Type 2 diabetes is a major health challenge globally. The quality of type 2 diabetes care can be evaluated using indicators that are based on clinical guidelines. This study links and analyses electronic health records of all diagnosed type 2 diabetes patients with geospatial and other register-based data from the health care district of Siun sote, in eastern Finland. This dissertation provides valuable information about the quality of type 2 diabetes care at different area-levels.
GEOSPATIAL VARIATIONS IN THE QUALITY OF TYPE 2 DIABETES CARE

EVIDENCE FROM ELECTRONIC HEALTH RECORDS IN NORTH KARELIA, FINLAND
Maija Toivakka

GEOSPATIAL VARIATIONS IN THE QUALITY OF TYPE 2 DIABETES CARE

EVIDENCE FROM ELECTRONIC HEALTH RECORDS IN NORTH KARELIA, FINLAND

Publications of the University of Eastern Finland
Dissertations in Social Sciences and Business Studies
No 221

University of Eastern Finland
Joensuu
2020
Author’s address:  Department of Geographical and Historical Studies
University of Eastern Finland
JOENSUU
FINLAND

Doctoral programme:  Past, Space and Environment in Society

Supervisors:  Professor Markku Tykkyläinen, Ph.D.
Department of Geographical and Historical Studies
University of Eastern Finland
JOENSUU
FINLAND

Professor Tiina Laatikainen, MD, Ph.D.
Institute of Public Health and Clinical Nutrition
University of Eastern Finland
KUOPIO
FINLAND

Professor Timo Kumpula, Ph.D.
Department of Geographical and Historical Studies
University of Eastern Finland
JOENSUU
FINLAND

Reviewers:  Professor Markku Löytönen, Ph.D.
Department of Geosciences and Geography
University of Helsinki
HELSINKI
FINLAND

Assistant Professor Usama Bilal, MD, MPH, Ph.D.
Department of Epidemiology and Biostatistics
Drexel University Dornsife School of Public Health
PHILADELPHIA
USA

Opponent:  Professor Markku Löytönen, Ph.D.
Department of Geosciences and Geography
University of Helsinki
HELSINKI
FINLAND
ABSTRACT

Increasing rates of obesity, lifestyle changes and longer life expectancy are the main driving forces behind the worldwide increase in the type 2 diabetes prevalence. While the prevention of type 2 diabetes is important, providing good quality care for people who have been diagnosed with the disease is at least equally important. In Finland, the care of type 2 diabetes is based on clinical treatment guidelines. However, it is poorly known how the treatment guidelines are implemented in practice and what the real outcomes of care are at the patient level and in different geographical contexts.

This dissertation explores the inequalities in the quality of type 2 diabetes care and in type 2 diabetes prevalence in the health care district of Siun sote and its subregions, located in North Karelia, Finland. The study design is multidisciplinary joining the fields of health geography and health sciences. The empirical part of the dissertation comprises four research articles that aim to investigate whether and how the inequalities in the quality of care are associated with several register- and geographic information system (GIS)- based factors. Electronic health records (EHRs) of type 2 diabetes patients are linked with the register based individual data on patient socioeconomics, the register- and GIS- based small-area factors on socioeconomics, the built environment and the accessibility in the patient’s residential neighbourhood.

This study shows that combining patient EHRs with geospatial perspectives, provides an evidence-based approach that could be utilized to support decision making in chronic disease care and in health care service planning. The use of EHRs from regional patient register is valuable and provides a good opportunity for disease management and monitoring. The results indicate that the type 2 diabetes care assessed by indicators of process of care and treatment outcomes and type 2 diabetes prevalence are not equally distributed in the study region. The results demonstrate that the place characteristics of the patients’ residential neighbourhood are related to the quality of care. Geospatial variation exists between different geospatial scales and areal classifications both in type 2 diabetes prevalence and its care. For example, patients in sparsely populated rural areas compared with patients elsewhere in the region achieve the treatment outcomes the weakest. In postal code areas where the areal educational attainment is low, the achievement of the treatment targets is also poorer.

The findings from this dissertation could be utilised to identify small-areas and settlement types where the disease burden is high and to show areas where patients are at risk of poor diabetes care outcomes. Register- and GIS- based indicators describing the quality of care or the population at various levels of geospatial detail provide
tailored and useful information to be used in type 2 diabetes management and in health care service planning. Subsequently, the management of type 2 diabetes care could be more effectively tailored and improved to the small areas, sub-regions and settlement types that are most in need, as well as to the socioeconomic groups at risk.

**Keywords:** type 2 diabetes, electronic health record, health geography, geospatial data, quality of care, process of care, treatment outcomes, geospatial analysis
Toivakka, Maija
Tyypin 2 diabeteksen hoidon laadun maantieteellinen vaihtelu – Sähköiset terveyskertomukset tiedon lähteenä Pohjois-Karjalassa Suomessa
Joensuu: Itä-Suomen yliopisto, 2020
Publications of the University of Eastern Finland
Dissertations in Social Sciences and Business Studies; 221
ISBN: 978-952-61-3360-7 (nid.)
ISSNL: 1798-5749
ISSN: 1798-5749
ISBN: 978-952-61-3361-4 (PDF)
ISSN: 1798-5757 (PDF)

TIIVISTELMÄ

Lisääntyvä lihavuus, elämäntapojen muutokset ja pidempi elämäntiekon, joissa on muutokset tyypin 2 diabeteksen esiintyvyydestä ja sen esiintyvyydessä. Liikunnan puutuminen ja sähköiset terveyskertomukset ovat tärkeitä terveydenhuollon suunnittelussa. Tyypin 2 diabeteksen hoidon laadun ja sairauden esiintyvyyden tarkistaminen on tärkeää terveydenhuollon tehtävänä.

Tässä väitöskirjassa on käsitelty tyypin 2 diabeteksen hoidon laadun ja sairauden esiintyvyyden tutkimusta Suomessa. Tutkimus on keskustellut terveyskertomuksien käyttöä ja sähköisten terveyskertojen kautta saaduista tietoista. Tutkimus on keskustellut diabeteksen hoidon laadun ja sairauden esiintyvyyden tutkimuksista ja niiden tärkeysä terveydenhuollon tehtävänä.

Tämä väitöskirja on tarkoitus osoittaa, että sähköiset terveyskertojen käyttö on tärkeää terveydenhuollon tehtävänä. Sähköiset terveyskertojen käyttö on tärkeää terveydenhuollon tehtävänä. Sähköiset terveyskertojen käyttö on tärkeää terveydenhuollon tehtävänä.
hoidon hallintaa voidaan muokata ja kohdistaa tehokkaammin eniten apua tarvitselle pienalueille, osa-alueille ja muille alueyhteisöille sekä sosioekonomisille ryhmille.

Avainsanat: tyypin 2 diabetes, sähköinen terveyskertomus, terveysmaantiede, paikkatieto, hoidon laatu, hoitoprosessit, hoidon toteutuminen, geospatiaalinen analyysi
I have lived in Joensuu since 2006—the year when I started to study geography. Initially, my plan was to apply for a master’s degree elsewhere after getting my bachelor’s degree. However, the people, the place and studies made me stay. Years later, I found myself completing my master’s thesis and obtaining my master’s degree here. Next, my plan was to get a job and move away from Joensuu. However, once again, my plans were altered, and my journey as a PhD student started in the spring of 2013.

I am the most grateful for my supervisors: Professor Tiina Laatikainen, Professor Markku Tykkyläinen and Professor Timo Kumpula. I would like to express my sincere gratitude to Tiina and Markku: your support during the entire process has been amazing, thank you. Markku, you have always found time to answer my questions and to read my article drafts and other texts—one after another. Tiina, you have provided inspiring guidance, and from you I have learned a lot about health sciences and type 2 diabetes care. Timo, it was you who introduced me to the field of geoinformatics. Thank you for your time and valuable advice, especially during the years when I was teaching. Your help was priceless.

I would like to thank the reviewers of this thesis: Professor Markku Löytönen who agreed also to serve as an opponent, and Assistant Professor Usama Bilal for kind and valuable comments. I am grateful for all my co-authors, whose contributions to the articles have been irreplaceable. I would like to particularly thank Professor Lauri Mehtätalo and Aki Pihlapuro for providing new insights into statistical methods. I also want to thank Doctor Kati Pitkänen, who introduced me to the academic career and made it possible for me to work as a research assistant and complete my master’s thesis in a research project. Kati, and later Doctor Olga Hannonen, have taught me a lot about second home research.

Thank you to all the staff and PhD students whom I have encountered at the Department of Geographical and Historical studies during these years, especially my colleagues in our Geospatial Health research group. Thank you, Teppo, for all the comments and conversations in these years related to chronic disease care and my synopsis. Thanks, Aapeli and Mikko for sharing an office for past two years and listening to my ups and downs when writing the synopsis. To the “HiMa girls”: Eerika, Eliisa and Olga—thank you for all the snowboarding trips and other activities outside of academics, but also the academic support and advice you have given me during these years. A special thanks to Eliisa for the support in my personal life. To my friends, Anna, Enna, Jenni, Laura and Sanna: you were the reason why I enjoyed my life in Joensuu in the early days. Thank you for still being in my life.

This study was funded by the Strategic Research Council at the Academy of Finland (project IMPRO, 312703, 312704), a six month grant by the North Karelia Regional Fund of the Finnish Cultural Foundation, the Juho Vainio Foundation, the Research Committee of the Kuopio University Hospital Catchment Area for State Research Funding, the Finnish Foundation for Cardiovascular Research, and the Finnish Diabetes Association. I would like to thank all the funders of my work.

Lastly, my warmest thanks go to my family. For my parents, mom and deceased father, I would like to say thank you for always supporting my choices and encouraging me to study what I wanted. Thanks, Juha for listening, supporting and sharing my
moments of despair and happiness. Thank you Aada, for your smile and joy after days at work in the recent years of my PhD journey.

In Joensuu, February 2020
Maija
CONTENTS

ABSTRACT ................................................................................................................................. 7
TIIVISTELMÄ ............................................................................................................................. 9
ACKNOWLEDGEMENTS ........................................................................................................... 11

1 INTRODUCTION .................................................................................................................. 17
  1.1 Type 2 diabetes as a public health challenge ................................................................. 17
  1.2 Electronic health records (EHRs) can be linked with geospatial data .................. 18
  1.3 Research aims .................................................................................................................. 20

2 QUALITY OF CARE IN GEOGRAPHY OF HEALTH ...................................................... 22
  2.1 Evolving medical and health geography ........................................................................ 22
  2.2 Geospatial health geography ......................................................................................... 23
  2.3 Factors associated with the quality of type 2 diabetes care .......................................... 24
  2.4 Place effects on health and health care ......................................................................... 27
    2.4.1 The concept of neighbourhood and health inequalities ...................................... 27
    2.4.2 Built and social neighbourhood environments .................................................. 30
    2.4.3 Accessibility to health care services ....................................................................... 31
  2.5 Use of EHRs and geospatial data in studying diabetes care ......................................... 32

3 MATERIALS AND METHODS .............................................................................................. 34
  3.1 Study region ..................................................................................................................... 34
  3.2 Overview of the data ........................................................................................................ 37
  3.3 EHR data from Mediatri to assess the quality of care .................................................. 37
  3.4 Statistical data .................................................................................................................. 38
    3.4.1 Patient-level socioeconomic data ......................................................................... 38
    3.4.2 Socioeconomic data by postal code area .............................................................. 39
    3.4.3 Data from grid database ....................................................................................... 39
  3.5 Other data on areal characteristics ................................................................................ 39
    3.5.1 Urban-rural classification ..................................................................................... 39
    3.5.2 Urban settlements ................................................................................................... 40
    3.5.3 Built environment characteristics ........................................................................... 40
  3.6 Methods ............................................................................................................................ 40
    3.6.1 Linking EHRs with other data .............................................................................. 40
    3.6.2 Statistical methods .................................................................................................. 42
    3.6.3 Neighbourhood definition ....................................................................................... 42

4 RESULTS .................................................................................................................................. 43
  4.1 Small-area-based SES factors are associated with type 2 diabetes care ....... 43
  4.2 Detailed urban-rural settlement typology reveals areal differences in type 2 diabetes care ................................................................. 44
  4.3 Valid small-area based SES factors provide cost-efficient means to predict type 2 diabetes treatment outcomes ................................................. 46
  4.4 Greenness in the built environment does not enhance the quality of type 2 diabetes care .................................................................................................................... 47
4.5 Summarising the results ............................................................................................................. 49
  4.5.1 Factors associated with the quality of type 2 diabetes care ........................................ 49
  4.5.2 Type 2 diabetes prevalence and quality of care illustrated by different areal classifications ................................................................. 51

