5){ beta[b] ~ dnorm(0,0.725) }
 for (s in 1:2){
  mu[s] ~ dflat()
  alpha[s,1] <- 0
  theta[s,1] <- 0
  for (a in 2:3){ alpha[s,a] ~ dnorm(0,0.725) }
  for (t in 2:5){ theta[s,t] ~ dnorm(0,0.725) }
 }

 # Leroux et al. (1999) prior <- adjIndex loaded as data
 for (i in 1:I_total){
  u[i] ~ dnorm( cond_mu[i] , cond_prec[i] )
  cond_mu[i] <- rho / (1-rho + rho * numNeigh[i]) * sum( adj_u[ adjIndex[i]:adjIndex[i+1]-1 ])
  cond_prec[i] <- tau_u * (1-rho + rho * numNeigh[i])
 }
 for (i in 1:sumNumNeigh){ adj_u[i] <- u[ adj[i] ] }
 tau_u ~ dgamma(1,0.01)
 #tau_u ~ dt(0,5,1)T(0,)
 rho ~ dunif(0,1)

 # PART E. Monitor odds ratios
 for (s in 1:2){
  mu_odds[s] <- exp( mu[s] )
  for (a in 1:3){ alpha_odds[s,a] <- exp( alpha[s,a] ) }
  for (t in 1:5){ theta_odds[s,t] <- exp( theta[s,t] ) }
 }
 for (b in 1:5){ beta_odds[b] <- exp( beta[b] ) }

} # End Model
```

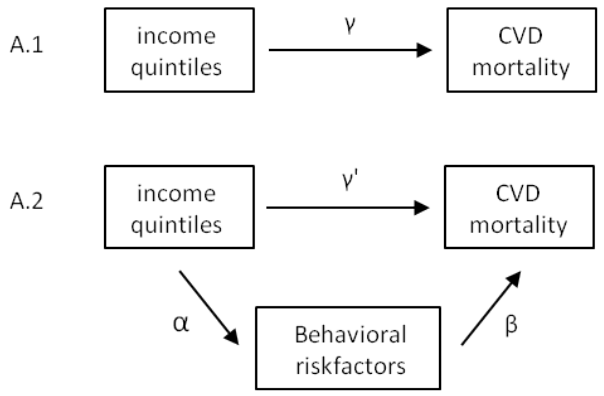
Appendix D. Implementation of the mediation test

To establish an approximation of the mediating effect of the behavioral risk factors on geographical variation in CVD mortality, two different approaches are used. The first approach uses the estimates of γ' and γ from the two HRR models, with γ' and γ denoting the income-quintile parameters derived from the models with and without the behavioral risk factors included, respectively. The overall mediating effect of the behavioral risk factors on income-related inequalities in CVD mortality is then determined as the difference (γ - γ') for the lowest neighborhood income quintiles—i.e., measuring the reduction in the income quintile estimates for the lowest relative to the highest income quintiles.

The second approach does not rely on the HRR estimates of γ and γ'. Instead, a measurement of the same mediation effect is obtained via the multiplication of the marginal effect of income on the prevalence of the behavioral risk factors (α) and the estimated log-odds ratios (β) of the behavioral risk factors, as detailed in Figure 7.3. This product ($\alpha \times \beta$) provides an alternative measurement of the mediation effect (γ - γ') and constitutes a substantially more powerful mediation test because it avoids the subtraction of two HRR estimates that are both measured with relatively low precision.

Because CVD mortality is a discrete event, the validity of the first approach relies on the included Leroux et al.[146] structured error terms to equalize the error variances in the HRR regressions. This is because the size of the estimated coefficients in logit and probit models depends on the error variance of the model, which will typically decrease when explanatory variables are added to the model; see e.g. Karlson et al.[158] and Mackinnon and Dwyer[159] for excellent overviews of the scaling problem in regressions with discrete dependent variables. In general, the non-collapsibility of odds-ratios also poses a problem,[160] but CVD mortality is a relatively rare event (see Appendix B) which implies that the odds-ratios approximate collapsible relative risks.[161] Of course, the Leroux et al. structured terms are only applicable to spatial regressions. Additionally, to the extent that the Leroux et al. terms do not completely equalize the error variances in the base-line and extended HRR model, the error variance in the latter will be smaller (due to the inclusion of additional explanatory variables) and the mediating effect of the behavioral risk factors will be slightly underestimated (cf. MacKinnon and Dwyer[159]).

Figure 7.3 *Illustration of the 2 mediation measurement approaches* *,**



** The mediating effect of the behavioral risk factors is approximated as (γ-γ') or, alternatively, as (α × β) where α represents the marginal effect of the behavioral risk factor prevalence with respect to neighborhood income. ** γ is obtained from the HRR model without behavioral risk factors included, γ' and β are obtained from the HRR model with behavioral risk factors included, and α is obtained from a third model in which the effect of neighborhood income on the prevalence of the behavioral risk factors is determined.*

The validity of both approaches in terms of capturing causal mediation also relies on the use of a correct model specification.[138] One important aspect is the absence of confounders, which we acknowledge to be violated. Indeed, as mentioned in the discussion section in the main text, there are a number of CVD mortality risk factors that are probably correlated with the included behavioral risk factors yet omitted from the model specification. As a result, the total mediation effect as estimated by both approaches will be an underestimation of the total mediating effect of behavioral risk factors on income-related differences in CVD mortality and the individual contribution of the included behavioral risk factors will capture some of the effects of the omitted behavioral risk factors.

Another important aspect is the potential interaction between the income quintiles and the effects of behavioral risk factors on the risk of CVD mortality. Due to sparse data problems, it is not feasible to estimate more elaborate HRR models and test whether the effect of the included behavioral risk factors on CVD mortality varies by income quintile. However, the most likely interaction effect is that lower income groups experience stronger impacts of the behavioral risk factors than higher income groups. For example, with the same base-line measurement, people in lower income groups are

potentially less effective in modifying unhealthy behavior after initial (nonfatal) CVD incidents and hence more likely to experience a subsequent (fatal) CVD incident. They might also be less effective at getting access to the best possible treatment options. If this is the case, the estimated mediation effect would again be an underestimation of the true mediating effect and the second mediation approach would constitute a conservative mediation test.

Taking these caveats into account, the formal mediation test based on the second approach is implemented as follows. First, the required estimate of α is obtained from a separate statistical model that closely resembles the model used to estimate the input for the Chow-Liu approximation. This model utilizes the same pooling of strength approach and similarly uses data from an additional age group (20-34 years) to stabilize the estimates. Here the number of risk factor exposed respondents as obtained from the GGD survey data are assumed binomial distributed:

$$\text{nrAbstainer}_{six} \sim \text{binomial}(\text{prob\_r}_{1\,six}\ ,\ \text{surveyAbstainer}_{six}), \tag{D.1}$$
$$\text{nrExcessive}_{six} \sim \text{binomial}(\text{prob\_r}_{2\,six}\ ,\ \text{surveyExcessive}_{six}), \tag{D.2}$$
$$\text{nrSmoker}_{six} \sim \text{binomial}(\text{prob\_r}_{3\,six}\ ,\ \text{surveySmoker}_{six}), \tag{D.3}$$
$$\text{nrObese}_{six} \sim \text{binomial}(\text{prob\_r}_{4\,six}\ ,\ \text{surveyObese}_{six}), \tag{D.4}$$
$$\text{nrInactive}_{six} \sim \text{binomial}(\text{prob\_r}_{5\,six}\ ,\ \text{surveyInactive}_{six}) \tag{D.5}$$

with the inputs specified by sex (s), area (i), and age group (x). The risk factor prevalences are predicted on the logit scale with an intercept (b_0), age-group specific term (b_1), a sex and quintile-specific income effect (b_2), and a spatial error term (u):

$$\text{logit}(\text{prob\_r}_{1\,six}) = b_0\_r_{1\,s} + b_1\_r_{1\,six} + b_2\_r_{1\,sq} + u\_r_{1\,i}, \tag{D.6}$$
$$\text{logit}(\text{prob\_r}_{2\,six}) = b_0\_r_{2\,s} + b_1\_r_{2\,six} + b_2\_r_{2\,sq} + u\_r_{2\,i}, \tag{D.7}$$
$$\text{logit}(\text{prob\_r}_{3\,six}) = b_0\_r_{3\,s} + b_1\_r_{3\,six} + b_2\_r_{3\,sq} + u\_r_{3\,i}, \tag{D.8}$$
$$\text{logit}(\text{prob\_r}_{4\,six}) = b_0\_r_{4\,s} + b_1\_r_{4\,six} + b_2\_r_{4\,sq} + u\_r_{4\,i}, \tag{D.9}$$
$$\text{logit}(\text{prob\_r}_{5\,six}) = b_0\_r_{5\,s} + b_1\_r_{5\,six} + b_2\_r_{5\,sq} + u\_r_{5\,i}. \tag{D.10}$$

As before, the b_0 terms are assigned weakly informative Normal(0,0.01) priors and the b_1 terms intrinsic conditional autoregressive (CAR) random-walk (1) priors with the adjacency matrices defined based on age-group contiguity. The spatial error terms are assigned Leroux et al.[146] conditionally autoregressive (CAR) priors, and the b_2 terms weakly informative Normal(0,0.725) priors, which represent the 95% prior believe that the true odds ratios are between 1/10 and 10. Note that separate b_2 terms are specified

for the first age group (20-34 years) to avoid potential bias, and that the b_2 terms for the highest income quintile are set to 0 because they serve as the reference category.

The next step is to calculate α, which is the marginal effect of income on the prevalence of the behavioral risk factors. Because the behavioral risk factors are discrete and modelled via a logit-link function, no straight-forward (linear) relationship exists between b_2 and α. Accordingly, α is determined by taking the partial derivative of the inverse-logit function of the right-hand side of equations D.6 – D.10 with respect to b_2, which equals:

$$\alpha\_r_{1\,sqix} = b_2\_r_{1\,sq} \times \text{prob\_r}_{1\,six} \times (1 - \text{prob\_r}_{1\,six}), \tag{D.11}$$

$$\alpha\_r_{2\,sqix} = b_2\_r_{2\,sq} \times \text{prob\_r}_{2\,six} \times (1 - \text{prob\_r}_{2\,six}), \tag{D.12}$$

$$\alpha\_r_{3\,sqix} = b_2\_r_{3\,sq} \times \text{prob\_r}_{3\,six} \times (1 - \text{prob\_r}_{3\,six}), \tag{D.13}$$

$$\alpha\_r_{4\,sqix} = b_2\_r_{4\,sq} \times \text{prob\_r}_{4\,six} \times (1 - \text{prob\_r}_{4\,six}), \tag{D.14}$$

$$\alpha\_r_{5\,sqix} = b_2\_r_{5\,sq} \times \text{prob\_r}_{5\,six} \times (1 - \text{prob\_r}_{5\,six}) \tag{D.15}$$

for income quintiles $q \in [1, 2, 3, 4]$ (the fifth quintile is the reference category) and age groups $x \in [2, 3, 4]$ (the first age group is irrelevant since it is not included in the HRR regressions). In a linear regression these partial derivatives would be constant and equal to the regression parameters, but here the marginal effects depend on the level of the behavioral risk factors, which means that the marginal effects vary by stratum. The average marginal effect of income on the prevalence of the behavioral risk factors is therefore calculated as a weighted average of the risk factor-specific α-terms over all areas i and relevant age groups x, with the available sample size in each stratum used as weights:

$$\overline{\alpha}\_r_{1\,sq} = \frac{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveyAbstainer}_{six} \times \alpha\_r_{1\,sqix})}{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveyAbstainer}_{six})}, \tag{D.16}$$

$$\overline{\alpha}\_r_{2\,sq} = \frac{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveyExcessive}_{six} \times \alpha\_r_{2\,sqix})}{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveyExcessive}_{six})}, \tag{D.17}$$

$$\overline{\alpha}\_r_{3\,sq} = \frac{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveySmoker}_{six} \times \alpha\_r_{3\,sqix})}{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveySmoker}_{six})}, \tag{D.18}$$

$$\overline{\alpha}\_r_{4\,sq} = \frac{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveyObese}_{six} \times \alpha\_r_{4\,sqix})}{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveyObese}_{six})}, \tag{D.19}$$

$$\overline{\alpha}\_r_{5\,sq} = \frac{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveyInactive}_{six} \times \alpha\_r_{5\,sqix})}{\sum_{i=1}^{I}\sum_{x=2}^{4}(\text{surveyInactive}_{six})}. \tag{D.20}$$

The overall mediating effect of the included behavioral risk factors can then be calculated as the sum of the individual contributions of the behavioral risk factors:

$$\begin{aligned}\text{mediated effect}_{sq} = \overline{\alpha}\_r_{1\,sq} \times \beta_{\text{abstainer}} + \overline{\alpha}\_r_{2\,sq} \times \beta_{\text{excessive}} + \overline{\alpha}\_r_{3\,sq} \times \beta\_\text{smoker} + \\ \overline{\alpha}\_r_{4\,sq} \times \beta_{\text{obese}} + \overline{\alpha}\_r_{5\,sq} \times \beta_{\text{inactive}}\end{aligned} \tag{D.21}$$

with posterior means and 95% credible intervals obtained by post-processing the MCMC output.

Table 7.11 provides the estimation results. As can be seen, the mediated effects of the behavioral risk factors have 95% credible intervals that are strictly larger than 0 for the two lowest income quintiles (for males) and three lowest income quintiles (for females). Accordingly, the null-hypothesis of no mediating effect can be rejected for these income quintiles. To summarize the overall mediating effect of the included behavioral risk factors, Table 7.12 shows the decomposition of the combined mediating efect for the lowest income quintile with respect to the individual behavioral risk factors. As can be seen, smoking and obesity are the major mediators for income-related inequality in CVD mortality for males, whereas obesity, physical inactivity and smoking are the important mediators for females. Interestingly, physical inactivity is no major mediator for males; but this concurs with the descriptive statistics provided in Table 7.9.

Table 7.11 *The combined mediating effect of the 5 behavioral risk factors (relative to the highest income quintile), by income quintile and sex**

Income quintile	Sex	
	Male	Female
1	0.18 (0.10—0.27)	0.21 (0.13—0.29)
2	0.11 (0.04—0.19)	0.14 (0.07—0.21)
3	0.04 (-0.04—0.11)	0.12 (0.06—0.19)
4	0.02 (-0.06—0.10)	0.06 (-0.00—0.12)
5	0.00 n/a	0.00 n/a

** Mean posterior estimates on the log-odds scale with 95% credible intervals in parenthesis*

Table 7.12 *The mediating effect of the 5 behavioral risk factors (in the lowest relative to the highest income quintile), by risk factor and sex**

Risk factor	Sex			
	Male		Female	
1. Alcohol abstinence	0.03	(-0.04—0.08)	0.03	(-0.04—0.09)
2. Excessive drinking	-0.00	(-0.03—0.02)	-0.01	(-0.05—0.04)
3. Smoking	0.10	(0.05—0.16)	0.04	(0.01—0.08)
4. Obese	0.05	(0.02—0.09)	0.09	(0.04—0.16)
5. Physical inactivity	0.01	(-0.02—0.04)	0.05	(0.02—0.10)

** Mean posterior estimates on the log-odds scale with 95% credible intervals in parenthesis*

Appendix E. Sensitivity analyses

As explained in the discussion section, the sample size of the GGD health survey is too small to include additional covariates in the HRR model. Still, it is possible to include less risk factors in the HRR models, and in different combinations. To some extent, the outcomes of these analyses are provided by the correlation matrices presented in Table 7.10 but we consider these analyses informative nonetheless.

Here we show the estimation results of five additional models, each with the univariate inclusion of one of the selected behavioral risk factors. For each of these models, an informative prior is estimated using an approach that is similar to the one described in Appendix A. First, the behavioral risk factor data are assumed binomial distributed:

$$\text{nrAbstainer}_{sxae} \sim \text{binomial}(\text{prob\_r}_{1\,sxae}\ ,\ \text{surveyAbstainer}_{sxae}), \tag{E.1}$$
$$\text{nrExcessive}_{sxae} \sim \text{binomial}(\text{prob\_r}_{2\,sxae}\ ,\ \text{surveyExcessive}_{sxae}), \tag{E.2}$$
$$\text{nrSmoker}_{sxae} \sim \text{binomial}(\text{prob\_r}_{3\,sxae}\ ,\ \text{surveySmoker}_{sxae}), \tag{E.3}$$
$$\text{nrObese}_{sxae} \sim \text{binomial}(\text{prob\_r}_{4\,sxae}\ ,\ \text{surveyObese}_{sxae}), \tag{E.4}$$
$$\text{nrInactive}_{sxae} \sim \text{binomial}(\text{prob\_r}_{5\,sxae}\ ,\ \text{surveyInactive}_{sxae}) \tag{E.5}$$

yet without the additional stratification of the risk factors that is required for the Chow-Liu approximation. The risk factor prevalances are again predicted on the logit scale using an intercept and an age-group specific term that takes correlations between behavioral risk factor exposures in adjacent age groups into account:

$$\text{logit}(\text{prob\_r}_{1\,sxae}) = b_0\_r_{1\ sae} + b_1\_r_{1\,sxae}, \tag{E.6}$$
$$\text{logit}(\text{prob\_r}_{2\,sxae}) = b_0\_r_{2\ sae} + b_1\_r_{2\,sxae}, \tag{E.7}$$
$$\text{logit}(\text{prob\_r}_{3\,sxae}) = b_0\_r_{3\ sae} + b_1\_r_{3\,sxae}, \tag{E.8}$$
$$\text{logit}(\text{prob\_r}_{4\,sxae}) = b_0\_r_{4\ sae} + b_1\_r_{4\,sxae}, \tag{E.9}$$
$$\text{logit}(\text{prob\_r}_{5\,sxae}) = b_0\_r_{5\ sae} + b_1\_r_{5\,sxae}. \tag{E.10}$$

Here the b_0 terms are again assigned weakly informative Normal(0,0.01) priors and the

b_1 terms the intrinsic conditional autoregressive (CAR) random-walk (1) prior with an adjacency matrix based on age-group contiguity. Because the joint distribution of the risk factors is not required, $\phi\text{-prior}_{sixk}$ is calculated directly, without making use of a Chow-Liu approximation, as:

$$\phi\text{-prior\_r}_{1\,six1} = \sum_{e=1}^{E}\sum_{a=1}^{A} \text{prob\_}r_{1\,sxea} \times \pi_{sixea}, \tag{E.11}$$

$$\phi\text{-prior\_r}_{2\,six1} = \sum_{e=1}^{E}\sum_{a=1}^{A} \text{prob\_}r_{2\,sxea} \times \pi_{sixea}, \tag{E.12}$$

$$\phi\text{-prior\_r}_{3\,six1} = \sum_{e=1}^{E}\sum_{a=1}^{A} \text{prob\_}r_{3\,sxea} \times \pi_{sixea}, \tag{E.13}$$

$$\phi\text{-prior\_r}_{4\,six1} = \sum_{e=1}^{E}\sum_{a=1}^{A} \text{prob\_}r_{4\,sxea} \times \pi_{sixea}, \tag{E.14}$$

$$\phi\text{-prior\_r}_{5\,six1} = \sum_{e=1}^{E}\sum_{a=1}^{A} \text{prob\_}r_{5\,sxea} \times \pi_{sixea} \tag{E.15}$$

and

$$\phi\text{-prior\_r}_{1\,six2} = 1 - \phi\text{-prior\_}r_{1\,six1}, \tag{E.16}$$

$$\phi\text{-prior\_r}_{2\,six2} = 1 - \phi\text{-prior\_}r_{2\,six1}, \tag{E.17}$$

$$\phi\text{-prior\_r}_{3\,six2} = 1 - \phi\text{-prior\_}r_{3\,six1}, \tag{E.18}$$

$$\phi\text{-prior\_r}_{4\,six2} = 1 - \phi\text{-prior\_}r_{4\,six1}, \tag{E.19}$$

$$\phi\text{-prior\_r}_{5\,six2} = 1 - \phi\text{-prior\_}r_{5\,six1} \tag{E.20}$$

for $k \in [1, 2]$.

Having estimated ϕ-prior, the next step is to translate the posterior of ϕ-prior into an informative prior that can be used by OpenBUGS to estimate ϕ in the HRR model. Because there are only two categories in each stratum (i.e. smoker/non-smoker, obese/non-obese, etc.) a standard Dirichlet distribution is sufficiently flexible. Starting from the mean and variance of the Dirichlet distribution, which are given by:

$$\mu[\phi\text{-prior}_k] = \alpha_k/\alpha_0 \tag{E.21}$$

and

$$\sigma^2[\phi\text{-prior}_k] = \frac{\alpha_k \times (\alpha_0 - \alpha_k)}{\alpha_0^2 \times (\alpha_0 + 1)} \tag{E.22}$$

with α_k representing the $1, \ldots, K$ parameters of the Dirichlet distribution and α_0 denoting the sum of these parameters (i.e. $\sum_{k=1}^{K} \alpha_k$), the parameters of the informative Dirichlet distribution can be derived from the posterior mean and standard deviations of ϕ-prior as follows. First, $\alpha_{0\,sixk}$ is calculated as:

$$\alpha_{0\,sixk} = \frac{\mu_{sixk} \times (1 - \mu_{sixk})}{\sigma^2_{sixk}} - 1 \tag{E.23}$$

after which the individual input parameters can be determined as:

$$\alpha_{sixk} = \mu_{sixk} \times \frac{\sum_{k=1}^{K} \alpha_{0\,sixk}}{K} \tag{E.24}$$

with $\mu_{six\,k}$ and $\sigma_{six\,k}$ in equation E.23 and equation E.24 denoting the posterior means and standard deviations of the estimated ϕ-prior, respectively.

Compared to the informative prior for the full HRR model, the informative Dirichlet based on the posterior distribution of ϕ-prior is substantially more informative (the same amount of data is used to estimate both priors, yet the first requires 9 and the second only a single parameter per stratum). At the same time, the need for an informative prior is much smaller (the same amount of area-specific survey data is available to estimate ϕ directly, with ϕ consisting of 2 rather than 24 categories). Hence, to reduce the impact of the informative prior, we downgraded the precision of the informative Dirichlet prior by a factor of 10. We also implemented an improved non-conjugate Dirichlet update algorithm to estimate the HRR models in OpenBUGS, which reduced the run-time of the models considerably.

Table 7.13 contains the estimation results of the behavioral risk factors in each of the five additional HRR regressions. As can be seen, the size of the coefficients of the behavioral risk factors has changed compared to those in the full HRR model, reflecting correlations between the behavioral risk factors (cf. Table 7.10). At the same time, the ranking of the behavioral risk factors in terms of impact on CVD mortality remains unchanged: smoking remains the most important factor (4.5 times higher risk of CVD mortality), being obese (3.3 times higher risk of CVD mortality) and physically inactive (2.4 times higher risk of CVD mortality) are also important factors, whereas excessive drinking and alcohol abstinence clearly less important and have 95% credible intervals that contain 1).

Table 7.13 *Hierachical Related Regression behavioral risk-factor estimates (based on the univariate inclusion of each risk factor)*

Risk factor	Estimate*
1. Is abstainer	1.35 (1.10 – 1.64)
2. Is excessive drinker	1.11 (0.46 – 1.84)
3. Is smoker	4.49 (3.36 – 5.97)
4. Is obese	3.33 (2.44 – 4.30)
5. Is physically inactive	2.36 (1.83 – 2.99)

** Mean posterior estimates on the log-odds scale with 95% credible intervals in parenthesis*

8

Discussion

8.1 INTRODUCTION

In this thesis, a Bayesian random-effects methodology for the estimation of small-area indicators of population health has been developed and validated using Monte Carlo simulations. Furthermore, the feasibility and merit of the selected indicators to map, rank, and select areas for policy purposes has been investigated. Finally, geographic differences in small-area population health have been investigated taking spatial dependencies and sparse data problems adequately into account.

In what follows, the main findings are summarized in **section 8.2**, the strengths and limitations of the presented methodologies discussed in **section 8.3**, technical limitations summarized in **section 8.4**, the scientific contribution, new insights and policy relevance summarized in **section 8.5** before the final **section 8.6** concludes.

8.2 SUMMARY OF FINDINGS

The first research aim was to develop and validate new methodology to estimate indicators of small-area population health.

At the small-area level, there is no straight-forward benchmark or "golden" measurement by which different methodologies to estimating measures of population health can be compared. Accordingly, to be able to compare the performance of existing and newly developed methodologies, a benchmark has been proposed in **chapter 2** that mimics the exact male and female population age structures, age-specific mortality rates, and geographic locations of a large number of European countries. This benchmark encompasses relatively large differences in population health and has the advantage that it is based on real populations and thereby avoids the use of arbitrary correlations. Using this benchmark, different methodologies for estimating small-area indicators of population health can be compared using Monte Carlo simulations.

Based on the Monte Carlo simulation evidence presented in **chapter 2** and **chapter 3** of this thesis, it was established that the traditional life table approach to estimating indicators of population health cannot accommodate the sparse data problems that are typically encountered in small-area applications. For populations smaller than approximately 5,000 person years at risk, and survey sample sizes smaller than approximately 100 persons per life table, the traditional approach should not be used to estimate LE and HLE. The resulting mean errors, standard errors, and root mean

squared errors become too large for reliable use and the coverage of the 95% confidence intervals drops below 92% (LE) and 89% (HLE).

The Monte Carlo simulation evidence presented in **chapter 2** and **chapter 3** also established that a Bayesian random-effects approach to estimating indicators of small-area population health performs better than the traditional approach. As shown, the proposed Bayesian random-effects models can be used to reliably estimate LE for populations as small as 2,000 person years at risk, and HLE for populations as small as 1,750 person years at risk when combined with a minimum survey sample size of approximately 45 persons per life table. Moreover, even for the smallest recommended population and survey sample sizes, the proposed Bayesian random-effects approaches produce almost unbiased estimates and exhibit 95% credible intervals with close to perfect coverage. The latter applies both to estimations of LE and HLE.

The second research aim was to investigate the feasibility and merit of the selected indicators to map, rank, and select areas in real-life analyses for policy purposes.

When older persons move to a nursing home, it is usually in the final years of their lives and it often involves a re-location to a different neighborhood. As shown in **chapter 4**, the latter can strongly distort small-area health comparisons, which makes the location of nursing homes an important confounder in real-life small-area estimations. More specifically, **chapter 4** shows that neighborhoods with nursing homes can have uncorrected life expectancies that are up to six years too low due to the selective migration of frail individuals to nursing homes. Conversely, neighborhoods without nursing homes have uncorrected life expectancies that are up to five years too high. Accordingly, a nursing home correction for small-area life table estimations is required if the estimates are to be interpreted as indicators of average population health.

The ideal correction for the migration of frail elderly to nursing homes is to re-assign all nursing home residents to their previous residential address prior to the calculation of the summary indicators of population health. This correction, however, requires detailed individual-level and longitudinal data, which are often unavailable. Fortunately, one of the advantages of the Bayesian random-effects approach is that it is easily extendible to incorporate corrections for area and age-group specific confounders, including the spatial distribution of nursing homes, as an integral part of the life table estimations. As shown in **chapter 4**, such a Bayesian model-based correction removes the impact of nursing homes from the age-specific mortality rates. This provides a reliable alternative to the ideal nursing home correction and can do

so with relatively modest data requirements: only the (cross-sectional) percentage of nursing home deaths per area is required.

Another important aspect when conducting small-area health comparisons is the use of an indicator that can provide balanced attention to the effects of fatal as well non-fatal health outcomes on overall population health. When the required data are available, health-adjusted life expectancy (HALE) is more sensitive to changes in the severity of morbidity within a population than other health expectancy measures, such as healthy life expectancy or disability-free life expectancy. Accordingly, in **chapter 5** a Bayesian random-effects approach to estimating HALE at the small-area level has been developed. The modeling approach is an extension of the methodology as developed in **chapter 2** and **chapter 3** and its application is illustrated in a real-life example, which confirmed the feasibility and reliability of the HALE estimations. Furthermore, **chapter 5** demonstrated how the estimated indicators can be graphically presented and how the full uncertainty of the estimates can be taken into account when creating rankings of the area's HALE. The latter is particularly useful for the selection of areas for policy purposes, such as, for example, area based programs that only target a small number of disadvantaged neighborhoods.

The third research aim was to explain geographic small-area differences in population health, making use of the methodologies as developed in this thesis.

Urban regeneration and neighborhood renewal programs often comprise urban green space developments. These are primarily aimed at creating more attractive neighborhoods, but increasingly also at affecting and improving population health. Based on the developed Bayesian random-effects methodology to estimate small-area LE and HLE, the impact of three different measures of urban green on small-area population health was evaluated in **chapter 6**. In contrast to previous epidemiological studies, the statistical methodology simultaneously accounted for spatially correlated unobserved determinants, measurement uncertainty (both in the dependent and independent variables) and the migration of frail elderly to nursing homes. Adjusted for differences in average household income, it was found that the quantity and quality of urban green space are modestly related to small-area LE and HLE: an increase of 1 standard deviation in the percentage of urban green is associated with a 0.1-year higher LE, and, in the case of the perceived quality of green, with an approximately 0.3-year higher LE and HLE.