5 DISCUSSION ........................................................................................................................ 57
  5.1 Management of type 2 diabetes care could benefit from utilising EHRs and geospatial data ................................................................. 57
  5.2 Small-area-based socioeconomics provide cost-efficient first-hand information for health care service planning ........................................ 59
  5.3 Should health behaviours and experiences be included in EHRs in the future? ......................... 60

6 CONCLUSIONS .................................................................................................................... 61
REFERENCES .................................................................................................................. 64
ARTICLES .................................................................................................................... 73
LIST OF TABLES
Table 1. Summary of patient, statistical, and GIS data used in the study. ....... 37
Table 2. The title, electronic health record (EHR) data, methods and main findings of each research article. ................................................................. 41
Table 3. Comparison of pros and cons among different geospatial scales and areal classifications. ................................................................. 56

LIST OF FIGURES
Figure 1. An illustration on how patient electronic health record (EHR) data, including reference to location, is linked with external geospatial and statistical data ........................................................................... 19
Figure 2. The author’s own perception of the three traditions of medical and health geography and how they overlap. ........................................ 22
Figure 3. A conceptual model that connects individual characteristics, socioeconomic factors, built environment characteristics and access to care with the process of care and treatment outcomes in type 2 diabetes patients. ................................................................. 27
Figure 4. Diabetes reimbursement index by hospital districts (map on the left) and by municipalities (map on the right) in 2017 ......................... 35
Figure 5. The study region in eastern Finland; the health care district of Siun sote was covered by 22 health care centres in 2017 ........................... 36
Figure 6. Individual characteristics and socioeconomic factors that were studied in article I. ............................................................................. 43
Figure 7. Individual characteristics, socioeconomic factors, built environment factors and access to care that were studied in article II. ......... 45
Figure 8. Individual characteristics and socioeconomic factors that were studied in article III. ................................................................. 46
Figure 9. Individual characteristics, socioeconomic factors and built environment factors that were studied in article IV .............................. 48
Figure 10. Summary of the associations between individual characteristics, socioeconomic factors, built environment characteristics and access to care with process of care and the treatment outcomes in type 2 diabetes patients. ................................................................. 50
Figure 11. Spatial distribution of the age-adjusted prevalence (A), percentage of type 2 diabetes patients whose HbA1c was measured (B) and percentage of type 2 diabetes patients who achieved the recommended HbA1c level from those whose HbA1c was measured (C) in 2017. The data are presented by municipalities ....... 52
Figure 12. Spatial distribution of the age-adjusted prevalence (A), percentage of type 2 diabetes patients whose HbA1c was measured (B) and percentage of type 2 diabetes patients who achieved the recommended HbA1c level from those whose HbA1c was measured (C) in 2017. The data are presented by postal code areas .......... 53
Figure 13. Spatial distribution of the age-adjusted type 2 diabetes prevalence (A), percentage of type 2 diabetes patients whose HbA1c was measured (B) and percentage of measured type 2 diabetes patients who achieved the recommended HbA1c level (C) in 2017. The data are presented by urban-rural classification. ........................................54

Figure 14. Spatial distribution of age-adjusted type 2 diabetes prevalence (A), percentage of type 2 diabetes patients whose HbA1c was measured (B) and percentage of measured type 2 diabetes patients who achieved the recommended HbA1c level (C) in 2017. The data are presented by 2 km x 2 km grids. ........................................................55
1 INTRODUCTION

1.1 TYPE 2 DIABETES AS A PUBLIC HEALTH CHALLENGE

Diabetes mellitus is a major global public health challenge. The International Diabetes Federation estimates that 8.4% of the adult population aged 18–99 years lived with diabetes in 2017, and the global prevalence is predicted to rise to 9.9% by 2045 (Cho et al. 2018). Prevalence is the number of existing cases of disease at a particular point in time. The diabetes prevalence varies by age group, gender, income group and also geographically (Cho et al. 2018). Geographical variations exist in diabetes prevalence among continents and countries (Cho et al. 2018), across regions within countries (see for example Bocquier et al. 2011; Zhou et al. 2015; Gurka et al. 2018) and within regions and neighbourhoods (see for example Green et al. 2003; Liu & Núñez 2014; Spratt et al. 2015; Smurthwaite & Bagheri 2017; Wikström et al. 2019; Dekker et al. 2020). This rise in diabetes rates originates from socioeconomic and life-style changes, such as alterations towards a sedentary lifestyle and urbanisation, but also better healthcare, which improves life expectancy, better diagnostics and the availability of newer and higher quality data (Cho et al. 2018).

Diabetes is a metabolic disorder characterised by chronic hyperglycaemia and the long-term effects of the disease include the development of complications, such as nephropathy that may lead to renal failure and retinopathy that can potentially cause blindness (Alberti & Zimmet 1998; Fowler 2008). In addition, people with diabetes are at increased risk of cardiovascular disease (Alberti & Zimmet 1998; Fowler 2008). Type 2 diabetes accounts for 90–95% of all diabetes; the risk of developing type 2 diabetes increases with age, obesity and lack of physical activity (American Diabetes Association 2018). Although these individual-level risk factors are important contributors to type 2 diabetes, there has been growing recognition that the features of residential neighbourhoods and neighbourhood environments may also affect type 2 diabetes (Diez Roux & Mair 2010; Bilal et al. 2018a).

While the prevention of type 2 diabetes is important, it is also crucial to provide good quality care for people who have already been diagnosed with the disease. Ensuring that the patient has a good treatment balance will improve the patient’s quality of life, reduce complications (Zoungas et al. 2012), decrease the risk of comorbidities (Rossi et al. 2011) and reduce the economic burden on public health care (Dall et al. 2014; Keng et al. 2019).

In Finland, public health care services comprise primary and specialised health care that are available to all residents. Primary health care is mainly delivered in public health care centres by general practitioners (GPs). Current Care Guidelines by the Finnish Medical Society Duodecim form the basis of the treatment and management of diseases and risk factors in health care (Current Care Guidelines 2015). The primary aims for the Current Care Guidelines for type 2 diabetes are to provide the means to prevent diabetes and early screening, prevent complications and ensure a balanced treatment and a good life quality for the patients (Type 2 diabetes: Current Care Guidelines 2018). The general aims for the treatment and self-management of the disease are to ensure a good and normal length of life and avoid complications. The guidelines for type 2 diabetes recommend that glycated haemoglobin A1c (HbA1c) should be followed-up regularly, every 6–12 months, and HbA1c should be lower
than 53 mmol/mol (7.0 %). In addition, it is recommended that low-density lipoprotein (LDL) cholesterol should be less than 2.5 mmol/l and blood pressure lower than 140/80 mmHg. The management of patients balanced glycaemic control and other risk factors is extremely important to prevent complications and comorbidities.

Although care in Finland is based on guidelines, no follow-up systems exist to ensure that treatment targets are achieved. However, other countries, such as the United Kingdom (NHS 2016), the United States (National Quality Forum 2011) and Australia (RACGP 2017), have developed measures and standards for monitoring health care quality and achievement of treatment targets for diseases, such as diabetes. Despite the guidelines for type 2 diabetes treatment and management in Finland, it is poorly known how the clinical guidelines are implemented in practice and what are the real outcomes of care in different geographical contexts. This research gap between evidence and treatment in type 2 diabetes care observed in different geographical contexts was the starting point for my dissertation.

1.2 ELECTRONIC HEALTH RECORDS (EHRS) CAN BE LINKED WITH GEOSPATIAL DATA

Electronic health records (EHRs) are digital versions of a patient’s medical records. The use of patient EHRs in epidemiological research, as well as combining EHRs with geospatial data and approaches have increased in the past decade (Casey et al. 2016; Xie et al. 2017; Schinasi et al. 2018; Bravo et al. 2019). Geospatial data is information that includes a reference to a certain location or place on Earth. Electronic health records can be linked to external contextual geospatial data (Figure 1), such as socioeconomic and sociodemographic data on population, as well as data on physical and built environment, by using geographic information systems (GIS). Geographic information systems are computer-based systems for the integration, management, analysis, and visualisation of geospatial data. Geospatial analysis aims to produce new information and additional meaning as a result of the subset of techniques and operations that are applicable to geo-referenced data (De Smith et al. 2009: 26; National Land Survey of Finland 2018a). Although EHRs are collected for clinical purposes, they usually contain a geographic component like municipality, postal code area and residential address. Further on the patient’s address can be geocoded to coordinates for more detailed analysis. Based on these references to a patient’s residential information, patient records can be integrated with external location-specific geospatial data.

EHRs of diabetes patients and geospatial approaches have been used in previous studies for diabetes detection, management and surveillance or monitoring at least in the United States (Geraghty et al. 2010; Liu et al. 2013; Spratt et al. 2015; Gabert et al. 2016; Richardson et al. 2017), Sweden (Sundquist et al. 2015; Mezuk et al. 2016), and Spain (Bilal et al. 2018b). Studies have linked EHRs with small-area information on socioeconomic status (SES) or other neighbourhood characteristics at the zip code, census tract and block group level or by using other geographic areas created for administrative purposes.

EHRs provide some crucial information to providers who are treating individual patients, to public health officials about the health of populations and to researchers about the determinants of health and the effectiveness of treatment (Institute of Medicine 2014). The use of EHRs provides a good opportunity for chronic disease monitoring at a small-area level, outcomes reporting and evidence-based health care (Birkhead et al. 2015; Namulanda et al. 2018).
EHRs contain individual-level patient-provider data that is captured during health care encounters (Casey et al. 2016). EHR data can be both unstructured and structured (Pendergrass & Crawford 2019). The unstructured format means that the health care professional can type free text into the records. EHRs are not designed for research purposes, and utilizing them in research have challenges (Casey et al. 2016; Farmer et al. 2018). First, the data quality can be poor from a research perspective. For example, records might have missing data values or records, or might be in an unstructured format (clinical free text), such as blood pressure. Second, protecting a patient’s privacy is important, and thus there are time-consuming permission processes to complete in order to analyze the data. In addition, researchers must think about a relevant way to report sensitive EHR data. Nevertheless, using EHRs for research purposes has several advantages: large longitudinal datasets, cost-effective data acquisition, objectivity of the datasets, comparability to international studies when using international disease classifications, such as International Classification of Disease (ICD) codes, and the possibility to link EHR data to contextual data (Casey et al. 2016; Farmer et al. 2018).

Currently, the study area of this thesis, the region of joint municipal authority for North Karelia social and health services (Siun sote), utilizes the only regional jointly used electronic patient register in Finland. The regional electronic register covers both primary and specialized health care for 14 municipalities and encompasses approximately 166,000 inhabitants. This regional electronic register provides an excellent opportunity to link patient electronic health records to contextual geospatial data and to study the gap between treatment and evidence at several geospatial scales and areal classifications (Figure 1).

![Figure 1. An illustration on how patient electronic health record (EHR) data, including reference to location, is linked with external geospatial and statistical data (based on and modified from Toivakka et al. 2018). Researchers and health care professionals or health care planners can choose the relevant geospatial scale or context for analysis and reporting.](image-url)
1.3 RESEARCH AIMS

This thesis focuses on the analysis of inequalities in type 2 diabetes prevalence and the quality of care in the entire health care district of Siun sote, located in North Karelia, Finland, and its subregions. The study design is multidisciplinary: it joins the fields of health geography and health sciences. Health geography is a subdiscipline of human geography. A Dictionary of Human Geography (Castree et al. 2013) defines health geography as the study of the relationship between health and place. As Kwan (2012) notes: “it is widely recognized that geographic variations in health cannot be explained exclusively in terms of the characteristics of individuals, as specific characteristics of place or neighbourhoods also exert significant influence on health”.

Health geography emerged in the 1980s as a terminology and has subsequently coexisted with the longer-established field of medical geography (Moon 2009). The distancing of health from medical geography is generally the idea of increased interest in well-being and broader social models of health and health care rather than concerns with disease alone (Kearns & Moon 2002; Moon 2009). A central feature of medical geography has been the use of biomedical models that view humans and diseases in biological terms (King 2010). In social models of health, there is attention to outside the human body, namely to the social and geographical context in which health and disease exist (Moon 2009). Health geographers have used both qualitative and quantitative approaches, including GIS methods for integrating, mapping and analysing spatial data, as well as various statistical techniques (Dummer 2008; Rosenberg 2016). Studies on human health from geographic perspectives, as well as research on applying geospatial analyses to health problems, have developed greatly over the past two decades (Richardson et al. 2013; Lyseen et al. 2014; Shi & Kwan 2015).

The objective of this thesis is to determine whether and how inequalities in the quality of care are associated with socioeconomic factors, built environment factors and the access to care in the patient’s residential neighbourhood. These associations can be useful predictors for health care planning. I link patient-based EHR data with patient and small-area-based socioeconomic data from statistical databases, as well as built environment data with geospatial data of different areal classifications (Figure 1). Combining data from various registers and databases provides a cost-effective way to examine the associations between selected factors in the quality of type 2 diabetes care and type 2 diabetes prevalence by using several geospatial scales and areal classifications. These associations are analysed with GIS and statistical analysis methods. Moreover, the study compares the usability of different register data and areal classifications.

This thesis seeks answers to the following research questions:

1. How are the type 2 diabetes prevalence and quality of care manifested by different geospatial scales and areal classifications?
2. How are various patient-based and small-area factors associated with the quality of type 2 diabetes care?
3. How can different areal classifications be utilised for planning and managing type 2 diabetes care?

Previous studies have used various approaches to define and measure health care quality among type 2 diabetes patients: process indicators (e.g. the frequency of HbA1c and lipid measurements carried out according to the clinical guidelines);
intermediate outcome indicators (e.g. achievement of HbA1c cut-off values); and diabetes-specific complications (e.g. the higher prevalence of retinopathy) (Grintsova et al. 2014). Following these practices and Finnish Current Care Guidelines for type 2 diabetes, assessing the quality of care in this thesis is twofold. First, the process of type 2 diabetes care is assessed through the measurement activity of a certain indicator. Second, the treatment outcomes of type 2 diabetes are assessed through achieving a certain cut-off value for the indicator. Process and treatment outcome indicators are objectively measured laboratory measurements from patient EHRs.
2 QUALITY OF CARE IN GEOGRAPHY OF HEALTH

2.1 EVOLVING MEDICAL AND HEALTH GEOGRAPHY

“[...] ‘places matter’ with regard to health, disease, and health care” (Kearns & Moon 2002).

This thesis has its theoretical background in both medical and health geography. Medical and health geography still work in parallel and it is often difficult to distinguish any difference between the traditions (Moon 2009; Dorn et al. 2010). Thus, it is prudent to explain how medical and health geography have evolved, and how they overlap (Figure 2).

![Figure 2. The author's own perception of the three traditions of medical and health geography and how they overlap.](image)

Two major approaches in medical geography since the mid-twentieth century have been disease ecology and health care service geography (Dorn et al. 2010) (Figure 2). Disease ecology has commonly been understood as complex interactions among the environment, population and culture in explaining and producing disease patterns (May 1960; Mayer 2010). It seeks answers to the questions “why is this disease here” or “why is this disease in places like this” (Mayer 2010). Disease ecology or disease geography focuses largely on infectious diseases and the impact of the natural environment on disease (Mayer 2010; Oppong & Harold 2010); it is closely tied to epidemiology and public health practice (Dorn et al. 2010). However, Oppong and Harold (2010) advocate that rather than viewing the environment in narrow physical terms, social, economic and other factors that characterise spatial variations in disease should be combined in disease ecology approaches. Mayer (2010) argues that the geography of the disease side of medical geography may have better been labelled
“epidemiologic geography”, “public health geography” or “spatial epidemiology”. Health care services geography, on the other hand, focuses on the planning of health care services (Dorn et al. 2010).

In the 1980s and 1990s, disease ecology and health service planning in medical geography were joined by a third tradition: health geography (Dorn et al. 2010). In 1993, Robin Kearns first called for a “post-medical” geography of health (Kearns 1993). He suggested that two interrelated streams should be identified: medical geography and the geography of health. Medical geography would involve the spatial and ecological aspects of disease, such as the disease ecology, and spatial aspects of health care. The latter (health geography) “would consider the dynamic relationship between health and place and the impacts of both health services and the health of population groups on the vitality of places”. Kearns’ ideas were criticised—for example, by Mayer and Meade (1994), who were concerned about the neglect of disease ecology tradition of medical geography. However, it was not Kearns’ intent to discard medical geography; rather, he envisioned health geography as an additional stream (Kearns 1994). The shift from medical to health geography has been described by Kearns and Moon (2002) “as indicative of a distancing from concerns with disease and the interests of the medical world in favour of an increased interest in well-being and broader social models of health and health care”. The study of disease and health care requires integrated rather than separate attention, and this orientation can be linked to the emergence of geography of health (Moon et al. 1998). As Andrews (2018: 44) put it: “understanding has since developed in geography that health and health care are deeply affected by places and the ways in which places are reacted to, felt and represented”.

2.2 GEOSPATIAL HEALTH GEOGRAPHY

Traditional directions in health (and medical) geography such as disease ecology, healthcare access and provision, disparities and contextual effects of a place, especially neighbourhoods, continue to hold their importance (Rosenberg 2014; Grady & Wadhwa 2015). During the last decade, several researchers (Nykiforuk & Flaman 2011; Lyseen et al. 2014; Shi & Kwan 2015; Rosenberg 2016) have reviewed or classified health geography and public health research with GIS approaches. These studies are shortly reviewed in the subsequent paragraphs.

Nykiforuk and Flaman (2011) identified four themes with regard to how GIS approaches have been used to inform decision making in public health from 621 articles and book chapters from 2002–2007. These categories include: disease surveillance or monitoring encompassing disease mapping (illustrations of the distribution of a disease) and disease modelling, risk analysis typically linked with environmental health, health access and planning and community profiling. These themes are not distinct from one another and often overlap. Lyseen and colleagues (2014) reviewed 865 articles from 2000 to 2012 and found four distinct categories within health geography and GIS. These four categories are: the spatial analysis of disease (disease mapping and modelling), the spatial analysis of health service planning, public health (e.g. spatial analysis of health outcomes) and health technologies and tools, including health data collection and manipulation. Within the category of spatial analysis of diseases, the majority of articles have focused on infectious rather than non-infectious diseases. Studies on non-infectious diseases have examined patterns at the prevalence or incidence rates in relation to a geographical component. Shi and
Kwan (2015) classified health research that has applied geospatial analyses into two broad areas: studies on disease and well-being of humans and studies about health care services. The former are referred to as spatial epidemiology. They conclude that geographic perspectives and geospatial methods advance our understanding about the complex interactions between social and physical environments and health outcomes. Rosenberg (2016), on the other hand, grouped recent health geography research that has taken a quantitative or GIS approach into five topics: the geographies of chronic and infectious diseases; access to health services; the food-obesity-built environment nexus; health inequalities; and mental health.

All of these above-mentioned reviews and classifications of health studies with GIS approaches seem to share two common aspects: studies on geospatial analysis on various diseases and health outcomes and studies on health care services. These areas have been the main themes in medical geography (see Figure 2) and are currently referred to as health geography. In the Finnish language, there is only one word for this branch of human geography: “terveysmaantiede”, which refers to health geography. In the Finnish context, geospatial analysis on various diseases and health outcomes (see for example Löytönen 1994; Rytkönen et al. 2001; Rytkönen et al. 2003; Tynelä et al. 2010; Hjort et al. 2016; Repo et al. 2018) and studies related to health care services (Lankila et al. 2016; Huotari et al. 2017) have also been conducted. These aspects are interweaved in my thesis.

Central to my thesis is the linkage of patient electronic health records with geospatial data on small-area socioeconomic and built environment factors. In addition, the aspects of accessibility to health care services are considered. Disease mapping is used to illustrate the geospatial distribution of type 2 diabetes prevalence and quality of care indicators at different local scales. The main focus is on the quality of care and management of the disease instead of incidence or mortality, both of which have been studied more often. In the thesis, I concentrate on several factors that can impact the quality of type 2 diabetes care (see Figure 3). Mapping, analysing associations between small-area factors and the quality of care and the usage of various areal classifications frame the theoretical background into health geography with an emphasis on the geospatial approach.

2.3 FACTORS ASSOCIATED WITH THE QUALITY OF TYPE 2 DIABETES CARE

Individual socioeconomic characteristics, such as educational level, occupation and income, serve as a sign of the prevalence and the risk of developing type 2 diabetes (Espelt et al. 2008; Agardh et al. 2011; Gary-Webb et al. 2013). Moreover, studies indicate that individual socioeconomic status (SES) is associated with the achievement of control targets among type 2 diabetes patients (Sundquist et al. 2011; Grintsova et al. 2014; Bijlsma-Rutte et al. 2018; Ibáñez et al. 2018). SES refers to the position in society that an individual has. More broadly, the concept can refer to the placement of households, census tracts or other aggregates with respect to the capacity to create or consume goods (Miech & Hauser 2001). SES is inversely related to health outcomes. Thus, the higher the socioeconomic status, such as a high level of education or occupation status, the less likely an individual is to suffer from chronic illness, disability, accidents or early death, among other conditions (Kulkarni & Subramanian 2010: 381).
Two systematic reviews have shown that care outcomes in people with type 2 diabetes vary depending on their individual SES, as well as regional deprivation (Ricci-Cabello et al. 2010; Grintsova et al. 2014). Type 2 diabetes patients with a lower SES or a higher area-level deprivation are often associated with worse process indicators of care and worse intermediate outcomes leading to an increased risk of diabetes-associated complications (Grintsova et al. 2014). In addition to individual SES, regional deprivation or low neighbourhood SES has been associated with an increase in type 2 diabetes prevalence (Connolly et al. 2000; Maier et al. 2013; Grundmann et al. 2014; Bilal et al. 2018b), risk of developing the disease (Cox et al. 2007; Krishnan et al. 2010; Bilal et al. 2018b) or worse diabetes care outcomes (Grintsova et al. 2014; Kowitt et al. 2018; Bilal et al. 2018b). Thus, health inequalities are caused by the characteristics of individuals, such as gender, age and SES, but also by the setting in which individuals are located (Curtis & Rees Jones 1998; Gatrell & Elliot 2015a: 125).

The framework of the social determinants of health has been used in geographic health research to elucidate the relationship between context and health outcomes (Curtis 2004; Anthamatten & Hazen 2011: 83). The World Health Organization (WHO) defines the social determinants of health as the conditions in which people are born, grow, live, work and age (WHO 2008). The social determinants of health is a term used as a shorthand to encompass the social, economic, political, cultural and environmental determinants of health (WHO 2011).

Dahlgren and Whitehead (1991) developed a model to assess health inequalities. Their model illustrates the main determinants of health—encompassing individual constitutional factors, individual lifestyle factors, social and community networks and general socioeconomic, cultural and environmental factors. Ansari et al. (2003) argue that a theoretical framework is needed to envelop the social determinants of health, the importance of behaviour and biology and the inter-connectedness of all these factors. They divide social determinants into three components: socio-economic determinants (e.g. age, gender, education), psychosocial risk factors (e.g. social support, chronic stress) and community and societal characteristics (e.g. income inequality, urban or rural residence).

The concept of the social determinants of health has also been utilised when assessing diabetes care. The prevention of or the risk of developing type 2 diabetes have mainly been the focus when studying the impact of social determinants on type 2 diabetes (Walker et al. 2014c). Hill and colleagues (2013), for example, examined the socioecological determinants of health (biological, geographic and built environment factors) that influence risk for prediabetes and type 2 diabetes. Less evidence exists on the associations of the social determinants of health on the progression of type 2 diabetes (Walker et al. 2014c). Gary-Webb and colleagues (2013) stated that “broadening the study of social determinants is a necessary step toward improving the prevention and treatment of type 2 diabetes”.