Despite their advantages, the use of summary measures of population health in explanatory research has proven to be of mixed value because the observed associations cannot be interpreted as individual-level associations (due to the ecological nature of the indicators, which already is a problem for aggregated individual-level indicators but proved to be an even more articulate problem with summary measures of population health that are—by definition—only defined for populations as a whole). Hence in **chapter 7** a different methodological approach was used to investigate the contribution of five important behavioral risk factors to small-area differences in mortality from cardiovascular disease (CVD). The implemented Hierarchically Related Regression methodology combines the benefits of both an ecological analysis (in terms of data availability and statistical power) and an individual-level analysis (in terms of identification of the parameters and interpretation of the results), and was used to investigate the extent to which income-related inequalities in small-area CVD mortality are mediated by several behavioral risk factors. After correcting for differences in age and sex, it was found that individuals who reside in the poorest neighborhoods have on average a 66% and 34% higher risk of CVD mortality than those who reside in the richest neighborhoods (for men and women, respectively). Furthermore, accounting for differences in five behavioral risk factors (i.e. smoking, excessive drinking, alcohol abstinence, obesity, and physical inactivity) reduced income-related inequalities in CVD mortality by approximately 30%. Additionally, given the geographic clustering of unhealthy risk factor behavior, area-based targeting of deprived neighborhoods was confirmed to be a relevant approach to address income-related inequalities in CVD mortality. Particularly smoking, physical activity, and overweight are identified as interesting candidates for interventions, as they have the largest impact on CVD mortality and exhibit substantial prevalence differences across the income quintiles.

8.3 METHODOLOGICAL CONSIDERATIONS

In this thesis, several innovative methods and methodological innovations have been introduced, evaluated, and applied in real-life analyses. These include the introduction of a new benchmark to evaluate small-area estimators of population health, Bayesian random-effects methodology to estimate various small-area indicators of population health, (ecological) regression methodology to explain spatial variation in small-area indicators, and hierarchically related regression methodology that combines ecological and individual-level survey data. In all of these contributions, methodological considerations have played an important role.

Comparing small-area estimation approaches using Monte Carlo simulations

Starting with the introduction of a new small-area estimation benchmark, an important consideration was that various standard statistical measures of fit, such as the Aikaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), all involve different trade-offs between goodness-of-fit and model parsimony. Whereas such indicators are informative about relative model fits, they require a subjective choice about the degree to which achieving a higher likelihood at the expense of increasing the number of parameters in the model is penalized. Additionally, statistical measures of fit are uninformative about the absolute fit of any particular model. Accordingly, the use of Monte Carlo simulations based on a pre-defined benchmark provided a better strategy to compare different methodologies for estimating small-area indicators of population health. The problem, however, is choosing a benchmark that adequately reflects the conditions encountered in small-area estimations.

One possibility to construct the benchmark is to rely on actual small-area data. Under this option, the benchmark could either be based on the results of a small-area estimation model or on the pooling of several years of data and calculating the required age-specific mortality rates directly. Both approaches, however, are problematic. The use of a model-based estimation implies smoothed age-specific mortality rates that reflect the specification of the model used to estimate them, resulting in a benchmark that is poorly suited for model comparisons (i.e. models with a different structure and/or different priors). Additionally, such a smoothed benchmark provides an unfair advantage for random-effects and mixed-effects models over traditional fixed-effects models that do not perform better with smoothed mortality rates and morbidity prevalences. The pooling of multiple (i.e. more than 4 or 5) years of data and calculating the age-specific mortality rates directly is also problematic for areas with changing populations, e.g. due to urban gentrification or increasing segregation and deprivation. Finally, in both cases the resulting benchmarks would be area-specific and leave the generalizability of the model comparisons contingent upon the similarities between the small-area characteristics of the benchmark and the area of implementation.

In contrast, the benchmark based on a large number of European countries is based on life table data obtained from millions of inhabitants and requires no pooling of data to obtain a reliable benchmark. Furthermore, the correlations between age groups, sexes, and areas are 'biologically feasible', something which is easily violated with artificially imposed correlations. Finally, the observed differences in population health are substantial and considerably larger than typically found in small-area analyses.

The latter ensures that Monte Carlo simulation evidence based on the European benchmark provides a worst-case scenario for the random-effects and mixed-effects models, which perform better with more similar areas included in the benchmark. It also implies that the presented evidence can safely be interpreted as a lower-estimate of the true performance of these models, with the determined sample size reflecting a conservative (i.e. minimum) sample size requirement for real-life applications.

The disadvantage of using Monte Carlo simulations is the large number of repetitions required to obtain reliable simulation results. Particularly when using Bayesian methods with spatial distributions, which typically require several minutes of runtime per model, the estimation of thousands of models is cumbersome and requires considerable computational power. This is further discussed in the practical considerations (i.e. section 8.3) and constitutes, for example, the reason that a) the sensitivity of the Monte Carlo results for the Bayesian random-effects models with respect to the chosen priors, and b) the performance of several more computationally demanding Bayesian random effects specifications were not investigated. Given the (a priori) known benchmark that is used in Monte Carlo simulations this does not imply that the performance of the random-effects specifications that were evaluated in **chapter 2** and **chapter 3** are potentially mis-interpreted or that their performance is imprecisely calculated. However, it does imply that the Bayesian random-effects could perform slightly better (or worse) with a different specification of the chosen priors, and that more complex, for example, spatially adaptive specifications could potentially further improve the performance of the Bayesian random-effects vis-a-vis the traditional life table approach.

Ecological regressions

Turning to the (ecological) regression methodology that was used to investigate associations between the estimated small-area indicators of population health and several important determinants, the first important methodological consideration was measurement uncertainty, or more precisely, differences in error variability of the life table estimates. As shown in the Monte Carlo simulation studies, the standard errors and root mean squared errors of the life table estimates are strongly influenced by the population and survey sample sizes of the neighborhoods. Given that population sizes vary between neighborhoods, a linear regression with LE or HLE as the dependent variable exhibits heteroskedastic error terms. This results in unreliable estimates of the standard errors and 95% credible intervals and in a reduction of statistical precision, which is an equally important problem in small-area applications given the

usually limited number of observations. Accordingly, to solve the heteroskedasticity and improve the precision of the estimates, the ecological regressions included a weighting approach that is very similar to that of standard weighted least squares regressions. More specifically, the individual LE and HLE estimates were weighted with the (reciprocal of) the estimated variances in the random-effects models used as weights, which ensures that areas with more reliable life table estimates carry more weight in the regression analysis.

Another methodological consideration was the spatial dependency of the regression errors between geographically adjacent areas. When not properly accommodated, the statistical significance of the predictors are inflated, resulting in a higher probability of finding statistically significant effects than actually supported by the data.[162] Moreover, parameter estimates can change and even result in reversals of direction if the spatial error terms pick up previously omitted confounders.[163] For these reasons, all models included in this thesis included spatial error terms, which increases the robustness of the results and corrects for omitted confounders that are spatially structured, and thereby improves upon a large proportion of the public health literature concerning spatial (determinants of) health.

The last important consideration was the selection of independent variables. Initially, a broad range of determinants was selected based on the Five Capitals model. This model, originally developed within the field of sustainable development, provides a theoretical framework for the types of assets and resources that are needed for a community to attain and sustain good health outcomes. Following Poortinga,[164] the Five Capitals model suggests that healthy neighborhoods require a balanced combination of 1. human capital (e.g. skills and education), 2. economic capital (e.g. income and wealth). 3. social capital (e.g. social networks), 4. natural capital (e.g. clean air and access to green space), and 5. built capital (e.g. access to amenities and healthcare). However, when used to explain spatial patterns of LE and HLE, the explanation of the results proved complicated. For example, the effects of human capital and economic capital were strongly correlated, making their individual contributions difficult to determine. Second, air pollution at the neighborhood level exhibited strong spatial autocorrelation, making it difficult to separate from other spatially structured determinants and sensitive to the specification of the spatial error terms. Third, when looking at the provision of health care, it was found that a higher concentration and/or closer average distance to the nearest health care provision was associated with lower life expectancy – which cannot be interpreted as a direct causal effect and probably reflects a mixture of other mechanisms, e.g. the financial incentive for

general practitioners in the Netherlands to practice in disadvantaged neighborhoods, the tendency to build new hospitals at the cities' outskirts where land-prices are lower and access-routes are better, and a 'cross-over effect' in the sense that those who need care tend to locate themselves dynamically in places where that care is available (see e.g. Starfield and Shi[165]). For green and social capital plausible results were found, the first of which was further investigated in this thesis.

Hierarchically related regressions

Turning to the hierarchically related regression (HRR) methodology, the reason for switching to this approach was the inherent limitation in the interpretation of regressions with LE and HLE as outcome measures with respect to individual-level associations. In contrast, the hybrid HRR methodology combined the benefits of both an ecological analysis (in terms of data availability and statistical power) and an individual-level analysis (in terms of identification of the parameters and interpretation of the results). A major methodological consideration in the application of the hierarchically related regression methodology was the incorporation of behavioral risk factors, which introduced a new layer of complexity to the regression framework. Most importantly, the limited number of survey respondents was accommodated using a carefully constructed informative prior, which highlights one of the advantages of the Bayesian estimation framework. Also, the construction of the informative prior relied on a pooling of strength approach that closely resembles the Bayesian random-effects methodology of the LE and HLE models and which provided a methodological consistency between the different chapters of this thesis.

A final consideration was the estimation of the mediating effect of the behavioral risk factors on geographical variation in mortality from cardiovascular disease (CVD). Because CVD mortality is a binary outcome, this proved to be rather complicated. The main problem is that the estimated coefficients in logit and probit models that are used for statistical analyses of binary dependent outcomes depend on the error variance of the model, which typically decreases when explanatory variables are added to the model. Hence the coefficients of two (nested) Hierarchical Related Regression models will not be on the same scale, which invalidates standard mediation analyses. One of the implemented solutions was the inclusion of a comprehensive error term (i.e. the conditionally autoregressive prior as proposed by Leroux et al.[146]), which equalizes the error variances and makes the estimated parameters roughly comparable. Although not perfect, some preliminary tests indicated that the Leroux et al. term actually performed very well. Also, to the extent that the error variance

would not be completely captured by the included error term, the mediation analysis provides a lower estimate of the true mediation effect. The second approach that was implemented was a formal mediation test; basically the Bayesian equivalent of the bootstrap test as proposed by Preacher and Hayes,[144,145] with a modification to take the binary structure of the behavioral risk factors in **chapter 7** into account. This test required a separate estimation of the marginal effect of income on the prevalence of the behavioral risk factors, which was based upon the same techniques as used in the Bayesian random-effects models to handle sparse data problems.

8.4 TECHNICAL LIMITATIONS

To handle the sparse data problems in the various chapters of this thesis, Bayesian methods proved to be convenient. They are unbiased with respect to sample size, allow for a relatively straight-forward estimation of complicated spatial models using MCMC methods, and adequately approximate the posterior distribution of the parameters without relying on normality assumptions that are often inappropriate given the sparse data. An important disadvantage of the Bayesian approach, however, was the required computation time. For a single LE or HLE model, the computation time is orders of magnitude larger when performed using Bayesian MCMC compared to the standard life table approach as implemented e.g. in Excel of MATLAB. As an indication, 5 – 15 minutes for the Bayesian estimation of the Bayesian random-effects models as used in this thesis (using OpenBUGS on a single core CPU) compared to an almost instantaneous result for the traditional life table estimations in Excel.

For a single estimation, the computation time is inconvenient but not a serious limitation. On the other hand, for the Monte Carlo simulations conducted in this thesis, the computation time was a major limitation of the Bayesian MCMC estimation approach. Here hundreds of thousands of models needed to be fit, which implies a total computation time of many CPU-years. Alternative Bayesian estimation packages, such as JAGS or STAN, do not include dedicated spatial distributions and are consequently much slower than BUGS for spatial models. Only a problem-specific implementation of the models, for example in MATLAB, Julia, or C would have substantially reduced the required computation time. Unfortunately, the latter requires substantial technical expertise that was, certainly in the earlier stages of the research, beyond the scope of this thesis. Instead, the simulations were conducted using OpenBUGS on a number of workstations and, when necessary, distributed on the Dutch Life Science Grid. The latter allowed the calculations to be performed in a number of weeks.

Another practical limitation was the requirement of accessing large datasets via the remote-access secure data infrastructure of Statistics Netherlands. Initially, the secure infrastructure was limited to 1 Gigabyte (Gb) random-access memory (RAM) per scientific user. As a result, editing the Dutch population registry with a size of approximately 7 Gb's, was difficult to program and time consuming to conduct. Merging the Dutch population registry (GBA) with external information was even more problematic. Only towards the end of the research project, when the CBS data infrastructure was upgraded to 64-bits servers, individual scientific users were allowed to address more than 1 Gb of RAM on the remote-access infrastructure. This made it considerably easier and faster to edit the GBA and merge it with additional data sources and thereby effectively facilitated the longitudinal merges of registry information as required for the Hierarchical Related Regression model in **chapter 7**.

Around the same time when the remote-access infrastructure of Statistics Netherlands was updated, OpenBUGS was granted permission to run on the secure data infrastructure with administrative rights. Because the individual-level and ecological CVD mortality data were not allowed to be exported and analyzed externally, and since administrative rights were required to be able to compile and use the custom functions, distributions, and samplers for the Hierarchical Related Regression models on the remote access infrastructure, this decision essentially accommodated the estimation of the Hierarchical Related Regression models in the first place. However, the secure servers were still quite old (slow) and not particularly suitable for computationally intensive estimations. The Hierarchical Related Regression models, for example, took 3.5 weeks of run-time to complete, during which (manual) logins were required every third day to keep the estimations running. Especially during the holidays the latter proved to be a major practical limitation.

8.5 SCIENTIFIC CONTRIBUTION AND PRACTICAL RELEVANCE

The results presented in this thesis are in various ways related to previously published results. At the same time, each chapter has extended and/or improved upon previous findings. The following sections cover both the scientific contribution as well as the practical relevance and policy implications.

Traditional life table approach

First, the use of traditional life table methodology for estimating small-area LE and HLE was initially considered by Eayres and Williams,[37] who found that the traditional methodology should not be used for populations smaller than 5,000 person years at risk. These results, which were based on Monte Carlo simulations with UK population data, were subsequently confirmed by Scherbov and Ediev[36] in more elaborate Monte Carlo simulations. The results of the Monte Carlo simulations presented in **chapters 2** and **3**, which were based on 33 and 28 European countries, again confirmed these findings using a more diverse benchmark. Additionally, the presented Monte Carlo simulations showed that the coverage of the 95% confidence intervals of the LE and HLE estimates are too optimistic regarding the actual performance of the traditional life table approach, even so for populations as large as 25,000 person years at risk.