Brown and colleagues (2004) developed a conceptual framework for the mechanisms that connect socioeconomic factors and health in individuals with diabetes. They discuss three main mechanisms posited to influence this relation: health behaviours, access to care and processes of care. They argue that to reduce health disparities, we should have an understanding about the individual and contextual factors that may influence health outcomes, such as diabetes outcomes, and the associations among these factors. This understanding can be achieved by examining individual, system-level and area-level factors and their relation to access to care, health behaviours and quality of care.
Walker and others (2014b) modified the model proposed by Brown et al. The revised model hypothesises the direct effects of socioeconomic variables on diabetes outcomes (glycaemic control) and indirect effects through mediators of health behaviours, access to care and processes of care. Based on their findings, there are direct and indirect pathways through which social determinants influence diabetes outcomes. For example, employment and lower diabetes distress are directly associated with lower HbA1c. On the other hand, higher income is associated with greater access and lower processes of care. Further, Walker et al. (2014a) studied the socioeconomic and psychological social determinants of health on diabetes knowledge, self-care, diabetes outcomes and quality of life. They hypothesise that lower levels of socioeconomic factors and psychological factors will be associated with poor self-care behaviours (e.g. diet, blood sugar testing), worse diabetes outcomes (HbA1c, cholesterol, blood pressure) and lower quality of life. Based on their results, socioeconomic factors are most often associated with diabetes outcomes and knowledge, while the psychological factors of self-efficacy and perceived stress are most often associated with the self-care and quality of life. Their results suggest that social determinants of health are associated with diabetes outcomes and should be considered in clinical care.

Gonzalez-Zacarias and others (2016) recommend multifaceted approaches for assessing glycaemic control among type 2 diabetes patients. They argue that understanding the social determinants that affect diabetes care, such as the interaction among demographics, knowledge, environment and other diabetes-related factors, may provide insight for improving glycaemic control. In addition, the neighbourhood social environment may influence medication adherence among type 2 diabetes patients (de Vries McClintock et al. 2015).

Given that the concept of the social determinants of health does not describe the whole spectrum that the concept encompasses, I will use an alternative term in this thesis. Mayer (2010: 44) suggests that the social determinants of health should be called the social influences on health. Following Mayer’s idea—and in order to better describe the factors I empirically study—I define the concept as the socioeconomic and environmental influences on health.

Adapting and following the idea of Brown and colleagues’ and Walker and colleagues’ model, Figure 3 illustrates the conceptual model of my thesis. This model describes the relationship between the socioeconomic and environmental influences on the quality of type 2 diabetes care. Further, the quality of care is assessed through indicators related to the process of care and treatment outcomes at the individual patient level. I divided the socioeconomic and environmental influences on health into four categories: individual characteristics, socioeconomic factors, built environment characteristics and access to care. Factors in these four categories are used as independent variables in statistical analyses. The dependent variables are the process of care and treatment outcomes variables. The composition in statistical analyses is correlative. The factors on individual characteristics, socioeconomic factors, built environment characteristics and access to care are used as predictors for the quality of care, but causal inferences cannot be made. Then, the results can be reported at the individual level or by choosing the desired geospatial scale or areal classification, as demonstrated in Figure 3. The small pictures of layers in Figure 3 represent the GIS data used in the analyses. The arrows in the figure represent the tendency for which way or how it is assumed that the relationship between the studied factors and type 2 diabetes care might act. It has to be noted that the arrows do not represent causality. The characteristics and factors that are studied in the thesis are presented in black.
However, I consider that it is important to demonstrate how complex the system is: other factors are also related to the quality of type 2 diabetes care. Thus, the factors that are not studied empirically in my thesis are presented in Figure 3 in grey.

Figure 3. A conceptual model that connects individual characteristics, socioeconomic factors, built environment characteristics and access to care with the process of care and treatment outcomes in type 2 diabetes patients. The model helps to assess the relationships of the socioeconomic and environmental influences in the local environment on the quality of type 2 diabetes care. The arrows do not indicate causality.

The study design in my thesis is cross-sectional, and therefore causal interpretation of the studied associations cannot be made. I use several socioeconomic factors, built environment factors, accessibility and patient characteristics as predictors of quality of type 2 diabetes care. Nevertheless, it is important to acknowledge that the mechanisms behind the associations of contextual characteristics and individual outcomes are unclear (Monden et al. 2006). Further, some other factors, such as the health behaviour of the patients, may be the root causes that have an effect on type 2 diabetes care.

2.4 PLACE EFFECTS ON HEALTH AND HEALTH CARE

2.4.1 The concept of neighbourhood and health inequalities

Interest in place, area or neighbourhood health effects has been a popular field of study since the beginning of the 1990s in health geography, epidemiology and public health
(see for example Macintyre et al. 1993; Pickett & Pearl 2001; Macintyre et al. 2002; Riva et al. 2007; Yen et al. 2009; Diez Roux & Mair 2010; Oakes et al. 2015). Residential neighbourhoods, environments or areas have emerged as potentially relevant contexts because they possess both physical and social characteristics that plausibly influence the health of individuals (Diez Roux & Mair 2010). Place in my thesis is understood as the residential neighbourhood or the neighbourhood environment where type 2 diabetes patients live. I study the associations of several place characteristics on the quality of type 2 diabetes care.

Diez Roux and Mair (2010) summarised the trends that have driven the increasing interest in neighbourhoods and health. The first trend has been the growing recognition that beyond only considering individual characteristics, features of the groups and contexts to which individuals belong need to be considered. Otherwise, one might miss important features that might be associated with health outcomes. A second trend has been the interest in understanding the causes of social inequalities and ethnic differences in health. Neighbourhood characteristics might contribute to inequalities in health because the place of residence can be strongly patterned by social position. A third trend has been the need to consider the health effects of policies because they might impact the contexts in which individuals live. A fourth trend has been the availability and popularity of methods, such as multilevel analysis, GIS and geospatial analysis techniques, all of which allow for a more detailed examination of place. However, one should remember that neighbourhood factors may not affect everyone equally. Further, neighbourhood context may play a limited role in behavioural choices that are for the outcomes of a complex set of processes (Diez Roux 2016). Socioeconomic characteristics of the residential neighbourhood might affect the lifestyles or service-seeking behaviours of individuals that are not representative of the area by non-existing or congested services. In this case, the neighbourhood of the individual can indirectly affect ones health.

A number of studies have investigated whether area differences in health outcomes were due to the composition of the resident population or to the features of place not captured by individual characteristics (Macintyre et al. 2002). Multilevel modelling became a key method for researching the role of geographical context in influencing health outcomes (Owen et al. 2016). However, Macintyre and others (2002) suggests that “the distinction between composition and context may be more apparent than real”. The characteristics of individuals are shaped by the features of the place. This problem is the first of three they identify with context versus composition approach. Second, individual characteristics, such as diet or physical activity, may be intervening variables on the pathways between place and health. These individual confounding variables might have already been influenced by features of the place. Third, there is a lack of clear theorising about the mechanisms that might link the area of residence and health, and which might form the basis for the selection and interpretation of variables. Context (or residential neighbourhood and neighbourhood environment) is a black box that influences some aspects of health, health-related behaviours or health risks, but we do not know how (Macintyre et al. 2002). Smyth (2008) also notes in her report on the geographies of health inequalities that plenty of the research on health inequalities focus on either the role of context or composition in explanation, but it would be important to gain a real understanding of the underlying causes of health inequalities. However, causal pathways are not straightforward to address, and it should be noted that many studies of neighbourhood health effects do not claim that the observed associations would be causal (Diez Roux 2004).
The terminology used in studying neighbourhood and health is inconsistent. Phrases of ‘area’ and ‘ecological effects’, ‘place’, ‘neighbourhood’, ‘context’ and ‘environmental’ are used for place effects on health in scientific literature (Cummins 2018: 141). Typically, health geography researchers define neighbourhoods by using different administrative units: census based definitions, those used in local government or by postal services (Gatrell & Elliot 2015b: 158). These readily available administrative geographical units may not be appropriate scales to use for different types of human activities (Macintyre et al. 2002) and may not coincide with the neighbourhoods that have an effect on health (Flowerdew et al. 2008). Uncertainties related to neighbourhood effects have been raised by van Ham & Manley (2012) and Kwan (2009; 2018). They note that much of the research on neighbourhood effects assumes that the individuals’ residential neighbourhood is the most relevant context that affect their health. This supposition ignores the role of time and human mobility. They highlight that there is a challenge to move away from single point-in-time measures of neighbourhood characteristics and to consider people’s neighbourhood histories. Furthermore, due to personal characteristics, the way an individual perceives, understands and reacts to factors in the neighbourhood might lead to distinct behaviours and outcomes.

As noted, neighbourhoods can be defined in many ways, and it is important to remember that conclusions may differ depending upon how the boundaries are drawn and the data aggregated (Gatrell & Elliot 2015b: 158). This issue is an example of the modifiable areal unit problem (MAUP)—a classic problem in the statistical analysis of geographical data (Flowerdew et al. 2008). The analytical results for the same data in the same study area can be different if they are aggregated in dissimilar ways. Kirby and others (2017) encourage researchers to conduct analyses “at different scales to test the robustness of the spatial relationships and the effect of different artificial boundaries”. Flowerdew and others (2008) note that it is important to consider whether the chosen areal unit is the best way to represent the processes that generate the data. Areas that range from small to large with varying geographic definitions may be important for different health outcomes or mediating mechanisms (Diez Roux 2001). Macintyre and others (2002) point out that there is a need to think of ways of modifying measures and spatial scale to consider rural or sparsely populated areas, given that much of the research on neighbourhoods and health relates to urban neighbourhoods. Meijer and colleagues (2012) encourage researchers to include multiple area levels in future investigations of neighbourhoods, morbidity and mortality, as people engage in different contexts, all of which contribute to their health (for example, a small-scale neighbourhood, a municipality and a region).

Macintyre and colleagues (1993) describe five contextual or place characteristics that might explain health inequalities. The first is physical features of the environment shared by all residents in a locality (e.g. air and water quality) and that are likely to be shared by neighbourhoods across wide areas. The second is the availability of healthy environments at home, work and play (e.g. decent housing, nutritious food, and healthy recreation). These environments in the second category are opportunities that may or may not be taken, with various degrees of choice. The third category includes the provided services to support people in their daily lives, including education, transport and health services. The fourth category includes the socio-cultural features of a neighbourhood (e.g. the political and economic history, and the current characteristics of the community). The fifth category is the reputation of an area (e.g. how the area is perceived by the residents); this factor might influence who
moves in or out of the area. These five features of local areas may be health promoting or health damaging.

In the next two sections of this thesis, I will introduce findings from previous studies that study features of place related to categories two, three, four and five in relation to type 2 diabetes. All of these aspects are not covered empirically in my thesis. I concentrate my empirical analyses (Figure 3 and Tables 1–2) on neighbourhood socioeconomic factors, built environment and accessibility, all of which can be placed in categories two, three, and four.

2.4.2 Built and social neighbourhood environments

The physical environment encompasses traditional environmental exposures, such as air pollution and noise, as well as features of the built environment (Diez Roux & Mair 2010; Cummins 2018: 144). The built environment refers to the man-made environment or surroundings of a neighbourhood, including land use and transportation (e.g. density of fast food restaurants or intersections), features of public spaces and access to resources such as recreational opportunities (Diez Roux & Mair 2010; Piccolo et al. 2015). The social environment includes characteristics related to the social life of the neighbourhood, such as the social relationships between the residents, presence of social norms and levels of safety and violence (Diez Roux & Mair 2010; Cummins 2018: 144). These neighbourhood environment features may affect health in general and diabetes related outcomes through several potential mechanisms: diet, physical activity, stress and social cohesion.

Food environments have been operationalised as favourable for health, such as access to healthier foods, or unfavourable for health, such as access and density of fast food restaurants (Bilal et al. 2018a). The results between the food environment and diabetes have been mixed. den Braver and colleagues (2018) reviewed built environmental characteristics and diabetes risk and prevalence. They found no consistent evidence for an association between the food environment and type 2 diabetes risk and prevalence. Bilal et al. (2018a) also found conflicting results between food environments and diabetes risk in their review. For example, there were no associations between fast food restaurant, convenience store, super store or grocery store densities and the prevalence of type 2 diabetes at the county level in South Carolina in the US (AlHasan & Eberth 2016). However, food insecurity (limited food access owing to cost) has been associated with poor glycaemic control among type 2 diabetes patients (Walker et al. 2018) and diabetes patients in general (Berkowitz et al. 2018), but not living in an area with low physical food access (Berkowitz et al. 2018). In addition, losing or gaining a supermarket in a neighbourhood has not been associated with meaningful change in HbA1c when studying the impact of food environment on glycaemic control (Zhang et al. 2017). A lower type 2 diabetes risk has been associated with features of neighbourhood environment that support both healthy foods and physical activity (Auchincloss et al. 2009; Christine et al. 2015). On the other hand, no association have been found between risk of type 2 diabetes and geographic proximity to supermarkets (Christine et al. 2015).

Built environments that affect physical activity are more consistently associated with diabetes. A review by Dendup and others (2018) suggests that higher level of walkability and green space are associated with a lower risk of type 2 diabetes. Similarly, den Braver et al. (2018) conclude in their review that a built environment
including walking and access to green space is associated with reduced diabetes risk and prevalence. Furthermore, they found that urbanisation is associated with higher type 2 diabetes risk and prevalence. Higher levels of green space in built environments have been associated with lower type 2 diabetes risk (Astell-Burt et al. 2014) and prevalence (Bodicoat et al. 2014; Lee et al. 2017; Müller et al. 2018). Walkability has been used in several studies of the built environment and diabetes (Bilal et al. 2018a). For example, there was a negative association between neighbourhood walkability and incidence of type 2 diabetes in studies from Australia (Müller-Riemenschneider et al. 2013), Sweden (Sundquist et al. 2015) and Canada (Creatore et al. 2016). Residential walkability has also been positively associated with glycaemic control in a longitudinal study in New York city in adults with diabetes (Tabaei et al. 2018). In addition, a systematic review by Chandrabose et al. (2019) on the built environment and cardio-metabolic health of longitudinal studies concludes that living in more walkable areas is likely to have protective effect against the development of type 2 diabetes. However, it is important to note that better opportunities for physical activity do not necessarily mean that people exploit them: some make use of the possibilities, but others will probably not.

Studies on neighbourhood social environments are less common compared to investigations of neighbourhood built environments (Diez Roux & Mair 2010), but they have increased in recent years. Neighbourhood social environment assessed by safety and social cohesion have not been associated with the development of type 2 diabetes (Christine et al. 2015). However, higher neighbourhood social cohesion has been associated with a lower incidence of type 2 diabetes in African Americans (Gebreab et al. 2017). Diabetes control (HbA1c > 9 % or no record of HbA1c) has not been associated with neighbourhood social environment assessed by violent crime rate, perceived safety, social capital and African American residential segregation (Lê-Scherban et al. 2019). Gariepy and others (2013) reported that neighbourhood characteristics, such as perceived order, social and cultural environment and access to services and facilities, can affect diabetes distress (worry, frustration and discouragement that may accompany life with diabetes) in adults with type 2 diabetes.

2.4.3 Accessibility to health care services

Type 2 diabetes care requires frequent visits to health care services. Therefore, accessibility to health care services may be associated with type 2 diabetes care. Accessibility—measured as distance, transportation, travel time or cost—is one of the five dimensions of access (Penchansky & Thomas 1981). The other dimensions are: availability (the supply of services), accommodation (hours of operation, waiting times), affordability (price of services) and acceptability (clients’ satisfaction). It is commonly thought that health care service utilisation decreases as distance increases.

Liese and others (2019) hypothesised that accessibility measured as road distance between young type 1 and type 2 diabetes patients and the health care provider is inversely associated with glycaemic control. They found no significant association. Similarly, Butalia and colleagues (2014) found that driving distance from home to diabetes care sites was not associated with glycaemic control in an urban setting among patients with type 1 diabetes. However, increased driving distance from the patient’s home to the primary care facility has been associated with poor glycaemic control in rural areas (Strauss et al. 2006; Zgibor et al. 2011) and lower use of insulin.
among type 2 diabetes patients (Littenberg et al. 2006). In addition, remote dwellers with diabetes and chronic kidney disease were less likely to receive recommended quality care compared with those living within 50 km of a kidney specialist (Bello et al. 2012). These mixed findings between accessibility and diabetes care outcomes might be at least partly due to varying sample sizes, different studied outcome measures, the distinct definitions of patient groups, differing health care organisations among countries and the kinds of areas (urban or rural) explored.

### 2.5 USE OF EHRs AND GEOSPATIAL DATA IN STUDYING DIABETES CARE

Approaches that utilise the electronic health records (EHRs) of diabetes patients and geospatial data have been employed in some previous studies (Geraghty et al. 2010; Spratt et al. 2015; Gabert et al. 2016; Richardson et al. 2017; Bilal et al. 2018b; Hirsch et al. 2018; Lê-Scherban et al. 2019) when assessing the quality of diabetes care. These studies focus on socioeconomic status and place effects and are shortly reviewed in following paragraphs.

Geraghty and others (2010) combined the EHRs of 7,288 type 2 diabetes patients from California, the United States, with neighbourhood SES variables using GIS methods. They found that neighbourhood SES at the census tract level was a barrier to optimal glucose control but not to lipid control. Lower income neighbourhoods had higher HbA1c, a finding that indicate less controlled diabetes. The Euclidean distance from a patient’s home address to their primary care clinic was not related to their HbA1c level. They conclude that GIS methods are an important tool for primary care and can provide guidance for disease management at a local level.

In Durham County, North Carolina, the United States, the clinical data of 22,982 patients with type 2 diabetes from EHRs was connected with social and environmental data to identify specific risk profiles of patients and neighbourhoods (Spratt et al. 2015). EHR data was connected with variables such as neighbourhood characteristics, census data, environmental data, and areas for outdoor or indoor recreation. However, these neighbourhood and environmental variables were not analysed with diabetes outcomes. A countywide HbA1c measurement was not available for 29.2 % of patients, and the HbA1c was outside of recommended level (> 7 %) for 39 % of the patients. The authors argue that linking patient data to neighbourhood-level characteristics can describe patient populations over space and time, and assist the decision-making and evaluation of clinical care.

Gabert and others (2016) examined small-area variation in HbA1c levels in three counties in Minnesota, the United States, among 63,053 diabetes patients. They identified zip code areas where targets for HbA1c, blood pressure, LDL-cholesterol and tobacco cessation were least commonly achieved. These clinical measures were strongly correlated with the average zip code area level of income, education and insurance coverage. The proportion of patients attaining HbA1c < 8.0 % ranged from 59–90 % across zip code areas. The authors argue that EHR data may be a useful, low-cost approach for identifying high risk neighbourhoods.

The HbA1c test results of 18,131 type 2 diabetes patients from EHRs were combined with United States census tract data on income and ethnicity to assess disparities in diabetes prevalence in California, the United States (Richardson et al. 2017). Their aim was to assess the validity of HbA1c test results for the public health surveillance
of diabetes, including the assessment of geographic, ethnic and income disparities in diabetes control. The prevalence of 5-year maximum HbA1c ≥ 6.5 % decreased with increasing median family income and increased with greater proportions of residents who were either non-Hispanic black or Hispanic.

In the city of Madrid, Spain, data from EHRs, including 23,908 patients with diabetes, was combined with information on neighbourhood SES (Bilal et al. 2018b). The aim was to study the association between neighbourhood SES (an index of seven indicators) and diabetes prevalence, incidence and control. Diabetes was uncontrolled (HbA1c ≥ 7 %) in 39 % of the patients. When neighbourhood SES improved, diabetes prevalence, lack of control and incidence decreased.

Primary care patients with type 2 diabetes and elevated HbA1c measures (n = 15,308) in townships, boroughs and census tracts in cities were identified from EHRs across Pennsylvania and New Jersey, the United States (Hirsch et al. 2018). Elevated HbA1c levels were identified ≥ 7.5 %. The objective was to evaluate whether treatment intensification and community factors (food environment, physical activity environment or community socioeconomic deprivation) were associated with HbA1c levels over time. Community socioeconomic deprivation was derived from United States census data and included the proportion of the population in poverty, unemployed on public assistance, with less than a high school education and not in the labour force. The average reduction in HbA1c was less pronounced in townships with a high level of community socioeconomic deprivation. Community socioeconomic deprivation was not associated with an HbA1c change in boroughs and cities. Treatment intensification occurred for 34.9 % of elevated HbA1c values. They authors conclude that the observed associations provide evidence to support the study of disease management strategies that take community factors into consideration.

Lê-Scherban and colleagues (2019) studied the associations between characteristics of patients’ residential neighbourhoods (census tracts) and diabetes control (n = 1,061) among EHRs of African American type 1 and type 2 diabetes patients in Philadelphia, the United States. They examined associations with survey-based neighbourhood SES (poverty, education and deprivation index), social environment (violent crime, perceived safety and social capital and racial segregation) and built environment (land-use mix and intersection density) characteristics. They found that poor diabetes control (HbA1c > 9 % or no record of HbA1c) was more common in highly segregated neighbourhoods and less common in neighbourhoods with more retail land use. Other measures of the neighbourhood SES, social environment, and built environment were not associated with diabetes control. They conclude that it is crucial to consider how specific community-level social determinants of health affect patients beyond individual-level determinants. This information provides evidence for policymaking to improve population health and health equity.

Although the above mentioned studies use a different granularity of geographic scales, distinct patient group definitions and specific thresholds for glycaemic control, the results indicate that when small-area SES increases diabetes is better controlled. In addition to SES measures, only one study used accessibility in analyses (Geraghty et al. 2010), three mentioned or analysed other neighbourhood characteristics—but one only mapped them instead of studying the associations (Spratt et al. 2015)—and one had relatively small sample size and specific patient group (Lê-Scherban et al. 2019). Furthermore, only one study out of seven was conducted in Europe.
3  MATERIALS AND METHODS

3.1  STUDY REGION

Eastern Finland and the county of North Karelia suffer from a high disease burden. For example, the region has suffered from extremely high cardiovascular mortality after the Second World War (Vartiainen et al. 2016; Vartiainen 2018). Puska (2016) and Vartiainen (2018) have described the results and experiences from the North Karelia project that was started in 1970s to carry out a community-based cardiovascular prevention programme. The results indicate major declines in cholesterol, blood pressure and smoking levels. Coronary mortality was reduced by 84 % from 1972 to 2014. Following in the footsteps of the North Karelia Project and prevention of cardiovascular disease, activities related to the prevention of type 2 diabetes have been implemented (Lindström et al. 2016). The increasing type 2 diabetes prevalence has been suggested to underlie a new upward turn in cardiovascular diseases as cardiovascular disease is one complication of type 2 diabetes (Lindström et al. 2016).

Although the incidence and mortality of coronary heart disease has declined in North Karelia, the overall morbidity in the region has remained high. According to the morbidity index of Finnish institute for health and welfare (THL 2019) the population is healthier in western and southern parts of Finland compared with the population in the east and north. When comparing the Finnish counties with the entire country (the morbidity index for 2014–2016 is 100), morbidity is the second highest in the county of North Karelia (morbidity index: 122). In order to prevent the complications of type 2 diabetes, such as cardiovascular disease, it is important to explore the state of the quality of type 2 diabetes care in the region.

Figure 4 illustrates a coarse estimate of the diabetes burden (the diabetes type is not separated) in Finland using the diabetes reimbursement index by hospital districts and by municipalities. This representation also shows the high disease burden in eastern Finland. The health care district of Siun sote (joint municipal authority for North Karelia social and health services) had the second largest diabetes reimbursement index (114.2) in Finland in 2017. While this figure presents a coarse estimate for the diabetes burden in Finland, a data comparison by Wikström and others (2019) demonstrated that the number of diabetes patients from regional EHRs and number of medication reimbursement rights were similar.
Figure 4. Diabetes reimbursement index by hospital districts (map on the left) and by municipalities (map on the right) in 2017. Kela (the Social Insurance Institution of Finland) includes diabetes as one of the diseases of public health importance. Diabetes is a disease where the medication is entitled to special reimbursements. The index is calculated and linked to the average for the entire country. The index indicates how common the special reimbursements are in the area compared to whole country’s average (= 100). The values are standardised for age and sex.

The health care district of Siun sote (joint municipal authority for North Karelia social and health services) provides both primary and specialised health care services for 14 municipalities in the region (Figure 5). The health care district of Siun sote comprises 13 municipalities that belong to the county of North Karelia and the municipality of Heinävesi which is part of the county of South Savo. In 2017, the total population of the health care district was 166,441, with a population density of 8.9 people per square kilometre (Official Statistics of Finland 2019). The region is characterised by a regional centre, a commuting zone of Joensuu, with approximately 75,000 inhabitants. The region is rather rural; in 2017, 72.0 % of the population lived in urban settlements compared with the Finnish total urban population of 85.9 % (Official Statistics of Finland 2018). The population is older on average (mean age: 45.3 years, share of people aged 65 years or over: 25.1 %) compared with entire Finnish population (mean age: 42.7 years, share of people aged 65 years or over: 21.4 %) (Official Statistics of Finland 2019).
Figure 5. The study region in eastern Finland; the health care district of Siun sote was covered by 22 health care centres in 2017. The region is characterised by a regional centre: a commuting zone of Joensuu. The population in the commuting zone of Joensuu is younger than elsewhere in the health care district of Siun sote.
3.2 OVERVIEW OF THE DATA

Electronic health records of type 2 diabetes patients, statistical data on several spatial scales and built environment characteristics are the main data sources of this thesis (Table 1). All data are extracted from different registers and databases.