The first important practical implication of **chapters 2** and **3** is that the traditional life table approach should not be used for populations that are smaller than 5,000 person years at risk or survey sample sizes smaller than approximately 100 respondents per life table. In the Dutch context, this implies that traditional life table methods should not be used to calculate summary measures of population health at the neighborhood level, while even estimations at the district-level ("wijk") should be conducted with careful consideration of the established minimum required sample sizes. Additionally, for those districts with populations smaller than approximately 25,000 person years at risk, the practical implication of **chapters 2** and **3** is that an alternative method should be used to calculate the standard errors and 95% confidence intervals of the life table measures—a method that does not rely on asymptotic normality assumptions, such as a (non-parametric) bootstrap approach.[166]

Bayesian random-effects approach

For example Congdon,[57,67,167] Srebotnjak et al.,[78] Kulkarni et al.[75] and Wang et al.[81] already considered the use of random and mixed-effects methodology to estimating small-area life expectancy measures. However, the performance of the various specifications relative to the traditional life table estimation methodology was not properly investigated; this was a specific contribution of **chapters 2** and **3**. Interestingly, only Congdon uses a Bayesian estimation framework. In the simulations presented in **chapter 3** it was found that the non-Bayesian models lacked flexibility and parsimony and performed worse than the traditional life table methodology. This highlights the advantage of the Bayesian estimation framework to conveniently fit parsimonious yet flexible model specifications, which is essential for outperforming the traditional life

table methodology in small-area applications.

The second important practical implication of **chapters 2** and **3** is the minimum population size and survey sample size that is established for the reliable estimation of LE and HLE in small-area applications. For the recommended random-effects specifications, the minimum population and survey sample sizes are set at 2,000 person years at risk for the LE estimations and 1,750 person years at risk and 45 survey respondents per life table for the HLE estimations, respectively. These are substantially smaller than the minimum population sizes required for the traditional life table approach, which reflects the superior performance of the Bayesian random-effects models vis-à-vis the traditional life table methodology. Consequently, when making use of the Bayesian random-effects models, many more areas can be included and life expectancies can be calculated at much smaller intervals than previously possible; in the Dutch context, the latter has made it possible to reliably estimate (healthy) life expectancy at the neighborhood level, which was previously not possible.

The effect of nursing homes on small-area mortality outcomes

Third, the effect of the location of nursing homes on small-area mortality outcomes was already described by Garnder and Winter[168] and has subsequently been confirmed in various settings and with various health outcomes (see e.g. Chellini et al.,[169] Congdon,[170] Williams et al.,[64] and Gandarillas et al.[83]). However, the data required for a nursing home correction involves historical information about the last-known residential address of all nursing home inhabitants, which is often unavailable. Hence such corrections are seldom used, with the exception of Veugelers and Hornibrook,[85] who obtained previous residential address information for approximately 80% of the nursing home population in Nova Scotia, Canada. In this respect, the Bayesian model-based correction as proposed in **chapter 4** provides an interesting alternative with much more modest data requirements: only a single (cross-sectional) percentage of nursing home deaths per area is required.

The practical implication of **chapter 4** is therefore that the impact of nursing homes on small-area life expectancy estimations is substantial and that uncorrected life expectancies should not be used for inter-area comparisons if the estimated life expectancies are to be interpreted as indicators of average population health. Instead, nursing home correct summary measures should be used; ideally based on a previous residential address correction, but when the required longitudinal information is unavailable, the presented Bayesian model-based nursing home correction provides a reliable alternative.

Methodology to map, rank, and select areas for policy purposes

Fourth, the methodology to estimate health-adjusted life expectancy at the small-area level is a novel contribution of **chapter 5**. Furthermore, the methodology to map, rank and identify areas for policy purposes originated in the Bayesian disease mapping literature (see e.g. Lawson[171]) and in the institutional performance literature (e.g. the ranking of schools and hospitals, see Goldstein and Spiegelhalter[172]). The application of these methods in a small-area public health setting is again a contribution of **chapter 5**, which enables researchers to select, measure, monitor, and evaluate area-based interventions with an attractive summary measure of population health.

The practical contribution of **chapter 5** is therefore that it provides researchers and policy makers an additional (or, alternative) indicator to select deprived neighborhoods for inclusion in area-based programs. Furthermore, given the modest impact that area based programs can realistically be expected to exert on average population health within a relatively short follow-up period (i.e. health impact assessments are usually conducted less than 10 years after the initial implementation of the programs), the ability of the proposed methodology to evaluate health impacts at a higher frequency while establishing statistical significance for much smaller impacts is also an important benefit. Finally, for Dutch policy makers the required data are already routinely collected and available for policy-related research via Statistics Netherlands and the Dutch health authorities (GGD). Admittedly, whereas the required mortality information is available annually, large-sample survey data with sufficient respondents at the neighborhood level are currently only available at 4-year intervals. In principal, the latter suffices for inclusion in research that comprises a larger time span, such as, for example, the Dutch regional and national Public Health Status and Foresight Reports (i.e. "volksgezondheid toekomstverkenningen"), which are published once every 4 years. However, if higher precision or shorter measurement intervals are required, these data can be supplemented with intermediary survey data. As shown in **chapter 3** and **chapter 5**, the Bayesian random-effects approach only requires modest survey sample sizes, which would make the latter a feasible approach to establish more frequent estimates.

The impact of urban green on small-area population health

Fifth, regarding urban green, there is ample evidence that contact with green can have substantial health benefits.[173] Whereas the available evidence is relatively strong at the individual level and often based on randomized case-control studies, much less is known about the effect of the proximity, quantity and quality of green spaces on

population health. The available evidence, however, seems to suggest that urban green constitutes an independent factor in the explanation of spatial health inequalities, which implies that it can contribute towards a reduction in socioeconomic inequalities in health.[120,174] The results presented in **chapter 6** confirm these findings with LE and HLE as outcome measures while taking spatial autocorrelation and the migration of frail elderly to nursing homes into account. Furthermore, **chapter 6** also takes the quality of urban green into account, which was found to be stronger related to small-area health than traditional measures that only take the proximity to and amount of urban green into account.

The first policy implication of **chapter 6** is that urban green development seems to deserve a more prominent place in urban planning, with the results of **chapter 6** indicating that not only the quantity but also the quality of urban green should be considered in future interventions. As indicated, this recommendation is based on comparatively strong evidence with respect to individual-level health outcomes, together with preliminary yet growing evidence of the positive impact of green on population health outcomes, of which **chapter 6** is a good example.

At the same time, there might still be some skepticism with respect to the current evidence base, particularly because cross-sectional research designs are notoriously problematic for establishing reliable causal effects. The second policy implication of **chapter 6** is therefore related to the observation that further cross-sectional research will add little to no additional weight to the current evidence base. Instead, in order to establish stronger (causal) effect estimates that can further substantiate the impact of urban green on small-area population health, the appropriate next step would be to implement and evaluate urban green improvements in a randomized controlled trail, or conduct similar longitudinal assessments based on natural experiments if the latter can be identified. Of course, in order to isolate the impact of green on population health as much as possible, these interventions and/or natural experiments should be as isolated and independent as possible from (pre-)existing area-based programs that target a broad range of public health determinants.

The impact of behavioral risk factors on geographic inequality in cardiovascular disease mortality

Finally, the Hierarchical Related Regression approach was introduced by Jackson et al.[20] with further applications of Jackson et al.[21,130] The methodology is elegant and combines the strengths of both individual-level analyses and ecological analyses. However, the methodology is also complicated and the required data difficult to obtain

(since it requires matching individual-level and aggregate data), which implies that there are few substantive applications in the literature. In addition to providing such an application, **chapter 7** shows how behavioral risk factor data can be integrated into the Hierarchical Related Regression methodology and how mediation analyses with the Hierarchical Related Regression methodology should be conducted. Both are methodological challenges that constitute important contributions. At the same time, also the presented results are important: it was found that approximately 1/3 of the income related inequality in small-area CVD mortality is related to the five included behavioral risk factors, i.e. alcohol abstinence, excessive drinking, smoking, obesity, and physical inactivity. Using a small simulation study, it was also established that these behaviors cluster geographically with a higher concentration of behavioral risk factors in low-income areas. Finally, a decomposition of the overall effect showed that smoking (men) and being obese (women) were the most important mediators.

The policy implications of **chapter 7** are at least twofold: first, the direct targeting of excess prevalence of unhealthy behaviors in deprived neighborhoods appears to be a relevant strategy to reduce income-related inequalities in CVD mortality. Particularly smoking and overweight are interesting candidates for interventions, as they have a large impact on CVD mortality and exhibit substantial differences in prevalence between lower and higher income neighborhoods. Second, given the established geographic clustering of unhealthy risk factor behavior, area-based targeting of deprived neighborhoods seems to be a relevant and potentially effective approach to address income-related inequalities in CVD mortality. The possibility of the HRR approach to identify specific neighborhoods and evaluate potential interventions before they are actually implemented would nicely complement such area- based strategies. The latter was briefly mentioned in the discussion section but was otherwise beyond the scope of **chapter 7** and forms one of the interesting potential avenues for future research.

8.6 FUTURE RESEARCH

There are many directions for future research, ranging from the valorization of the developed methods to additional investigations using the developed methodologies and further methodological improvements that can refine the small-area LE and HLE estimations.

Methodological improvements

Starting with further methodological improvements, there are several interesting avenues for further refinements of the small-area LE and HLE estimations. These include, for example, spatially adaptive prior structures,[63] more flexible Wishart specifications,[175] and Bayesian variable selection techniques.[176] These technical improvements could improve the accuracy of the presented methodologies and accommodate more flexible longitudinal extensions of the Bayesian random-effects methodology. Such longitudinal extensions would provide a welcome opportunity to monitor small-area population health over time with higher sensitivity and at smaller time intervals. Additionally, the same techniques could be used to develop methodology that is better suited to accommodate the simultaneous inclusion of smaller and larger areas. Such circumstances are, for example, encountered in New South Wales in Australia[177] and the models as developed in this thesis could be adapted to better accommodate such applications.

Valorization

The Bayesian random-effects models are relatively difficult to implement for the intended audience (i.e. public health researchers, public health authorities, and demographers). Particularly the need to specify the spatial matrix, specify the life table data in the correct format, provide initial values, and monitor appropriate mixing and convergence of the MCMC chains constitute a steep learning curve. Accordingly, a dedicated small-area (healthy) life expectancy estimation tool would be essential to promote further implementation of the developed methodology. Such a tool would simplify the data management, include appropriate tests for monitoring convergence, and could be programmed in the C or C++ language to speed up the required computations. As an indication, a speed-up of up to 20x relative to OpenBUGS would be feasible. Altogether, such an estimation tool would make the (healthy) life expectancy estimations as described in this thesis considerably easier and faster to carry out, and

allow a non-technical audience to estimate small-area (healthy) life expectancies using the methodology as developed in this thesis.

Additional investigations

Turning to the additional investigations, several options have already been mentioned. For example, a longitudinal extension as well as accommodation of areas with non-uniform population sizes provides interesting new avenues for future research. Additionally, applications of the existing methodology with health gaps (instead of life expectancies), the application of life table dispersion measures to investigate within (cross-sectional) and between-area (longitudinal) inequalities in small-area health (similar to e.g. Vaupel et al.[88] and Goesling and Firebaugh[178] at the international level), investigations into the effect of social capital on small-area population health (e.g. with a differences-in-differences approach), the application of the developed hierarchically related regression methodology to all-cause mortality (similar to e.g. Nandi et al.[137]), possibly combined with area-specific simulations of specific policy interventions, would all provide interesting new insights into the measurement and explanation of differences in small-area health. Some of these will be carried out in the near future, whereas other ideas are left for others to complete.

9

Summary

9.1 SUMMARY

Urban policy is a major route through which governments and health authorities attempt to deliver improvements in living conditions, economic opportunities, and public health. These policies are often area-based and typically involve investments in key socioeconomic determinants of health (such as income, employment, and housing conditions) combined with efforts to create a diverse socioeconomic mix of residents and initiatives to promote healthy life styles.

Due to the inherently small populations and typically limited health information for individuals within each area, existing (traditional) methodology to measure summary measures of population health cannot be used for estimations at the neighborhood or even district-level. This makes it difficult to measure, monitor, and reliably evaluate the impact of area-based programs and other area-based interventions in terms of population health.

In this thesis, new methodology for the estimation of small-area indicators of population health has been developed, making use of Bayesian random-effects specifications that can pool strength and make use of correlations between the various dimensions of the required mortality and morbidity data; for example, between sexes, age groups, and contiguous areas. This methodology constitutes a new class of small-area estimations that forms a welcome addition to the existing small-area estimation literature due to its excellent accommodation of sparse data problems while being solely based on the observed small-area data, i.e. without reliance on area-level covariates or assumptions about the transferability of correlations between aggregate-level data to the small-area level.

The performance of the Bayesian random-effects models has been compared with that of the existing (traditional) methodology using Monte Carlo simulations. In the Monte Carlo simulations, a novel benchmark based on a large number of European countries was used. Essentially, the included European countries are shrunk until Europe is represented as a city, with the European countries as its neighborhoods. Based on the large differences between the European countries this benchmark represents a "worst-case" scenario for the Bayesian random-effects methodology, which pools strength from similarities between the areas under analysis. The latter does not apply for the traditional methodology, whose performance is independent from differences and similarities between the included areas.

An important conclusion of the Monte Carlo simulations is that the traditional methodology to estimate summary measures of population health cannot be used for populations smaller than approximately 5,000 person years at risk and survey sample sizes smaller than approximately 100 persons per life table. In the Dutch context, this implies that the traditional methodology is not suitable for the estimation of life expectancy measures at the neighborhood level. Even estimations at the district-level (i.e. "wijk-level") should be conducted with careful consideration of the established minimum population and survey sample sizes. Moreover, the standard approach to calculate standard errors and 95% confidence intervals is found to be unreliable for populations smaller than approximately 25,000 person years at risk. For these population sizes an alternative approach should be used that does not rely on asymptotic normality assumptions.

Another important conclusion of the Monte Carlo simulations is that the Bayesian random-effects models outperform the traditional approach in sparse data conditions. As shown, the proposed Bayesian random-effects models can be used to reliably estimate standard life expectancy (LE) for populations as small as 2,000 person years at risk, and healthy life expectancy (HLE) for populations as small as 1,750 person years at risk when combined with a minimum survey sample size of approximately 45 persons per life table. Moreover, the proposed Bayesian random-effects approaches are shown to produce unbiased estimates and reliable 95% credible intervals for all recommended population and survey sample sizes. In the Dutch context, this implies that the Bayesian random-effects approach can be used for the estimation of life expectancy measures at the district as well as the neighborhood level.

The Monte Carlo simulations conducted in this thesis did not consider the geographic location of nursing homes as a potential confounder. At the national level, where each country has many nursing homes and few people migrate to another country just prior to their time of death, this makes sense. At the small-area level, however, only a minority of neighborhoods has a nursing home and this induces a migration of frail elderly from neighborhoods that do not have a nursing home to neighborhoods that do. The latter can strongly distort small-area health comparisons. As shown in this thesis, neighborhoods with a nursing home can have uncorrected life expectancies that are up to six years too low due to the selective migration of frail individuals to nursing homes. Conversely, neighborhoods without nursing homes can have uncorrected life expectancies that are up to five years too high. Compared to the range of nursing-home corrected life expectancies, which is approximately 8 years, this impact is large. Accordingly, an important finding of this thesis is that a nursing home correction is

required if the small-area summary measures are to be interpreted as indicators of average population health.