Table 1. Summary of patient, statistical, and GIS data used in the study.

<table>
<thead>
<tr>
<th>Type of data</th>
<th>Data Source</th>
<th>Scale</th>
<th>Format</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient characteristics</td>
<td>Electronic health records (EHRs)</td>
<td>Individual</td>
<td>Tabular / Coordinates</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>(age, gender, diagnoses, laboratory results, home address)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical data</td>
<td>Statistics Finland, FIONA remote access (restricted access, chargeable)</td>
<td>Individual</td>
<td>Tabular</td>
<td>III</td>
</tr>
<tr>
<td>(income, education, unemployed)</td>
<td>Statistics Finland, Paavo-database (open data, free of charge)</td>
<td>Postal code area</td>
<td>Tabular / Polygon GIS data</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Statistical data</td>
<td>Statistics Finland, Grid Database (chargeable)</td>
<td>250 m x 250 m grid</td>
<td>Tabular / Polygon GIS data</td>
<td>IV</td>
</tr>
<tr>
<td>[median income (thousands / €), educated (%), unemployed (%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical data</td>
<td>Finnish Environment Institute (SYKE)</td>
<td>based on 250 m x 250 m grid</td>
<td>Polygon GIS data</td>
<td>II, IV</td>
</tr>
<tr>
<td>[educated (%), unemployed (%), average age of inhabitants]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban-rural settlement type classification</td>
<td>Finnish Environment Institute (SYKE)</td>
<td>based on 250 m x 250 m grid</td>
<td>Polygon GIS data</td>
<td>II</td>
</tr>
<tr>
<td>(inner urban area, outer urban area, peri-urban area, local centres in rural areas, rural areas close to urban areas, rural heartland areas, sparsely populated rural areas)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban settlements</td>
<td>Finnish Environment Institute (SYKE)</td>
<td>based on 250 m x 250 m grid</td>
<td>Polygon GIS data</td>
<td>II</td>
</tr>
<tr>
<td>Characteristics of the built environment</td>
<td>National Land Survey of Finland, Topographic database CORINE Land Cover 2018</td>
<td>1:10,000 / 1:100,000</td>
<td>Polygon GIS data</td>
<td>IV</td>
</tr>
<tr>
<td>(meadow, agricultural field, park, sport/recreational area) (forests)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.3 EHR DATA FROM MEDIATRI TO ASSESS THE QUALITY OF CARE

Electronic health records (EHRs) are digital versions of a patient’s medical records. Partly patient transactions, such as information on clinical visits and laboratory measurements are routinely collected into patient electronic health records but partly health care professionals are responsible for recording them. For example, professionals have to record the diagnosis of the visit, risk information, such as
smoking or alcohol consumption, as well as physiological measurements, such as the patient’s blood pressure, height and weight.

The Siun sote region utilises the only regional jointly used electronic patient register in Finland—called Mediatri. EHRs of patients who had a visit to health care with type 2 diabetes diagnosis (ICD10 code E11) were retrieved from the regional electronic patient database. Patient information comprised age, gender, date of birth, the place of domicile (municipality, postal code, address), laboratory data and clinical visit data. Initially, the data comprised all diagnosed type 2 diabetes patients (n = 10,204) at the end of 2012 (articles I–III) with a home municipality in the county of North Karelia. Subsequently, another data retrieval was performed, which consisted all diagnosed type 2 diabetes patients (n = 13,545) at the end of 2017 (Article IV) who had their home municipality in the health care district of Siun sote (municipalities in North Karelia and the municipality of Heinävesi).

The process of care and treatment outcomes were assessed by divergent laboratory results and information recorded in the clinical visits. The primary outcome measure (articles I–IV) was glycated haemoglobin A1c (HbA1c). HbA1c provides a long-term blood sugar value, and it indicates whether the patient has good or poor glycaemic control. The Current Care Guidelines recommends that HbA1c should be lower than 53 mmol/mol (7.0 %), and it should be measured every 6–12 months (Type 2 diabetes: Current Care Guidelines 2018). Secondary outcome measures (article IV) included low-density lipoprotein (LDL) cholesterol and blood pressure (BP). LDL cholesterol should be lower than 2.5 mmol/l and blood pressure lower than 140/80 mmHg (Type 2 diabetes: Current Care Guidelines 2018).

The process of care was assessed by follow-up rates. In articles I and II, whether HbA1c was measured during the years 2011–2012 was assessed. Only the latest measurement during 2011–2012 was used. In article IV, whether HbA1c, LDL, and blood pressure were measured during 2016–2017 was assessed, and again only the latest measurement was used. In order to specify if HbA1c, LDL and blood pressure were measured, the measurement had to be from the same date as the type 2 diabetes diagnosis or after the diagnosis. If the measurement was performed before the diabetes diagnosis, it was not included.

The achievement of treatment outcomes was assessed among those patients who had process indicators measured. Treatment outcomes were assessed through achieving a certain cut-off value for the indicator considering an appropriate period for treatment effect. When analysing the attainment of the recommended HbA1c level, only patients who had at least three months between their diabetes diagnosis and their last HbA1c measurement were included in the analyses. When analysing the attainment of the recommended LDL level, only patients who had at least one month between the diabetes diagnosis and their last LDL measurement were included in the analyses. For blood pressure, no period for the treatment effect was applied.

### 3.4 STATISTICAL DATA

#### 3.4.1 Patient-level socioeconomic data

Article III uses individual-level socioeconomic characteristics of type 2 diabetes patients. The data were provided by Statistics Finland via a protected remote FIONA
environment. Each patient’s earned income (€), educational attainment and whether the patient was unemployed were used in the analysis.

3.4.2 Socioeconomic data by postal code area

Socioeconomic variables at the postal code area level are used in articles I–III. Statistics Finland has provided open statistical data combined with map data by postal code areas from Paavo database since 2015 (Statistics Finland 2018c). Before 2015, statistical data by postal code areas were subject to a charge. The data contain information such as population structure, the degree of education, the income of the inhabitants and households and the main activities of the inhabitants. The data are protected if the population in the data group is less than 10 or 30 (depending on the data group). At the time of writing articles I–III, the postal code area data were used from the 2011 -database.

3.4.3 Data from grid database

The grid database is a chargeable product that contains Statistics Finland’s coordinate-based statistical data calculated by a map grid (Statistics Finland 2018b). Statistical data by map grids provides an opportunity to observe the population structure and socioeconomic variables independently of administrative boundaries. Article IV utilises data from the 2017 grid database. The grid database contains data groups such as population structure, the educational structure of the population and buildings and housing. These data are protected if the population in the grid is less than 3 or 10 (depending on the data group). The data on educational structure are confidential if the grid contains fewer than 10 people aged 18 years or over, for example.

3.5 OTHER DATA ON AREAL CHARACTERISTICS

3.5.1 Urban-rural classification

Information on regional development in Finland has traditionally been based on administrative regions (Environmental Administration 2018). The classifications of urban and rural areas for various policy purposes have been based on municipal borders. The former municipality-based classification method for urban and rural areas became problematic and outdated because many municipalities merged in 2009–2012, a change that created areas within the same municipality with urban and rural characteristics. The new grid-based classification was developed in 2014; it divides urban areas into three classes (inner, outer and peri-urban) and rural areas into four classes (local centres in rural areas, rural areas close to urban areas, rural heartland areas and sparsely populated rural areas) (Environmental Administration 2018). This urban-rural settlement typology, which depicts areal efficiency well, was used in articles II and IV.

The new urban-rural classification system uses geospatial data on population, labour, commuting, buildings, roads and land use calculated based on 250 x 250 m grids of cells (Environmental Administration 2018). The Environmental Administration
(2018) defines urban areas as urban settlements (the population centres) where the population exceeds 15,000 inhabitants. Inner urban areas are compact and densely built with high areal efficiency. Outer urban areas comprise suburbs and lower areal efficiency residential areas. Peri-urban areas are a part of intermediate zone between urban and rural and are directly linked to an urban area. Local centres in rural areas are urban settlements (population centres) located outside urban areas. Rural areas close to urban areas are characterised as locations with a rural character that are functionally connected and close to urban areas. Rural heartland areas have intensive land use with a relatively dense population and a diverse economic structure at the local level. Sparsely populated rural areas include dispersed small settlements that are located at a distance from each other and where most of the land areas are forested.

3.5.2 Urban settlements

The official definition of localities, population centres or urban settlements are defined as the clusters of buildings with at least 200 inhabitants (Statistics Finland 2018a). Urban settlements are produced, defined and delimited in co-operation with the Finnish Environment Institute. The definition utilises the building and population data of Statistics Finland’s 250 m x 250 m grid data. This categorisation of urban settlements was utilised in article II to describe a simple urban-rural typology.

3.5.3 Built environment characteristics

Green space information was extracted from the CORINE Land Cover 2018 dataset and the topographic database of National Land Survey of Finland from 2018. The former provides information on Finnish land cover and land use in raster or vector format (Finnish Environment Institute 2018) and the latter depicts the terrain of all of Finland (National Land Survey of Finland 2018b). Meadows, agricultural fields, parks, sport and recreational areas and forests in vector format were used to describe the greenness in the built environment in article IV.

3.6 METHODS

3.6.1 Linking EHRs with other data

The empirical part of my thesis comprises four research articles (Table 2). EHRs of type 2 diabetes patients from the regional electronic patient register were linked with external contextual geospatial and statistical data. This linkage is possible with geographic references. Articles I and III used patients’ postal code areas as a geographic reference and articles II and IV utilised patients’ residential addresses.
Table 2. The title, electronic health record (EHR) data, methods and main findings of each research article.

<table>
<thead>
<tr>
<th>Article</th>
<th>Title</th>
<th>EHR data</th>
<th>Methods</th>
<th>Main Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Type 2 diabetes care in North Karelia Finland: Do area-level socio-economic factors affect processes and outcomes?</td>
<td>n = 10,204 patients n = 131 postal code areas encompassing 10,067 patients</td>
<td>Principal component analysis Linear regression analysis Logistic regression analysis</td>
<td>Several small-area-level socioeconomic status (SES) factors are associated with treatment outcomes.</td>
</tr>
<tr>
<td>II</td>
<td>Do the classification of areas and distance matter to the assessment results of achieving the treatment targets among type 2 diabetes patients?</td>
<td>n = 9,606 patients</td>
<td>Chi square test Logistic regression analysis</td>
<td>It is more informative to apply a more refined area classification than a simple urban-rural dichotomy. Distance is not a barrier for achieving the treatment targets.</td>
</tr>
<tr>
<td>III</td>
<td>The usefulness of small-area-based socioeconomic characteristics in assessing the treatment outcomes of type 2 diabetes patients: a register-based mixed-effect study</td>
<td>n = 10,204 patients n = 131 postal code areas encompassing 10,067 patients</td>
<td>Mixed-effect modelling</td>
<td>Valid small-area-based SES variables (such as education) provide a useful way to predict the treatment outcomes by area.</td>
</tr>
<tr>
<td>IV</td>
<td>Association of built environment characteristics with type 2 diabetes care outcomes in North Karelia, Finland</td>
<td>n = 13,322</td>
<td>Chi square test Logistic regression analysis</td>
<td>Green land use in the patient’s residential neighbourhood does not enhance the quality of care for type 2 diabetes. Urban-rural settlement types are associated with the quality of type 2 diabetes care.</td>
</tr>
</tbody>
</table>

In articles I and III, data were joined based on postal code areas because the EHRs included in which postal code area the patient was living. For example, if a patient lived in postal code area 80100, socioeconomic information, such as educational information of that postal code area from Statistics Finland, was indicated for the patient. There were some patients who lacked the information about which postal code area they lived or there were spelling mistakes with no match to an existing postal code area. In addition, some patients lived in postal code areas with a very small population, and therefore SES data was protected. Eventually, statistical information was available for 131 postal code areas that encompassed 10,067 out of 10,204 patients.