The ideal correction for the migration of frail elderly to nursing homes is to re-assign all nursing home residents to their previous residential address prior to the calculation of the summary indicators of population health. This correction, however, requires detailed individual-level and longitudinal data, which are often unavailable. Fortunately, one of the advantages of the Bayesian random-effects approach is that it is easily extensible to incorporate a correction for the spatial distribution of nursing homes as an integral part of the life table estimations. As shown in this thesis, such a Bayesian model-based correction provides a reliable alternative to the ideal nursing home correction and can do so with relatively modest data requirements: only the (cross-sectional) percentage of nursing home deaths per area is required.

Another aspect when using the random-effects methodology to estimate small-area summary measures of population health for real-life applications is that, when the required data are available, health-adjusted life expectancy may constitute a more reliable and suitable outcome measure. Unlike standard LE, it combines both mortality and morbidity information, and in comparison to other health expectancy measures, HALE is more sensitive to changes in the severity of morbidity within a population. Accordingly, in this thesis a Bayesian random-effects approach to estimating HALE at the small-area level has also been developed. The modeling approach is an extension of the models that were evaluated in the Monte Carlo simulations and its application is illustrated in a real-life example, which confirmed the feasibility and reliability of the HALE estimations for the city of Rotterdam. Additionally, it was demonstrated how the information could be graphically presented and how the full uncertainty of the estimates can be taken into account when creating rankings of the area's outcome measures. The latter is particularly useful for the selection of areas for policy purposes, such as, for example, area based programs that only target a small number of disadvantaged neighborhoods.

The final part of this thesis was aimed at investigating spatial variation in small-area population health. Based on the developed Bayesian random-effects methodology, one example of such an investigation was a follow-up regression in which the impact of three different measures of urban green on two different measures of small-area population health (i.e. small-area LE and HLE) has been estimated. In contrast to previous epidemiological studies, the statistical methodology simultaneously accounted for spatially correlated unobserved determinants, measurement uncertainty,

and the migration of frail elderly to nursing homes. Adjusted for differences in average household income, it was found that the quantity and quality of urban green space are modestly related to small-area LE and HLE: an increase of 1 standard deviation in the percentage of urban green is associated with a 0.1-year higher LE, and, in the case of the perceived quality of green, with an approximately 0.3-year higher LE and HLE. In concordance with a growing body of evidence, it was concluded that urban green development seems to deserve a more prominent place in urban planning, with the results indicating that not only the quantity but also the quality of urban green should be considered in future interventions.

At the same time, despite their advantages, the use of summary measures of population health in explanatory research was found be of mixed value because the observed associations cannot be interpreted as individual-level associations. Of course, the latter is a general problem with statistical analyses based on ecological (aggregated individual-level) indicators, but it is an even more articulate problem for summary measures that are by definition only defined for populations as a whole. Hence a different approach was used in this thesis to investigate the contribution of behavioral risk factors to small-area differences in mortality from cardiovascular disease (CVD). That is, a Hierarchically Related Regression approach was implemented that combines the benefits of both an ecological analysis (in terms of data availability and statistical power) and an individual-level analysis (in terms of identification of the parameters and interpretation of the results), making use of similar methodological solutions as used in the Bayesian random-effects estimations to accommodate sparse data problems at the small-area level.

After correcting for differences in age and sex, it was found that individuals who reside in the poorest neighborhoods have on average a 66% and 34% higher risk of CVD mortality than those who reside in the richest neighborhoods (for men and women, respectively). Furthermore, accounting for differences in smoking, excessive drinking, alcohol abstinence, obesity, and physical inactivity reduced the income-related inequality in CVD mortality by approximately 30%. Finally, given the geographic clustering of unhealthy risk factor behavior, area-based targeting of deprived neighborhoods was confirmed to be a relevant and potentially effective approach to address income-related inequalities in CVD mortality. Particularly smoking, physical activity, and overweight are identified as interesting candidates for future interventions, as they have the largest impact on CVD mortality and exhibit substantial prevalence differences across the income quintiles.

10

Samenvatting

10.1 SAMENVATTING

Stedelijk beleid is voor overheden en gezondheidsautoriteiten een belangrijk instrument om verbeteringen in leefomstandigheden, economische mogelijkheden en volksgezondheid te realiseren. Dit beleid is vaak gebiedsgericht en omvat investeringen in de belangrijkste sociaal-economische determinanten van gezondheid (zoals inkomen, werkgelegenheid en huisvesting) in combinatie met zowel inspanningen om een diverse sociaal-economische mix van bewoners te bewerkstelligen als initiatieven om een gezonde leefstijl te bevorderen.

Vanwege de inherent kleine populaties en beperkte gezondheidsinformatie (van individuen) binnen wijken en buurten levert de bestaande, traditionele methodologie om indicatoren van populatiegezondheid te meten onvoldoende betrouwbare statistische schattingen op wijk- en buurtniveau op. Dit maakt het moeilijk om de impact van gebiedsgerichte programma's en andere gebiedsgerichte interventies te meten, te monitoren en te evalueren.

In dit proefschrift is een nieuwe methode voor het schatten van indicatoren van populatiegezondheid ontwikkeld waarbij gebruik gemaakt wordt van Bayesiaanse random-effects specificaties. Deze specificaties maken gebruik van correlaties in de data om tot betere schattingen te komen; denk hierbij aan correlaties tussen geslachten, leeftijdsgroepen, en aan elkaar grenzende gebieden. De ontwikkelde methode vormt een nieuwe groep (of klasse) van kleine-gebiedsschattingen die een welkome aanvulling vormt op de bestaande literatuur vanwege de uitstekende accommodatie van schaarse data problemen. Een belangrijk aspect hierbij is dat alleen gebruik gemaakt wordt van de beschikbare data in elk gebied, zonder gebruik te maken van predicties met behulp van verklarende variabelen of van veronderstellingen over de overdraagbaarheid van correlaties op geaggregeerd niveau naar gezondheidsuitkomsten in kleine geografische gebieden.

De prestaties van de Bayesiaanse random-effects modellen zijn in dit proefschrift vergeleken met die van de bestaande (traditionele) methode via Monte Carlo simulaties. Deze simulaties zijn gebaseerd op een nieuw ontwikkelde benchmark op basis van de mortaliteits- en morbiditeitsdata van een groot aantal Europese landen. In essentie worden de Europese landen verkleind totdat Europa de omvang heeft van een stad, met de Europese landen als buurten, waarvan de leeftijdsopbouw en gezondheid exact bekend zijn. Gebaseerd op de enorme verschillen tussen de Europese landen vormt deze benchmark een "worst-case" scenario voor de Bayesiaanse random-effects metho-

dologie, die kracht put uit overeenkomsten tussen de gebieden. Derhalve kunnen de prestaties van de Bayesiaanse random-effects modellen in de Monte Carlo simulaties worden geïnterpreteerd als een ondergrens van de werkelijke prestaties; immers, in 'real-life' toepassingen zullen de gezondheidsverschillen kleiner zijn dan in de benchmark en zullen de Bayesiaanse random-effects modellen nog beter presteren dan in de simulaties. Voor de traditionele methode, waarvan de uitkomsten onafhankelijk zijn van verschillen en overeenkomsten tussen gebieden, geldt dit niet.

Een belangrijke conclusie die op basis van de Monte Carlo simulaties getrokken kan worden is dat de traditionele methode om indicatoren van populatiegezondheid te meten onvoldoende betrouwbaar is voor populaties kleiner dan circa 5.000 persoonsjaren 'at risk' en voor enquête steekproeven kleiner dan circa 100 personen per geslacht. In de Nederlandse context betekent dit dat de traditionele methode ongeschikt is voor het schatten van levensverwachtingen op buurtniveau. Zelfs schattingen op wijkniveau zijn problematisch en moeten worden uitgevoerd met zorgvuldige inachtname van de vastgestelde minimale bevolking- en steekproefomvang. Bovendien is gebleken dat de standaard methode om standaardfouten en 95% betrouwbaarheidsintervallen te berekenen onbetrouwbaar is voor populaties kleiner dan circa 25.000 persoonsjaren. Voor deze populatiegroottes zou een alternatieve benadering gebruikt moeten worden die niet gebaseerd is op de aanname van asymptotische normaliteit.

Een andere belangrijke conclusie van de Monte Carlo simulaties is dat de Bayesiaanse random-effects modellen aanzienlijk beter presteren dan de traditionele methode in combinatie met schaarse (cq. kleine populaties en kleine steekproef) data. De simulaties laten zien dat de Bayesiaanse random-effects modellen gebruikt kunnen worden om een betrouwbare schatting van de standaard levensverwachting (LV) te verkrijgen bij populaties vanaf circa 2,000 persoonsjaren en van gezonde levensverwachtingen (GLV) bij populaties vanaf circa 1,750 persoonsjaren in combinatie met minimale steekproefgroottes van circa 45 personen per geslacht. Bovendien produceren de Bayesiaanse random-effects benaderingen zuivere schattingen en betrouwbare 95% waarschijnlijkheidsintervallen voor alle geschikte populatie- en steekproefomvangen. In de Nederlandse context betekent dit dat de Bayesiaanse random-effects benaderingen gebruikt kunnen worden voor het schatten van levensverwachtingen op zowel wijk als buurtniveau.

In de Monte Carlo simulaties is de geografische locatie van verpleeg- en verzorgingshuizen niet meegenomen als een potentiële confounder. Dit was ook niet nodig, want op internationaal niveau heeft elk land zijn eigen verpleeg- en verzorgingshuizen en

zullen relatief weinig mensen in hun laatste levensjaren emigreren naar een ander land. Echter, op klein geografisch niveau heeft slechts een beperkt aantal buurten deze faciliteiten, wat leidt tot een migratie van kwetsbare ouderen vlak voor hun overlijden vanuit buurten zonder naar buurten met verpleeg- en verzorgingshuizen. Dit kan de vergelijkbaarheid van populatiegezondheidsindicatoren sterk vertekenen. Zoals aangetoond in dit proefschrift kunnen buurten met verpleeg- en verzorgingshuizen ongecorrigeerde levensverwachtingen hebben die tot 6 jaar te laag zijn. Andersom kunnen buurten zonder verpleeg- en verzorzingshuizen ongecorrigeerde levensverwachtingen hebben die tot 5 jaar te hoog zijn. In vergelijking met de totale range van verzorgingshuisgecorrigeerde levensverwachtingen van ongeveer 8 jaar is deze vertekening groot. Vandaar dat een belangrijke aanbeveling van dit proefschrift is om verzorgingshuis-gecorrigeerde indicatoren te gebruiken wanneer deze interpreteerbaar moeten zijn als samenvattende indicatoren van volksgezondheid (i.e. van de gehele onderliggende populatie).

De ideale correctie voor de migratie van kwetsbare ouderen naar verpleeg- en verzorgingshuizen is om alle bewoners fictief toe te wijzen aan hun laatst bekende residentiele adres alvorens de benodigde data per buurt te aggregeren en indicatoren van populatiegezondheid te berekenen. Deze correctie vereist echter uitgebreide longitudinale gegevens op individueel niveau die vaak niet beschikbaar zijn. Een van de voordelen van de ontwikkelde Bayesiaanse random-effects benadering is echter dat deze eenvoudig uit te breiden is met een verpleeg- en verzorgingshuiscorrectie als integraal onderdeel van de schattingen. Zoals beschreven in dit proefschrift biedt zo'n modelmatige correctie een betrouwbaar alternatief voor de ideale correctie, en kan dit op basis van relatief eenvoudig beschikbare data: alleen de (cross-sectionele) percentages van sterfgevallen in verpleeg- en verzorgingshuizen per buurt zijn nodig.

Een belangrijk aspect bij gebruik van de random-effects methodologie voor het schatten van samenvattende indicatoren van populatiegezondheid in real-life toepassingen is dat, mits de benodigde gegevens beschikbaar zijn, de gezondheidsgecorrigeerde levensverwachting (GCLV) een betrouwbaardere en geschiktere uitkomstmaat zal zijn. In tegenstelling tot de standaard levensverwachting combineert deze maatstaf mortaliteit en morbiditeitsgevens in één maatstaf, en in vergelijking met andere gezondheidsverwachtingen is de GCLV gevoeliger voor veranderingen in de ernst van de ziektelast binnen populaties. Vandaar dat in dit proefschrift ook methodologie ontwikkeld is om GCLV te berekenen voor kleine geografische gebieden. De gebruikte methode is een uitbreiding van de modellen die eerder al werden geëvalueerd zijn in de Monte Carlo simulaties en de toepassing ervan is geïllustreerd in een real-life

voorbeeld waarin de haalbaarheid en de betrouwbaarheid van de GCLV schattingen voor de stad Rotterdam uiteengezet zijn. Tevens is gedemonstreerd hoe de geschatte indicatoren grafisch gepresenteerd kunnen worden en hoe de volledige onzekerheid van de schattingen meegenomen kan worden bij het opstellen van ranglijsten van de uitkomstmaten per wijk of buurt. Dit laatste is nuttig voor de selectie van gebieden voor beleidsdoeleinden, zoals, bijvoorbeeld, gebiedsgerichte programma's die zich specifiek richten op een beperkt aantal achterstandsbuurten.

Het laatste deel van dit proefschrift is gericht op het onderzoeken van ruimtelijke variatie in gezondheid op klein geografisch niveau. Op basis van de ontwikkelde Bayesiaanse random-effects methodologie is een follow-up regressie gebruikt om de impact van drie verschillende maatstaven van stedelijk groen op zowel de standard levensverwachting (LV) als gezonde levensverwachting (GLV) te bepalen. In tegenstelling tot eerdere epidemiologische studies corrigeert de gebruikte statistische methodologie voor ruimtelijk gecorreleerde confounders, meetonzekerheid en de migratie van kwetsbare ouderen naar verpleeg- en verzorgingshuizen. Gecorrigeerd voor verschillen in het gemiddelde huishoudinkomen blijkt de hoeveelheid en de kwaliteit van stedelijke groen gerelateerd te zijn aan LV en GLV op buurtniveau: een stijging van 1 standaarddeviatie van het percentage van de stedelijke groen wordt geassocieerd met een 0.1 jaar hogere LV en, in het geval van kwaliteit van groen, een 0.3 jaar hogere LV en GLV. Deze resultaten ondersteunen de bredere conclusie in dit proefschrift dat het gebruik van groen een prominentere plaats verdient in de stedelijke planning, waarbij de gepresenteerde resultaten laten zien dat niet alleen de kwantiteit maar ook de kwaliteit van het stedelijk groen meegenomen zou moeten worden in toekomstige interventies.

Ondanks de voordelen van de ontwikkelde indicatoren van populatiegezondheid in beschrijvend onderzoek, bleek hun toepassing in verklarend onderzoek relatief lastig. Dit komt omdat de statistische schattingen niet geïnterpreteerd konden worden als associaties op individueel niveau. Enerzijds is dit een algemeen probleem van statistische analyses op basis van ecologische data, anderzijds komt dit probleem extra sterk naar voren bij analyses op basis van indicatoren van populatiegezondheid welke per definitie alleen gedefinieerd zijn voor populaties in z'n geheel. Vandaar dat in dit proefschrift een Hierarchisch Gerelateerde Regressie (HGR) benadering gebruikt is om de bijdrage van gedragsgerelateerde risicofactoren op geografische verschillen in sterfte aan hart- en vaatziekten te onderzoeken. Deze methodode combineert de voordelen van een ecologische analyse (in termen van databeschikbaarheid en statistische power) en een analyse op individueel niveau (in termen van identificatie van de parameters en interpretatie van de resultaten). Bovendien is bij de toepassing van de HGR methode

gebruik gemaakt van vergelijkbare methodologische oplossingen voor schaarse data-problemen zoals eerder in dit proefschrift toegepast bij de Bayesiaanse random-effects modellen, wat zorgt voor een zekere mate van methodologische consistentie.

Gecorrigeerd voor leeftijd en geslacht bleek dat mensen die in de armste buurten wonen gemiddeld een 66% en 34% hoger risico op sterfte aan hart- en vaatziekten hebben dan degenen die in de rijkste wijken wonen (voor mannen en vrouwen, respectievelijk). Wanneer additioneel gecorrigeerd wordt voor verschillen in vijf gedragsgerelateerde risicofactoren (i.e. roken, overmatig drinken, alcohol onthouding, obesitas en lichamelijke inactiviteit) verminderde de inkomensafhankelijke ongelijkheid in sterfte aan hart- en vaatziekten met ongeveer 30%. Tevens bleek er een sterke geografische clustering van ongezond gedrag in de buurten met een lager gemiddeld inkomen te zijn. Dit betekent dat een gebiedsgerichte aanpak van ongezond gedrag potentieel een effectieve aanpak kan zijn om de aangetoonde inkomensongelijkheid in sterfte aan hart-en vaatziekten te reduceren. Vooral roken, lichamelijke activiteit en overgewicht vormen hierbij interessante handvatten voor toekomstige interventies, omdat zij de grootste impact op sterfte aan hart- en vaatziekten hebben en juist deze risicofactoren aanzienlijke verschillen in prevalenties tussen de laagste en hoogste inkomensgroepen laten zien.

11

Curriculum Vitae

Marcel Jonker (1982, Voorburg, The Netherlands) holds a propedeuse in Ecomomics (Cum Laude, 2001) from the University of Amsterdam, a B.Sc. in Social Science (Summa Cum Laude, 2004) from University College Utrecht, and a M.Sc. in International Economics and Finance (Cum Laude, 2007) and Economics (Cum Laude, 2008) from Tilburg University.

In 2006, during his graduate studies, Marcel started working at the Economics & Valuation group of PricewaterhouseCoopers (Amsterdam, The Netherlands) where he worked as a quantitative researcher and econometric consultant on many different projects, ranging from optimal pricing and market entry analyses to discounted cash flow valuations and mathematical simulations of regulatory frameworks.

In 2010, Marcel left PricewaterhouseCoopers to pursue a Ph.D. in Public Health at the Erasmus Medical Center (Rotterdam, The Netherlands), where he could focus on a project that involved the quantification and explanation of variation in small-area population health. In 2013, Marcel was awarded the Erasmus MC public health department's academic stimulation award for his first peer-reviewed scientific article, which was published in the American Journal of Epidemiology.

During his Ph.D. Marcel also worked on the adaptation of discrete choice experiments (DCE) for health-state valuations, supported by the Netherlands Scientific Organization (NWO) and the EuroQol Group. While finalizing his dissertation in 2015, Marcel continued his work on methodological improvements in DCE for health-state valuations at the Institute of Health Policy and Management (iBMG) and the Erasmus Choice Modeling Center (ECMC).

12

List of publications

List of publications

Jonker, M.F., F.J. van Lenthe, P.D. Congdon, B. Donkers, A. Burdorf, and J.P. Mackenbach. 2012. "Comparison of Bayesian random-effects and traditional life expectancy estimations in small-area applications." American journal of epidemiology 176(10):929-937.

Jonker, M.F., F.J. van Lenthe, B. Donkers, P.D. Congdon, A. Burdorf, and J.P. Mackenbach. 2013. "The impact of nursing homes on small-area life expectancies." Health & Place:1-8..

Jonker, M.F., P.D. Congdon, F.J. van Lenthe, B. Donkers, A. Burdorf, and J.P. Mackenbach. 2013. "Small-area health comparisons using health-adjusted life expectancies - A Bayesian random-effects approach." Health & Place:1-9.

Jonker, M.F., F.J.van Lenthe, B. Donkers, J.P. Mackenbach, A. Burdorf. 2014. "The effect of urban green on small-area (healthy) life expectancy." Journal of Epidemiology and Community Health; 68:999–1002

Jonker, M.F., B. Donkers, B. Chaix, F.J. van Lenthe, A. Burdorf, J.P. Mackenbach. 2015."Estimating the impact of health-related behaviors on geographic variation in cardiovascular mortality; a new approach based on the synthesis of ecological and individual-level data." Epidemiology; Vol.26.No.6:888-897

Jonker, M.F., B. Donkers, F.J. van Lenthe, F. Peters, A. Burdorf, J.P. Mackenbach. "Estimating healthy life expectancy for small geographic areas." (submitted)

Jonker, M.F., E. d'Ippolito, T. Eikemo, P.D. Congdon, N. Nante, J.P. Mackenbach. "The effect of regional politics on population health in Italy (1980-2010)." (submitted)

Jonker, M.F., B. Donkers, M. Bliemer, E.A. Stolk, E. de Bekker-Grob. "Estimating SF-6D health state utility values using an improved and efficient discrete choice experiment." (submitted)

Jonker, M.F. A.E. Attema, E.A. Stolk, M.M. Versteegh. "Are health-state valuations from the general public biased? A test for health-state reference dependency in the EQ-5D-5L using a discrete choice experiment." (submitted)

Willers, S.M., **M.F. Jonker,** L. Klok, M. Keuken, J. Odink, W.J. Okkerse, S. van den Elshout, J.P. Mackenbach; A. Burdorf. 2016. "High resolution exposure modelling of heat and air pollution and the impact on mortality." Environment International 89: 102-109.

De Bekker-Grob, E.W., B. Donkers, **M.F. Jonker,** E.A. Stolk. 2015. "Sample size requirements for discrete-choice experiments in healthcare: a practical guide." The Patient-Patient-Centered Outcomes Research, 1-12.

13

PhD Portfolio

PhD candidate:	Marcel Jonker
Department:	Public Health
Promotors:	Prof.dr. J.P. Mackenbach
	Prof.dr. A. Burdorf
Co-promotor:	Dr. F.J. van Lenthe
PhD period:	2010-2015

Courses	2010	Constructing Composite Indicators: Theoretical and Practical Aspects (EC Joint Research Centre, Ispra, Italy)
	2010	Social Epidemiology (EUR, Rotterdam, The Netherlands)
	2010	Introduction to Bayesian Modelling using WinBUGS (Imperial College, London, United Kingdom)
	2010	Bayesian Hierarchical Modelling using WinBUGS (Imperial College, London, United Kingdom)
	2013	Bayesian Inference for Latent Gaussian Models using INLA (RIVM, Bilthoven, The Netherlands by prof.dr. Havard Rue)
Seminars	2010—2015	MGZ public health, biostatistics, marketing, and ClubMeth meetings
Workvisits	2010	Prof.dr. Peter Congdon (Queen Mary University of Londen, UK)
	2013	Prof.dr. Havard Rue (Norwegian University of Science and Technology, Trontheim, Norway - incl. oral presentation)
	2014	Dr. Andrew Thomas (Cambridge, UK - OpenBUGS programmer)
Presentations	2011	GGD (oral presentation)
	2012	Cephir (oral presentation)
	2013	MGZ seminar (oral presentation)
	2011, 2014	ClubMeth (oral presentations)
	2010, 2014	Social Epidemiology (oral presentations)
Conferences	2008	WEON Preconference (oral presentaton)
	2012, 2015	EuroQol Conference (oral presentations)
	2014	Young Demographers Conference (oral presentation)
	2015	IAPHR (oral presentation)
	2015	SMDM (oral presentation)
Research support	2010—2014	GGD life expectancy calculations
	2010—2011	GGD health monitor
	2013—2014	DCMR thermal comfort & air pollution
Teaching	2011—2013	MGZ 3rd year research group supervisor (5x)
	2012—2015	Course: Patient Preferences in the Delivary of Health Care (3x)
	2013	Course: Discrete Choice Experiments for Quality of Life Evaluations
	2015	Course: Statistical Methods

14

Dankwoord

Met veel plezier kijk ik terug op de afgelopen jaren waarin ik het voorrecht heb gehad om op de afdeling Maatschappelijke Gezondheidszorg van het Erasmus MC te hebben mogen werken.

Allereerst wil ik mijn promotoren Johan Mackenbach en Lex Burdorf en mijn co-promotor Frank van Lenthe bedanken voor hun begeleiding en inzet om tot dit mooie resultaat te komen. Frank, je hebt alle belangrijke beslissingen in dit onderzoek begeleid en je constante interesse en deskundigheid heb ik enorm gewaardeerd. Lex, jouw aandacht voor de kern van het verhaal en de manier waarop je mij telkens wist uit te dagen om de belangrijkste boodschappen van ons onderzoek nog krachtiger te formuleren hebben enorme indruk en waardering achtergelaten. Johan, diep respect hoe je feilloos de pijnpunten in elke tekst, analyse en/of onderzoeksopzet weet te vinden en richting hebt weten te geven aan dit lastige onderzoek. Heel erg bedankt!

Voor statistische ondersteuning ben ik grote dank verschuldigd aan mijn statistische begeleiders Peter Congdon en Bas Donkers. Hun inbreng heeft mijn promotieonderzoek absoluut tot een hoger niveau gebracht. Peter, mijn dank voor je creatieve input, kritische vragen en statistische ondersteuning in een gebied waarin jij excelleert. Bas, je hebt met grenzeloos enthousiasme geholpen om waanzinnig complexe problemen op te lossen. Bedankt voor deze meer dan geweldige begeleiding.

Ook op ICT-gebied was mijn onderzoek uitdagend. Mijn dank gaat uit naar vele mensen die meegeholpen hebben om de simulaties en statistische berekeningen zowel intern als extern te faciliteren. Met name Kees Noordsij-Wagenaar en collega's van MGZ, Lykle Voort en collega's van SURF/SARA, Gea Schouten en collega's van de GGD Rotterdam-Rijnmond en de medewerkers van het Centrum voor Beleidsstatistiek van het CBS wil ik graag speciaal bedanken.

Iets minder direct maar desondanks niet minder belangrijk is de voorbereiding bij PricewaterhouseCoopers geweest. Deel uitmaken van zo'n geweldige groep mensen, met wisselende teams, uitdagende opdrachten en groeiende verantwoordelijkheden was een fantatische ervaring. De grote hoeveelheid mensen die nu nog als inspiratiebron fungeert zegt in dat opzicht genoeg! Gezien de lange lijst zonder naam en toenaam, maar heel veel dank aan alle voormalige collega's bij PwC.

Tenslotte wil ik natuurlijk al mijn familie, vrienden, kennissen en collega's (zowel bij het Erasmus MC als de Erasmus Universiteit) enorm bedanken. Met name dankzij hun inbreng en steun waren de afgelopen jaren zo geweldig. Dank ook aan mijn paranimfen (Nanda en Sonja) en in het bijzonder mijn vriendin Sjoanne voor alle steun en motivatie bij de totstandkoming van dit proefschrift.

15

References

[1] Thomson, H. (2008) "A dose of realism for healthy urban policy: lessons from area-based initiatives in the UK." *Journal of Epidemiology and Community Health* 62.10: 932-936.

[2] Tunstall, R. and Lupton, R. (2003) "Is targeting deprived areas an effective means to reach poor people? An assessment of one rationale for area-based funding programmes." *An Assessment of One Rationale for Area-Based Funding Programmes.* LSE Research Paper No. CASE070.

[3] Van Ham, M., Manley, D., Bailey, N., Simpson, L., and Maclennan, D. (2012). "Neighbourhood Effects Research: New perspectives" *Springer Netherlands*: 1-21.

[4] Sharpe, E.K. (2013) "Targeted neighbourhood social policy: a critical analysis." *Journal of Policy Research in Tourism, Leisure and Events* 5.2: 158-171.

[5] Maclennan, Duncan. (2013) "Neighbourhoods: Evolving ideas, evidence and changing policies." *Neighbourhood Effects or Neighbourhood Based Problems?.* Springer Netherlands: 269-292.

[6] McLoone, P. (2001) "Targeting deprived areas within small areas in Scotland: population study." *BMJ* 323.7309: 374-375.

[7] Séguin, A.M., Apparicio, P. and Riva, M. (2012) "The impact of geographical scale in identifying areas as possible sites for area-based interventions to tackle poverty: the case of Montréal." *Applied Spatial Analysis and Policy* 5.3: 231-251.

[8] Cheshire, P. (2007) "Are Mixed Communities the Answer to Segregation and Poverty?" *The Joseph Rowntree Foundation* 17.

[9] Manley, D., van Ham, M., Bailey, N., Simpson, L., and Maclennan, D. (2013). "Neighbourhood effects or neighbourhood based problems. A policy context." *Springer Netherlands.*

[10] Oakes, M.J. (2004) "The (mis) estimation of neighborhood effects: causal inference for a practicable social epidemiology." *Social science & medicine* 58.10: 1929-1952.

[11] Van Ham, M. and Manley, D. (2009) "The effect of neighbourhood housing tenure mix on labour market outcomes: a longitudinal investigation of neighbourhood effects." *Journal of Economic Geography*: lbp017.

[12] Atkinson R. (2008) "European urban policies and the neighbourhood: an overview." *Urban Design and Planning* 161:115-122

[13] Rothenberg, R., Stauber, C., Weaver, S., Dai, D., Prasad, A., and Kano, M. (2015) "Urban health indicators and indices current status." *BMC public health* 15(1): 494.

[14] Kramer, D., Jongeneel-Grimen, B., Stronks, K., Droomers, M., and Kunst, A.E. (2015) "Are area-based initiatives able to improve area safety in deprived areas? A quasi-experimental evaluation of the Dutch District Approach." *BMC public health* 15(1): 711.

[15] Stronks, K., Droomers, M., Jongeneel-Grimen, B., Kramer, D., Hoefnagels, C., van Oers, H., and Kunst, A.E. (2014) "Health impact of area-based interventions in Dutch deprived neighbourhoods: the URBAN40 study." *The European Journal of Public Health* 24: 164-044.