In article II, the residential locations of patients were geocoded based on the address information available from EHRs. Address data were provided in the autumn of 2013. With geocoding, each residential address can be converted to a point on a map. From 10,204 patients, 9,606 patients were geocoded by address matching in ArcGIS based on the Finnish Transport Infrastructure Agency’s Digiroad. Geocoding was not successful for all of the original patient group members due to insufficient address information or migration out of the study region. Thus, with more exact
patient locations, it was possible to study in which urban and rural settlement types the patients lived.

Article III utilised type 2 diabetes patients’ individual socioeconomic information from Statistics Finland in addition to postal code area-level socioeconomics. The linkage of patients’ individual SES information was done by Statistics Finland, and the data were provided via a protected remote FIONA environment.

In article IV, data on type 2 diabetes patients were updated. There were 13,545 diagnosed type 2 diabetes patients alive at the end of 2017 in the health care district of Siun sote. The geocoding of the patients’ home addresses was 98.4 % successful (n = 13,322). These geocoded patients were utilised in the analyses.

### 3.6.2 Statistical methods

The data (see Tables 1 and 2) were processed statistically by principal component analysis (PCA; article I), linear regression analysis (article I), the chi-square test (article II), logistic regression analysis (articles I, II and IV) and mixed-effect modelling (article III) with IBM SPSS Statistics software (articles I, II and IV) or R environment (article III). Different analysis methods were used to identify which individual and small-area based factors were associated with type 2 diabetes prevalence at the postal code area level (article I), the process of care at the individual patient level (articles I, II and IV) or the treatment outcomes at individual patient level (articles I–IV). These methods are described in more detail in the Results section of the thesis and in the specific research articles.

In addition, age standardisation for type 2 diabetes prevalence in Figures 11–14 was performed in the R environment with the epitools package. Standardisation is a process where the rates are adjusted to take into account differences, such as age distribution, between two groups, like geographic areas in my thesis.

### 3.6.3 Neighbourhood definition

In general, three approaches exist to define neighbourhoods in research concerned with contextual effects (Cromley & McLafferty 2012: 394). First, administrative boundaries like census tracts or postal code areas can be used. Second, GIS functions can be used to create neighbourhoods. Third, individuals’ perceptions and reports of conditions in their neighbourhoods can be employed. I utilised two neighbourhood definitions in my thesis. In articles I–III, neighbourhood is based on existing administrative boundaries: postal code areas. In article IV, neighbourhood is defined as a 1-km network buffer around each type 2 diabetes patient. Network buffers are based on the distance measured along the street network.
4 RESULTS

4.1 SMALL-AREA-BASED SES FACTORS ARE ASSOCIATED WITH TYPE 2 DIABETES CARE

The first article focuses on assessing the impact of small-area level (postal code area) socioeconomic factors on the prevalence, process and treatment outcomes of type 2 diabetes care in the health care district of Siun sote, Finland. The data consist of electronic health records of diagnosed type 2 diabetes patients (n = 10,204) alive at the end of 2012 in Siun sote region (not the municipality of Heinävesi). EHRs were combined with postal code level socioeconomic variables. As postal code areas are more socioeconomically homogenous areas compared to entire municipalities, postal code areas were chosen as the main statistical unit. Statistical information was available for 131 postal code areas in the study area: these areas encompassed 10,067 patients.

The process of care was assessed by whether HbA1c was measured from the type 2 diabetes patients in 2011–2012. The outcomes of care, on the other hand, was assessed by whether the recommended level of HbA1c < 7 % (< 53 mmol/mol) was achieved among patients who were measured. Principal component analysis (PCA) was used to compress similar information of the area-level socioeconomic variables into fewer factors. Eight socioeconomic variables were used in the PCA: educated (%), median income, lower and upper clerical employee (%), unemployed (%), income less than 12,000 € (%), labour force in primary production (%), labour force in manufacturing (%) and labour force in the service sector (%). Three components resulted from the PCA. According to the PCA scores, the postal code areas were divided into three groups; they were named well-paid knowledge communities, industrial communities and idle unemployment communities. To test how the PCA components at the small-area level affect the type 2 diabetes prevalence, linear regression analysis was applied. The process and outcomes of diabetes care were assessed with a logistic regression analysis, first with neighbourhood SES factors and then with the three PCA components.

![Figure 6. Individual characteristics and socioeconomic factors that were studied in article I.](image)

First, the descriptive analysis shows that the prevalence of diagnosed type 2 diabetes in the region is 6.2% in 2012. The age-adjusted prevalence varies from 2.4% to 11.5% among postal code areas. Young age structure and advanced socioeconomic
conditions in the regional centre of North Karelia and its commuting zone are especially favourable for low type 2 diabetes prevalence. According to the results, the socioeconomic characteristics of the postal code areas, described as three PCA components, were not associated with the age-adjusted type 2 diabetes prevalence.

Second, the results indicate that the follow-up and outcomes of care among type 2 diabetes patients are moderate in the region. The levels of HbA1c are measured in over 80% of the patients, and many of these patients (72%) reach the recommended HbA1c level. The female gender is associated with higher HbA1c follow-up rates and a higher proportion who achieve the recommended HbA1c level. In addition, a younger age increases the probability of achieving the recommended HbA1c level.

Third, the results depict the role of the neighbourhood, defined as postal code area, in type 2 diabetes care. From area-level SES factors, the unemployment rate is one of the major factors that lead to lower HbA1c follow-up activity in the study region. In addition, the higher the proportion of educated people in postal code areas, the more likely the recommended HbA1c level is achieved. As stated above, the PCA resulted in three areal components. Hence, it was possible to study the dependence of both follow-up rate and HbA1c level on these components. In addition, it was possible to depict where those community types located in the study area. When analysing the process of care, it is more probable that HbA1c is not measured in industrial and idle unemployment communities. In general, the patients in the well-paid knowledge communities—which are characterised by higher rates of work participation, better income and more educated people—are more likely to achieve the recommended HbA1c level. This study reveals that several socioeconomic factors at the small-area level are related to treatment outcomes, yet the effect is not very substantial, and other factors exist that are more related to the quality of care than the ones studied here.

The study revealed that information from the regional EHRs combined with area-level socioeconomic information could be utilised to identify small areas in which individuals are at particular risk of poor diabetes care outcomes and SES groups at risk. Thus, diabetes care can be tailored and improved more effectively to regions in most need.

### 4.2 Detailed Urban-Rural Settlement Typology Reveals Areal Differences in Type 2 Diabetes Care

The study region, the health care district of Siun soite, is sparsely populated: most of the population lives reasonably close to the nearest health care centre (for example, see Figure 5). However, in rural areas the distance to the health care service provider can be rather long. There is evidence of an association between longer distance or travel time to a healthcare service provider and poorer health outcomes (Kelly et al. 2016). It is commonly believed that poor geographical accessibility to health care services might lead to delayed care and underuse of health care, especially among residents living in rural areas (Syed et al. 2013). This underuse of health care in rural areas and accessibility to health care services was the starting point for the second article.

The second article focuses on two different urban and rural classification of areas, and whether these settlement type classifications give different results of achieving the targets of process and treatment outcomes among type 2 diabetes patients in Siun soite (not the municipality of Heinävesi). These differences in the treatment targets were
explored by using two grid based classifications of urban and rural areas: a simple urban-rural dichotomy and detailed settlement typology with seven area classes.

The data of the second article consists of the same EHR data of type 2 diabetes patients as the first article, with the exception that only patients with geocodeable address (n = 9,606) were part of the analysis. Patient health records were combined with socioeconomic factors at the postal code area level, 2-class (less detailed) and 7-class (detailed) grid-based classifications of urban and rural areas depicting the built environment and accessibility measure of travel distance from patient’s home to the health care centre (Figure 7). In this study, as in the article I, achievement of the clinical treatment guidelines was assessed by the realisation of a control HbA1c measurement and the attainment of the recommended HbA1c level. In addition, whether longer distance to the health care centre reduced the achievement of the treatment guidelines was investigated. The chi-square test of independence and logistic regression analysis were applied.

Figure 7. Individual characteristics, socioeconomic factors, built environment factors and access to care that were studied in article II.

The results show that when the simple dichotomy of urban and rural is used, there are no differences in the measurement activity of HbA1c and achieving the recommended HbA1c level. Instead, the detailed 7-class classification of urban and rural areas reveals geospatial variations. The best HbA1c measurement rates are found in peri-urban areas, rural heartland areas and rural areas close to urban areas. The weakest situation in the measurement activity is in local centres in rural areas and outer urban areas. The level of education in the neighbourhood measured at postal code area level increases the probability of attendance at HbA1c screenings. The best results for achieving the recommended HbA1c level are found in outer and inner urban areas and peri-urban areas. The worst treatment outcomes are again found in local centers in rural areas and especially in sparsely populated rural areas. Furthermore, the results reveal that differences in the measurement activity of HbA1c between urban and rural areas are not due to the remote location of the rural patients because the road distance from
patient’s home to the health care centre is not a significant factor in explaining the HbA1c measurement rates. In addition, the results suggest that when the patient’s travel distance increases the patient is more likely to achieve the recommended level of HbA1c.

First, the results strongly indicate that the travel distance from the patient’s home to the health care centre is not a barrier to the realization of HbA1c control measurements or to achieve HbA1c treatment targets at individual patient level. Second, the community type matters, and this aspect is more informative and crucial to apply a more refined area classification or settlement typology than a simple urban-rural dichotomy.

4.3 VALID SMALL-AREA BASED SES FACTORS PROVIDE COST-EFFICIENT MEANS TO PREDICT TYPE 2 DIABETES TREATMENT OUTCOMES

Area-level socioeconomic variables were utilised in the first two research articles (articles I and II) when exploring how SES is associated with type 2 diabetes care quality. Both studies noted and discussed the lack of patients’ individual SES information as a limitation. Thus, the third article compared the effects of type 2 diabetes patients’ individual SES variables with the respective SES variables of postal code areas on the treatment outcomes.

Similar to articles I and II, the treatment outcomes were assessed by the patients’ latest available HbA1c value (Figure 8). This measure was used as a continuous variable. Again EHRs of type 2 diabetes patients from the regional electronic patient database were used. In addition, the patients’ individual register-based SES information (earned income, educational attainment and employment status) from Statistics Finland, and the SES information (median income, educational attainment and the proportion of the unemployed) about the population of the postal code area of the patients from Statistics Finland was utilised (Figure 8). For the analysis mixed-effect modelling was applied.

Figure 8. Individual characteristics and socioeconomic factors that were studied in article III.

The results show that an increasing age and the male gender are associated with higher HbA1c values. Furthermore, less educated patients have a higher HbA1c value, as do those living in low-educated areas. Unemployment does not affect the HbA1c value either the patient or small-area level. Income is the only predictor that provided
divergent results: high HbA1c values are associated with patients’ low incomes, but these associations are not present at the small-area level. The main finding is that the educational attainment of a neighbourhood amidst the area-based socioeconomic variables can explain a major portion of such variability in HbA1c levels that is associated with socioeconomic characteristics of a neighbourhood. Patient-based information on SES provides only a slight improvement. This main result indicated that the small-area-based information on education attainments can be almost as useful as patient-based information when assessing the socioeconomic differences in the treatment outcomes.

When assessing the treatment outcomes of type 2 diabetes patients, small-area-based SES variables provide a useful way to predict the treatment outcomes by area. This possibility of using more small-area-based data would be valuable in first-hand planning of health care services, where access to individual-level information on socioeconomic characteristics is complicated and expensive.

4.4 GREENNESS IN THE BUILT ENVIRONMENT DOES NOT ENHANCE THE QUALITY OF TYPE 2 DIABETES CARE

Evidence exists that the built environment, and especially higher levels of green space, is associated with lower diabetes risk and prevalence (den Braver et al. 2018; Dendup et al. 2018). However, it is unclear whether the built environments around people with diabetes are related to diabetes control. The fourth article investigated the associations of built environment characteristics (green land use, neighbourhood socioeconomics and urban-rural status) with the quality of type 2 diabetes care at the individual patient level. The process of care and the treatment outcomes were assessed by divergent laboratory results: glycated haemoglobin A1c (HbA1c), low-density lipoprotein (LDL) cholesterol and blood pressure (BP). The process of care was assessed by whether HbA1c, LDL and BP was measured and the outcomes of care by whether the recommended levels of HbA1c < 53 mmol/mol, LDL < 2.5 mmol/l, and BP < 140/80 mmHg were achieved.