[16] Walthery, P., Stafford, M., Nazroo, J., et al. (2015) "Health trajectories in regeneration areas in England: the impact of the New Deal for Communities intervention." *Journal of Epidemiology and Community Health* 69(8): 762-768.

[17] Mackenbach, J.P. (2003) "Tackling inequalities in health: the need for building a systematic evidence base." *Journal of Epidemiology and Community Health* 57.3: 162-162.

[18] Mackenbach, J.P. (2011) "Can we reduce health inequalities? An analysis of the English strategy (1997–2010)." *Journal of epidemiology and community health*

[19] Mackenbach, J.P. and Karien Stronks. (2002) "A strategy for tackling health inequalities in the Netherlands." *BMJ* 325.7371: 1029-1032.

[20] Jackson, C., Best, N., and Richardson, S. (2006) "Improving ecological inference using individual-level data." *Statistics in Medicine* 25(12): 2136-2159.

[21] Jackson, C., and Richardson, S. (2008) "Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171.1: 159-178.

[22] Stafford, M., Badland, H., Nazroo, et al. (2014). "Evaluating the health inequalities impact of area-based initiatives across the socioeconomic spectrum: a controlled intervention study of the New Deal for Communities, 2002–2008." *Journal of Epidemiology and Community Health* 68(10): 979-986.

[23] Magnée, T., Burdorf, A., Brug, J., et al. (2013). "Equity-specific effects of 26 Dutch obesity-related lifestyle interventions." *American journal of preventive medicine* 44(6): 61-70.

[24] Murray, C., Salomon, J.A., and Mathers, C.D. (2000) "A critical examination of summary measures of population health." *Bulletin of the World Health Organization* 78.8: 981-994.

[25] Mathers, C.D., Salomon, J.A., Murray, C.J., Lopez, A., Murray, C., and Evans, D. (2003). "Alternative summary measures of average population health." *Health systems performance assessment: debates, methods and empiricism* 319-334.

[26] Brooks, R, and EuroQol Group. (1996) "EuroQol: the current state of play." *Health policy* 37.1: 53-72.

[27] Herdman, M., Gudex, C., Lloyd, A., et al. (2011) "Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L)." *Quality of life research* 20.10: 1727-1736.

[28] Feeny, D., Furlong, W., Boyle, M. and Torrance, G.W. (1995) "Multi-Attribute Health Status Classification Systems: Health Utilities Index." *PharmacoEconomics* Vol.7, No.6: 490-502.

[29] Furlong, W., Feeny, D., Torrance, G., et al. (1998) "Multiplicative multi-attribute utility function for the Health Utilities Index Mark 3 (HUI3) system: a technical report." *Centre for Health Economics and Policy Analysis (CHEPA)* McMaster University, Hamilton, Canada.

[30] Brazier, J.E., Roberts, J. and Deverill, M. (2002) "The estimation of a preference-based measure of health from the SF-36." *Journal of health economics* 21.2: 271-292.

[31] Brazier, J.E. and Jennifer Roberts. (2004) "The estimation of a preference-based measure of health from the SF-12." *Medical care* 42.9: 851-859.

[32] Microdata services: conduct your own research usingdata from Statistics Netherlands (http://www.cbs.nl/en-GB/menu/informatie/beleid/zelf-onderzoeken/default.htm) Accessed September 2015.

[33] Schoen, R. (1978) "Calculating life tables by estimating Chiang's a from observed rates." *Demography* 15.4: 625-635.

[34] Golbeck, A.L. (1986) "Probabilistic approaches to current life table estimation." *The American Statistician* 40.3: 185-190.

[35] Chiang C.L. (1968) "The life table and its construction." *In: Introduction to Stochastic Processes in Biostatistics.* New York, NY: John Wiley & Sons, Inc; 189–214

[36] Scherbov S. and Ediev D. (2011) "Significance of life table estimates for small populations: simulation-based study of estimation errors." *Demographic Researc*h 24(22): 527–550.

[37] Eayres D. and Williams E.S. (2004) "Evaluation of methodologies for small area life expectancy estimation." *Journal of Epidemiology and Community Health* 58(3): 243–249

[38] Bayes, T. (1763) "An essay towards solving a Problem in the Doctrine of Chances." *Bayes's essay as published in the Philosophical Transactions of the Royal Society of London* Vol.53: 370

[39] Congdon, P. (2007) "Bayesian statistical modelling." *John Wiley & Sons.*

[40] Hall, B. (2012) "Bayesian Inference." CRAN R-project, LaplacesDemon Package, (http://cran.rproject.org/web/packages/LaplacesDemon/vignettes/Bayesian Inference.pdf) Accessed September 2015.

[41] Samaniego, F.J. (2010) "A comparison of the Bayesian and frequentist approaches to estimation." *Springer Science & Business Media.*

[42] Elliot, P., Wakefield, J.C., Best, N.G., and Briggs, D.J. (2000) "Spatial Epidemiology: Methods and Applications." *Oxford University Press.*

[43] Salomon, J.A., Wang, A., Freeman, M., et al. (2013) "Healthy life expectancy for 187 countries, 1990–2010: a systematic analysis for the Global Burden Disease Study 2010." *The Lancet* 380.9859: 2144-2162.

[44] Worldbank Development indicators, (http://wdi.worldbank.org/table/1.1) Accessed September 2015.

[45] Wakefield, J. (2004) "A critique of statistical aspects of ecological studies in spatial epidemiology." *Environmental and Ecological Statistics* 11.1: 31-54.

[46] Public Health Status and Foresight Report 2014 (Volksgezondheid Toekomstverkenning 2014). RIVM (http://www.eengezondernederland.nl/Heden_en_verleden/

Levensverwachting/Regionale_verschillen_in_gezonde_levensverwachting) Accessed September 2015.

[47] Congdon, P. (2007) "A model for spatial variations in life expectancy; mortality in Chinese regions in 2000." *International Journal of Health Geographics* 6.1: 16.

[48] Canadian Life tables, Provinces and Territories (2013) *Statistics Canada* No.5. (http://publications.gc.ca/pub?id=9.507323&sl=0) Accessed September 2015.

[49] Public Health Status and Foresight Report 2014 (Volksgezondheid Toekomstverkenning 2014). RIVM (http://www.eengezondernederland.nl/Heden_en_verleden/Levensverwachting) Accessed September 2015.

[50] Tsai, S.P. and Wen, C.P. (1986) "A review of methodological issues of the standardized mortality ratio (SMR) in occupational cohort studies." *International Journal of Epidemiology* 15(1):8–21.

[51] Julious, S.A., Nicholl, J. and George, S. (2001) "Why do we continue to use standardized mortality ratios for small area comparisons?" *Journal Public Health Medicine* 23(1):40–46.

[52] Silcocks, P.B., Jenner, D.A. and Reza, R. (2001) "Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities." *Journal of Epidemiology and Community Health* 55(1):38–43.

[53] Tsai, S.P., Hardy, R.J. and Wen, C.P. (1992) "The standardized mortality ratio and life expectancy." *American Journal of Epidemiology* 135(7):824–831.

[54] Lai, D., Hardy, R.J. and Tsai, S.P. (1996) "Statistical analysis of the standardized mortality ratio and life expectancy." *American Journal of Epidemiology* 143(8):832–840.

[55] Toson, B. and Baker, A. (2003) "Life Expectancy at Birth: Methodological Options for Small Populations." *Office for National Statistics* Norwich, United Kingdom.

[56] Congdon, P. (2004) "Modelling trends and inequality in small area mortality." *Journal of Applied Statistics* 31(6):603–622.

[57] Congdon, P. (2009). "Life Expectancies for Small Areas: A Bayesian Random-effects methodology." *International Statistical Review* 77(2):222-240.

[58] The Human Mortality Database. University of California, Berkeley, USA, and Max Planck Institute for Demographic Research, Rostock, Germany (http://www.mortality.org) Accessed April 2011.

[59] European Commission Populat (Eurostat population) database. Brussels, Belgium.
European Commission (http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics) Accessed April 2011.

[60] Williams, E.S., Dinsdale, H., Eayres, D., et al. (2005) "Technical Report: Calculating Life Expectancies in Small Areas." *South East Public Health Observatory* Oxford, United Kingdom.

[61] Gelman, A., Carlin, J., Stern, H., et al. (1995) "Bayesian Data Analysis." Chapman & Hall Ltd., London, United Kingdom.

[62] SARA Computing and Networking Services. Life Science Grid. Amsterdam, the Netherlands: SARA Computing and Networking Services; (http://www.sara.nl/project/life-science-grid) Accessed June–July 2011 and December 2011–January 2012.

[63] Congdon, P. (2008) "A spatially adaptive conditional autoregressive prior for area health data". *Statistical Methodology* 5(6):552–563.

[64] Williams, E.S., Dinsdale, H., Eayres. D., et al. (2004) "Impact of nursing home deaths on life expectancy calculations in small areas." *Journal of Epidemiology and Community Health* 58(11):958–962.

[65] Cai, L. and Lubitz, J. (2007) "Was there compression of disability for older Americans from 1992 to 2003" *Demography* 44(3):479-495.

[66] Chiang, C. (1984) "The Life Table and Its Applications." *Krieger Publishing* Malabar, Florida.

[67] Congdon, P. (2006) "A model for geographical variation in health and total life expectancy." *Demographic Research* 14(9):157-178.

[68] Imai, K. and Soneji, S. (2007) "On the estimation of disability-free life expectancy: Sullivan's method and its extension." *Journal of the American Statistical Association* 102(480):1199-1211.

[69] Jagger, C., Cox, B., Le Roy, S. and EHEMU. (2007) "Health expectancy calculation by the Sullivan method: A practical guide." *Euro-REVES* Montpelier, VT.

[70] Jagger, C. and Robine, J.M. (2011) "Healthy life expectancy." *In International Handbook of Adult Mortality* Springer Netherlands, 551-568.

[71] Jonker, M.F., Congdon, P.D., Van Lenthe, F.J., et al. (2013) "Small-area health comparisons using health-adjusted life expectancies - A Bayesian random-effects approach." *Health & Place* 1-9.

[72] Jonker, M.F., Van Lenthe, F.J., Congdon, P.D., et al. (2012). "Comparison of Bayesian Random-Effects and Traditional Life Expectancy Estimations in Small-Area Applications." *American Journal of Epidemiology* 176(10):929-937.

[73] Jonker, M.F., Van Lenthe, F.J., Donkers, B., et al. (2013) "The impact of nursing homes on small-area life expectancies." *Health & Place* 1-8.

[74] Kibele, E. and Janssen, F. (2013) "Distortion of regional old-age mortality due to late-life migration in the Netherlands." *Demographic Research* 29(5):105-132.

[75] Kulkarni, S.C., Levin-Rector, A., Ezzati, M. and Murray, C.J.L. (2011) "Falling behind: life expectancy in US counties from 2000 to 2007 in an international context." *Population Health Metrics* 9(1):16.

[76] Lynch, S.M. and Brown, J.S. (2010) "Obtaining multistate life table distributions for highly refined subpopulations from cross-sectional data: A bayesian extension of sullivan's method." *Demography* 47(4):1053-1077.

[77] Mathers, C.D. and Robine J.M. (1997) "How good is Sullivan's method for monitoring changes in population health expectancies?" *Journal of Epidemiology and Community Health* 51(1):80-86.

[78] Srebotnjak, T., Mokdad, A.H. and Murray, C.J. (2010) "A novel framework for validating and applying standardized small area measurement strategies." *Population Health Metrics* 8:26.

[79] Stiefel, M.C., Perla, R.J. and Zell, B.L. (2010) "A healthy bottom line: healthy life expectancy as an outcome measure for health improvement efforts." *Milbank Quarterly* 88(1):30-53.

[80] Sullivan, D.F. (1971) "A single index of mortality and morbidity." *HSMHA health reports* 86(4):347.

[81] Wang, H., Schumacher, A.E., Levitz, C.E., et al. (2013) "Left behind: widening disparities for males and females in US county life expectancy, 1985–2010." *Population health metrics* 11(1):8.

[82] Wood, R., Sutton, M., Clark, D., et al. (2006) "Measuring inequalities in health: the case for healthy life expectancy." *Journal of Epidemiology and Community Health* 60(12):1089-1092.

[83] Gandarillas, A., Dominguez-Berjon, M.F., Zorrilla, B., et al. (2011) "Deaths in collective dwellings and inequalities in small-area mortality: an ecological study in the Madrid region (Spain)." *Journal of Epidemiology and Community Health* 65: 310–314.

[84] Nimmo, A.W., Peterkin, G., and Coid, D.R. (2006) "Mortality in general practice: problems concerning deaths in institutions." *Public Health* 120: 671–675.

[85] Veugelers, P.J. and Hornibrook, S. (2002) "Small area comparisons of health: applications for policy makers and challenges for researchers." *Chronic Diseases in Canada* 23: 100–110.

[86] Arcaya, M., Brewster, M., Zigler, C.M., and Subramanian, S.V. (2012) "Area variations in health: a spatial multilevel modeling approach." *Health and Place* 18(4): 824–831.

[87] Bonneux, L.G., Huisman, C.C. and De Beer, J.A. (2010) "Mortality in 272 European regions, 2002–2004. An update." *European Journal of Epidemiology* 25: 77–85.

[88] Vaupel, J.W., Zhang, Z. and Van Raalte, A.A. (2011) "Life expectancy and disparity: an international comparison of life table data." *BMJ Open* 1(1): e000128.

[89] Brazier, J.E. and Roberts, J. (2004) "The estimation of a preference-based measure of health from the SF-12." *Medical Care* 42(9): 851–859.

[90] Camacho, N., Donkers, B. and Stremersch, S. (2011) "Predictably non-Bayesian: quantifying salience effects in physician learning about drug quality." *Marketing Science* 30(2): 305–320.

[91] Congdon, P. (2011) "The spatial pattern of suicide in the US in relation to deprivation, fragmentation and rurality." *Urban Studies* 48(10): 2101–2122.

[92] Ferrari, S. and Cribari-Neto, F. (2004) "Beta regression for modelling rates and proportions." *Journal of Applied Statistics* 31(7): 799–815.

[93] Fiebig, D.G., Keane, M.P., Louviere, J., and Wasi, N. (2010) "The generalized multinomial logit model: accounting for scale and coefficient heterogeneity." *Marketing Science* 29(3): 393–421.

[94] Gakidou, E.E., Murray, C.J. and Frenk, J. (2000) "Defining and measuring health inequality: an approach based on the distribution of health expectancy." *Bulletin of the World Health Organization* 78(1): 42–54.

[95] Gandek, B., Ware, J.E., Aaronson, et al. (1998) "Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project." *Journal of Clinical Epidemiology* 51(11): 1171–1178.

[96] Ware Jr., J.E., Kosinski, M., and Keller, S.D. (1996). "A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity." *Medical care* 34(3): 220-233.

[97] Heijink, R., Van Baal, P., Oppe, M., et al. (2011) "Decomposing cross-country differences in quality adjusted life expectancy: the impact of value sets." *Population Health Metrics* 9(1): 17.

[98] Jia, H., Zack, M.M., and Thompson, W.W. (2011) "State quality-adjusted life expectancy for U.S. adults from 1993–2008." *Quality of Life Research* 20(6): 853–863.

[99] Leon, D.A. (2011) "Trends in European life expectancy: a salutary view." *International Journal of Epidemiology* 40(2): 271–277.

[100] Manuel, D.G., Goel, V., Williams, J.I., and Corey, P. (2000) "Health-adjusted life expectancy at the local level in Ontario." *Chronic Diseases in Canada* 21(2): 73–80.

[101] Mari-Dell'Olmo, M., Martiez-Beneito, M.A., Borrell, C., et al. (2011) "Bayesian factor analysis to calculate a deprivation index and its uncertainty." *Epidemiology* 22(3): 356–364.

[102] Mathers, C.D., Iburg, K.M., Salomon, J.A., et al. (2004) "Global patterns of healthy life expectancy in the year 2002." *BMC Public Health* 4: 66.

[103] Mathers, C.D., Salomon, J.A., Murray, C.J.L. and Lopez, A.D. (2003) "Alternative summary measures of population health." In: Murray, C.J.L. and Evans, D.B. (Eds.), "Health Systems Performance Assessment: Debates, Methods and Empiricism." *World Health Organization* Geneva, 319–334.

[104] Mathers, C.D., Murray, C.J., Salomon, J.A., et al. (2003) "Healthy life expectancy: comparison of OECD countries in 2001." *Australian and New Zealand Journal of Public Health* 27(1): 5–11.

[105] Parkin, D., Rice, N. and Devlin, N. (2010) "Statistical analysis of EQ-5D profiles: does the use of value sets bias inference?" *Medical Decision Making* 30(5): 556–565.

[106] Paolino, P. (2001) "Maximum likelihood estimation of models with beta-distributed dependent variables." *Political Analysis* 9(4): 325–346.

[107] Rao, J.N.K. (2003) "Small Area Estimation." John Wiley and Sons, Inc., New Jersey.

[108] Rahman, A. (2008) "A Review of Small Area Estimation Problems and Methodological Developments." *NATSEM Discussion paper*, University of Canberra.

[109] Rahman, A., Harding, A., Tanton, R. and Liu, S. (2010) "Methodological issues in spatial microsimulation modelling for small area estimation." *International Journal of Microsimulation* 3(2): 3–22.

[110] Smithson, M. and Verkuilen, J. (2006) "A better lemon squeezer Maximum-likelihood regression with beta-distributed dependent variables." *Psychological Methods* 11(1): 54–71.

[111] Verkuilen, J. and Smithson, M. (2012) "Mixed and mixture regression models for continuous bounded responses using the beta distribution." *Journal of Educational and Behavioral Statistics* 37: 82–113.

[112] Toson, B. and Baker, A. (2003) "Life Expectancy at Birth: Methodological Options for Small Populations." *National Statistics Methodological Series 33.* Office for National Statistics, Norwich, UK.

[113] Giles-Corti, B., Broomhall, M.H., Knuiman, M., et al. (2005) "Increasing walking: how important is distance to, attractiveness, and size of public open space." *American Journal of Preventive Medicine* 28:169–176.

[114] Stigsdotter, U.K., Ekholm, O., Schipperijn, J., et al. (2010) "Health promoting outdoor environments—associations between green space, and health, health-related quality of life and stress based on a Danish national representative survey." *Scandinavian Journal of Public Health* 38:411–417.

[115] Maas, J., Verheij, R.A., De Vries, S., et al. (2009) "Morbidity is related to a green living environment." *Journal of Epidemiology and Community Health* 63:967–973.

[116] Maas, J., Verheij, R.A., Groenewegen, P., et al. (2006) "Green space, urbanity, and health: how strong is the relation?" *Journal of Epidemiology and Community Health* 60:587–592.

[117] Mitchell, R. and Popham, F. (2007) "Greenspace, urbanity and health: relationships in England." *Journal of Epidemiology and Community Health* 61:681–683.

[118] Mitchell, R., Astell-Burt, T. and Richardson, E.A. (2011) "A comparison of green space indicators for epidemiological research." *Journal of Epidemiology and Community Health* 65:853–858.

[119] Takano, T., Nakamura, K. and Watanabe, M. (2002) "Urban residential environments and senior citizens' longevity in megacity areas: the importance of walkable green spaces." *Journal of Epidemiology and Community Health* 56:913–918.

[120] Mitchell, R. and Popham, F. (2008) "Effect of exposure to natural environment on health inequalities: an observational population study." *Lancet* 372:1655–1660.

[121] Lee, A.C.K. and Maheswaran, R. (2011) "The health benefits of urban green spaces: a review of the evidence." *Journal of Public Health* 33:212–222.

[122] Silcocks, P.B.S., Jenner, D.A. and Reza R. (2001) "Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities." *Journal of Epidemiology and Community Health* 55:38–43.

[123] Julious, S.A., Nicholl, J. and George S. (2001) "Why do we continue to use standardized mortality ratios for small area comparisons?" *Journal of Public Health* 23:40–46.

[124] Statistics Netherlands, Nabijheid Voorzieningen. (http://statline.cbs.nl/StatWeb/) Downloaded June 2013.

[125] Plummer, M., Best, N., Cowles, K., et al. (2006) "CODA: convergence diagnosis and output analysis for MCMC." *R News* 6:7–11.

[126] Groenewegen, P.P., Van den Berg, A.E., Maas, J., et al. (2012) "Is a green residential environment better for health? If so, why?" *Annals of the Association of American Geographers* 102:996–1003.

[127] Pugh, T.A., MacKenzie, A.R., Whyatt, J.D., et al. (2012) "Effectiveness of green infrastructure for improvement of air quality in urban street canyons." *Environmental Science & Technology* 46:7692–7699.

[128] Tzoulas, K., Korpela, K., Venn, S., et al. (2007) "Promoting ecosystem and human health in urban areas using green infrastructure: a literature review." *Landscape and Urban Planning* 81:167–178.

[129] Jongeneel-Grimen, B., Droomers, M., Stronks, K., et al. (2011) "Migration does not enlarge inequalities in health between rich and poor neighbourhoods in the Netherlands." *Health & Place* 17:988–995.

[130] Jackson, C.H., Richardson, S. and Best, N.G. (2008) "Studying place effects on health by synthesising individual and area-level outcomes." *Social Science & Medicine* 67.12:1995-2006.

[131] Pickett, K.E. and Pearl, M. (2001) "Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review." *Journal of Epidemiology and Community Health* 55(2):111-122.

[132] Riva, M., Gauvin, L., and Barnett, T.A. (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.

[133] Diez Roux, A.V. and Mair, C. (2010) "Neighborhoods and health." *Annals of the New York Academy of Sciences* 1186(1): 125-145.

[134] Morris, R.W., Whincup, P.H., Lampe, F.C., et al. (2001)"Geographic variation in incidence of coronary heart disease in Britain: the contribution of established risk factors." Heart 86(3):277-283.

[135] Stringhini, S., Sabia, S., Shipley, M., et al. (2010) "Association of socioeconomic position with health behaviors and mortality." *Journal of the American Medical Association*, 303(12): 1159-1166.

[136] Stringhini, S., Dugravot, A., Shipley, M., et al. (2011) "Health behaviours, socioeconomic status, and mortality: further analyses of the British Whitehall II and the French GAZEL prospective cohorts." *PLoS Medicine* 8(2): e1000419.

[137] Nandi, A., Glymour, M.M. and Subramanian, S.V. (2014) "Association Among Socioeconomic Status, Health Behaviors, and All-Cause Mortality in the United States." *Epidemiology* 25(2): 170-177.

[138] Greenland, S. and Morgenstern, H. (1989) "Ecological bias, confounding, and effect modification." *International Journal of Epidemiology* 18(1): 269-274.

[139] Greenland, S. and Robins, J. (1994) "Invited commentary: ecologic studies—biases, misconceptions, and counterexamples." *American Journal of Epidemiology* 139(8): 747-760.

[140] Richardson, S. and Monfort, C. (2000) "Ecological correlation studies." *Spatial Epidemiology: Methods and Applications*: 205-220.

[141] Greenland, S. (2001) "Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects." *International Journal of Epidemiology* 30(6): 1343-1350.

[142] Wooldridge, J. (2003) "Introductory Econometrics: A modern approach" *2ed. South-Western* Ohio, USA.

[143] Kaufman, J.S., MacLehose, R.F. and Kaufman, S. (2004) "A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation." *Epidemiologic Perspectives & Innovations* 1(1): 4.

[144] Preacher, K. and Hayes, A. (2004) "SPSS and SAS Procedures for Estimating Indirect Effects in Simple Mediation Models," *Behavior Research Methods, Instruments, and Computers* 36(4): 717–31.

[145] Preacher, K. and Hayes, A. (2008) "Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models," *Behavior Research Methods* 40(3): 879–91.

[146] Leroux, B., Lei, X. and Breslow, N. (1999) "Estimation of disease rates in small areas: A new mixed model for spatial dependence" *In Statistical Models in Epidemiology, the Environment and Clinical Trials* Halloran, M. and Berry, D. (Eds) Springer-Verlag, New York: 135-178.

[147] Chow, C. and Liu, C. (1968) "Approximating discrete probability distributions with dependence trees." *Information Theory, IEEE Transactions on* 14(3): 462-467.

[148] Lunn, D., Jackson, C., and Miller, S. (2010) "OpenBUGS Development Interface (Udev): Implementing your own functions"

[149] Statistics Netherlands "Kerncijfers Wijken en Buurten", (http://www.cbs.nl/nl-NL/menu/methoden/dataverzameling/kerncijfers-wijkbuurt-kob.htm) Accessed May 2014.

[150] Kemper H.G.C., Ooijendijk, W.T.M. and Stiggelbout, M. (2000) "Consensus over de Nederlandse Norm voor Gezond Bewegen." [Consensus on the Dutch Norm for Healthy Physical Activity]. *Tijdschrift voor Gezondheidswetenschappen* 78(3):180-183.

[151] Kuipers, M.A., Jongeneel-Grimen, B., Droomers, M., et al. (2013) "Why residents of Dutch deprived neighbourhoods are less likely to be heavy drinkers: the role of individual and contextual characteristics." *Journal of Epidemiology and Community Health* 67(7): 587-594.

[152] Atkinson, R. (2008) "European urban policies and the neighbourhood: an overview." *Proceedings of the ICE-Urban Design and Planning* 161(3): 115-122.

[153] Chaix, B., Evans, D. and Suzuki, E. (2014) "Commentary: socioeconomic status, health behavior, and mortality: old question plus modern methods equals new insights?" *Epidemiology* 25(2): 178-181.

[154] Graham, P. (2008) "Intelligent smoothing using hierarchical Bayesian models." *Epidemiology* 19(3): 493-495.

[155] O'Hagan, A., Forster, J. and Kendall, M.G. (2004) "Bayesian inference." London: Arnold.

[156] Aitchison, J. (1986) "Logistic normal distributions on the simplex." *In The Statistical Analysis of Compositional Data* (pp. 112-140). Springer Netherlands.

[157] Van den Brink, C., Uitenbroek, D., Van den Brink, M., et al. (2009) "Richtlijn wegen voor GGD-epidemiologen."

[158] Karlson, K.B., Holm, A. and Breen, R. (2012) "Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit; A New Method." *Sociological Methodology* 42(1): 286-313.

[159] MacKinnon, D.P. and Dwyer, J.H. (1993) "Estimating mediated effects in prevention studies." *Evaluation Review* 17(2): 144-158.

[160] Greenland, S., Robins, J. M. and Pearl, J. (1999) "Confounding and collapsibility in causal inference." *Statistical Science* pp.29-46.

[161] Cummings, P. (2009) "The relative merits of risk ratios and odds ratios." *Archives of Pediatrics & Adolescent Medicine* 163(5): 438-445.

[162] Lennon, J.J. (@000) "Red-shifts and red herrings in geographical ecology." *Ecography* 23(1): 101-113.

[163] Kühn, I. (2007) "Incorporating spatial autocorrelation may invert observed patterns." *Diversity and Distributions* 13(1): 66-69.

[164] Poortinga, W. (2012) "Community resilience and health: The role of bonding, bridging, and linking aspects of social capital." *Health & place* 18(2): 286-295.

[165] Starfield, B. and Shi, L. (2007) "Commentary: primary care and health outcomes: a health services research challenge." *Health Services Research* 42(6p1): 2252-2256.

[166] Carpenter, J. and Bithell, J. (2000) "Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians." *Statistics in Medicine* 19(9): 1141-1164.

[167] Congdon, P. (2001) "Health status and healthy life measures for population health need assessment: modelling variability and uncertainty." *Health & place* 7(1): 13-25.

[168] Gardner, M.J. and Winter, P.D. (1984) "Mapping small area cancer mortality: a residential coding story." *Journal of Epidemiology and Community Health* 38(1):81-4

[169] Chellini, E., Lo, P.E., Mazzoni, G. and Biggeri, A. (2004) "An unusual excess of mortality in a small Tuscan municipality and the "nursing home effect"." *Epidemiologia e Prevenzione* 28(2): 83-6.

[170] Congdon, P. (2002) "Nursing home mortality: patterns and trends." *Public Health Research Report No.141*, Essex Public Health Directorate, Barking and Havering Health Authority.

[171] Lawson, A.B. (2013) "Bayesian disease mapping: hierarchical modeling in spatial epidemiology." *CRC press.*

[172] Goldstein, H. and Spiegelhalter, D.J. (1996) "League tables and their limitations: statistical issues in comparisons of institutional performance." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* pp.385-443.

[173] Selhub, E.M. and Logan, A.C. (2012) "Your brain on nature." *Hoboken: Wiley.*

[174] Hartig, T. (2008) "Green space, psychological restoration, and health inequality." *The Lancet* 372(9650), 1614-1615.

[175] Huang, A. and Wand. M.P. (2013) "Simple marginally noninformative prior distributions for covariance matrices." *Bayesian Analysis* 8(2): 439-452.

[176] O'Hara, R.B. and Sillanpää, M.J. (2009) "A review of Bayesian variable selection methods: what, how and which." *Bayesian Analysis*, 4(1): 85-117.

[177] Stephens, A.S., Purdie, S., Yang, B. and Moore, H. (2013) "Life expectancy estimation in small administrative areas with non-uniform population sizes: application to Australian New South Wales local government areas." *BMJ Open* 3(12): e003710.

[178] Goesling, B. and Firebaugh, G.(2004) "The trend in international health inequality." *Population and Development Review* 30(1): 131-146.