The data for the fourth article comprises the updated type 2 diabetes patient group. These patients were alive at the end of 2017 and had their home municipality in North Karelia and in the municipality of Heinävesi. EHR data of 13,322 geocoded type 2 diabetes patients in the health care district of Siun sote was utilised. Factors that described the built environment were extracted from various databases. First, meadows, agricultural fields, parks, sport and recreational areas from the topographic database of National Land Survey of Finland from 2018 and forests from the CORINE Land Cover 2018 dataset were used to describe the greenness in the built environment. Second, given that article II found that the quality of type 2 diabetes care differs between urban and rural settlement types, the seven class urban-rural settlement type classification was also utilised. In addition, socioeconomic variables from the Grid Database of Statistics Finland were used. The following variables were extracted from the Grid Database: the percentage of people with at least an upper secondary qualification, the average age of inhabitants and percentage of unemployed from the labour force. Road network-based buffers of 1 km around patient home locations were calculated, and then the proportion of green land use and other neighbourhood measures within a 1-km network buffer was calculated. Further, 1-km network-based greenness was divided into tertiles among all patients. The chi-square test of
independence was used to study the initial associations between greenness in the built environment and type 2 diabetes process and treatment outcome indicators. Subsequently, logistic regression analyses were applied to ascertain the associations of age in four age groups, gender, obesity, green space tertiles, the percentages of educated and unemployed in the patient’s 1-km neighbourhood, the average age of residents in patient’s 1 km neighbourhood and urban-rural status on the likelihood that aforementioned process indicators were followed-up, and treatment outcomes were achieved.

The results revealed some similarities related to the measurement activity and achieving the treatment outcomes. HbA1c, LDL and BP are more likely to be measured in patients from the oldest age group (80 years and over). Patients aged 40–79 years achieve the recommended HbA1c level better than the very old (80 years and over). With LDL and BP, the situation is different: patients in the oldest age group achieve the recommended levels of LDL and BP better than patients belonging to other age groups. In sparsely populated rural areas, HbA1c and LDL are better measured from patients compared with patients in inner and outer urban areas. The greenness in the patient’s 1-km neighbourhood is not associated with the LDL or BP measurement activity or level. However, there are associations with greenness and achievement of the recommended HbA1c level. Increasing greenness, measured by tertiles, is associated with the increased odds of worse treatment balance. However, the highest green tertile mainly comprise of forest and agricultural land use, while the lowest green tertile includes more parks and recreational areas.

Although evidence exists that higher levels of green space are associated with lower diabetes risk and prevalence, this phenomenon is not the case for the quality of type 2 diabetes care in the health care district of Siun sote. The findings do not support the idea that increasing greenness in the residential neighbourhoods will enhance the quality of care among type 2 diabetes patients in the rather sparsely populated study region.

Figure 9. Individual characteristics, socioeconomic factors and built environment factors that were studied in article IV.
4.5 SUMMARISING THE RESULTS

4.5.1 Factors associated with the quality of type 2 diabetes care

Figure 10 illustrates and summaries how the individual characteristics of type 2 diabetes patients, socioeconomic factors at different levels, built environment factors and access to care that were empirically studied in this thesis are associated with the process of care and treatment outcomes. Glycated haemoglobin A1c (HbA1c) was studied in all four research articles. Additionally, article IV assessed low-density lipoprotein (LDL) cholesterol and blood pressure (BP). The associations for HbA1c are visible in the Figure 10. The green lines depict positive and red lines negative correlations.

From individual patient characteristics, the female gender is positively associated with achieving the recommended HbA1c (HbA1c < 53 mmol/l) level. This association existed in articles I, II and IV. In addition, in article III, the male gender is associated with higher HbA1c values. Increasing age is associated with better HbA1c measurement activity (articles I, II and IV) but with worse achievement of the recommended HbA1c level (articles I, II and IV) and with higher HbA1c values (article III). If the patient is obese, HbA1c is more likely above the recommended level (article IV). Less educated patients and patients with low incomes had higher HbA1c values (article III).

I used two neighbourhood definitions in this thesis: one based on administrative boundaries (postal code areas) and one based on a 1-km network buffer around each patient. Thus, neighbourhood socioeconomic factors were assessed at the postal code area level and the patient’s immediate residential neighbourhood. Postal code level socioeconomic variables were used either solely (articles I, II and III) or in the form of principal components (article I). Furthermore, socioeconomic variables in the patient’s 1-km network buffers were calculated from socioeconomic grid (250 m x 250 m) data (article IV). In general, better postal code level socioeconomic conditions are associated with better HbA1c follow-up (article I and II) or with the better achievement of the recommended HbA1c level (articles I, II and III). An increasing percentage of the unemployed in the 1-km buffer around the patient amplified the likelihood that the recommended level of HbA1c is not achieved (article IV).
Figure 10. Summary of the associations between individual characteristics, socioeconomic factors, built environment characteristics and access to care with process of care and the treatment outcomes in type 2 diabetes patients. The green lines depict positive correlations and the red lines negative correlations. The associations are statistically significant at the level of $p < 0.05$.

The settlement type where the patient lives is related to the quality of care. HbA1c is better measured from patients in sparsely populated rural areas compared to inner and outer urban areas (articles II and IV). However, patients in sparsely populated rural areas compared with patients in other settlement types were the weakest with regard to achieving the recommended treatment outcomes (articles II and IV). A longer travel distance from rural areas to the health care centre is not a barrier to balanced type 2 diabetes care (article II).

Lower green land use (meadows, agricultural fields, parks, sport and recreational areas and forests) in the patient’s residential neighbourhood was related to better achieving the recommended HbA1c level (article IV). This finding indicates that in the study region of the health care district of Siun sote increasing greenness does not necessarily enhance the quality of care.
4.5.2 Type 2 diabetes prevalence and quality of care illustrated by different areal classifications

I utilised several geospatial scales and areal classifications in articles I–IV (see Figure 3 and Table 1). Article I illustrated the type 2 diabetes prevalence at postal code level. Article II assessed the achievement of the HbA1c follow-up and treatment outcomes based on an urban-rural dichotomy as well as urban-rural settlement type classification. Additionally, article II used road distance from a patient’s home to their health care centre and thus included an idea of a service area. In general, one health care centre exists per municipality, with the exception of Joensuu commuter area and the municipalities of Joensuu and Kitee, where more health care units exist due to the consolidation of municipalities (see Figure 5). Article IV used grid-based data for the analyses. Therefore, this section of the thesis demonstrates how the type 2 diabetes prevalence and quality of care manifests itself when utilising the scales and classifications of municipality, postal code area, urban-rural settlement type classification and 2 km x 2 km grids (Figures 11–14). Municipality level and postal code area level are based on administrative boundaries, whereas urban-rural settlement type classification and 2 km x 2 km grids are based on geocoded exact patient locations. These illustrations (Figures 11–14) help to visualize and explore the geospatial variation in the prevalence and the quality of care as well as discuss the advantages and challenges of each areal classification (Table 3).

Map A in Figures 11–14 illustrates the spatial distribution of age-adjusted type 2 diabetes prevalence in the study region. Map B in Figures 11–14 shows the percentage of type 2 diabetes patients whose HbA1c was measured between 1.1.2016–31.12.2017 after the first recorded visit with type 2 diabetes diagnosis. Finally, map C in Figures 11–14 illustrates the percentage of type 2 diabetes patients who achieved the recommended HbA1c level from those whose HbA1c was measured. Only patients who had at least three months between their first diabetes diagnosis and their last HbA1c measurement were included to guarantee an appropriate period for treatment effect. The patient group in these maps are the type 2 diabetes patients who were alive at the end of 2017 and geocoded (n = 13,322). The number of type 2 diabetes patients slightly differs depending on the spatial scale or areal classification. The reasons for these differences are explained in the text.

First, Figure 11 shows the spatial distribution in 2017 at the municipality level. Age-adjusted prevalence among 13,322 type 2 diabetes patients was the highest in the municipalities of Rääkkylä (10.0), Outokumpu (9.5) and Heinävesi (9.4) (see Figure 5). The lowest prevalence was found in municipalities in the commuting zone of the regional centre: Kontiolahdi (6.7) and Joensuu (7.4). Overall, approximately 86 % of the type 2 diabetes patients had HbA1c measured (n = 11,430) in the health care district of Siun sote. In three municipalities (Juuka, Tohmajärvi and Outokumpu) HbA1c was measured in over 90 % of the type 2 diabetes patients. From the measured patients, there were 11,162 patients with an appropriate period for the treatment effect. Out of these patients, 69 % achieved the recommended HbA1c level (HbA1c < 53 mmol/mol). As a municipality, Outokumpu also performed well with regard to achieving the HbA1c control target: 75.7 % of the patients achieved the recommended level. Notably, while patients in Valtimo achieved the recommended HbA1c level well, the measurement activity was among the lowest quantile.
Figure 11. Spatial distribution of the age-adjusted prevalence (A), percentage of type 2 diabetes patients whose HbA1c was measured (B) and percentage of type 2 diabetes patients who achieved the recommended HbA1c level from those whose HbA1c was measured (C) in 2017. The data are presented by municipalities.

Next, the spatial distribution is illustrated at postal code area level (Figure 12). The age-adjustment was done for 157 postal code areas encompassing 13,288 type 2 diabetes patients (Figure 12, map A). The number of patients was smaller than at the municipality level because there were some patients with no matching postal code information with the GIS data on postal code areas or patients lived in postal code areas for which there was no information available for age groups. The postal code data on population structure is protected if the population in the area is less than 30. Age-adjusted prevalence at the postal code area level shows higher levels of geographic detail and more variation compared to the municipality level. Notably, the highest type 2 diabetes prevalence rates were found nearly in the areas of the same municipalities as where the rates were the highest at the municipality level. Again, the regional centre of the health care district of Siun sote had the lowest age-adjusted prevalence. As at the municipality level, approximately 86 % of the patients had HbA1c measured (n = 11,403) at the postal code level. The weakest measurement activity was in postal code areas in the north, south-west and middle of the Siun sote region. From the measured patients, there were 11,135 patients with an appropriate period for the treatment effect. Out of these patients, 69 % achieved the recommended HbA1c level (HbA1c < 53 mmol/mol).
Third, the spatial distribution is illustrated based on urban and rural settlement type classification. This classification divides urban areas into three classes and rural areas into four classes. Statistics Finland also provides population structure information in different age groups according to this urban-rural classification. Hence, it was possible to calculate the age-adjusted prevalence based on urban and rural settlement types. Although the prevalence map A in Figure 13 is age-adjusted, the regional centre and other urban areas with a younger population stand out as areas with the lowest type 2 diabetes prevalence. Inner urban area had the lowest (6.7) age-adjusted prevalence, while the highest (8.9) was found in sparsely populated rural areas. Measurement activity was the weakest in the regional centre and best in local centres in rural areas (88.7 %) and rural heartland areas (88.6 %). In rural areas, the population was older. The patients in sparsely populated rural areas were the worst with regard to achieving the recommended HbA1c level. The best achievement rates were in urban and rather urban areas: in the city area of Joensuu, local centres in rural areas as well as rural heartland areas.
Figure 13. Spatial distribution of the age-adjusted type 2 diabetes prevalence (A), percentage of type 2 diabetes patients whose HbA1c was measured (B) and percentage of measured type 2 diabetes patients who achieved the recommended HbA1c level (C) in 2017. The data are presented by urban-rural classification.

Finally, the spatial distribution is illustrated by 2 km x 2 km population grids. The Siun sote study region comprises 3,207 of these population grids, out of which 1,826 grids contained type 2 diabetes patients (n = 13,309). Map A in Figure 14 demonstrates more accurately—and truthfully—where type 2 diabetes patients are located and concentrated compared with the age-adjusted prevalence maps in Figures 11–13. Again, the regional centre is more visible with lower age-adjusted prevalence (more yellow and orange colour). Higher age-adjusted prevalence rates are shown in red and are found in urban settlements in the municipalities of Outokumpu, Liperi and Rääkkylä. Age-adjustment was successful for only 740 grids encompassing 11,494 type 2 diabetes patients because some grids had zero population in certain age groups. Map B in Figure 14 shows the measurement activity for 13,309 patients. HbA1c was measured in 86 % of the patients (n = 11,415). The weakest measurement activity (dark red grids) is scattered, but there is some concentration in grids in areas belonging to the municipality of Heinävesi. From the measured patients, there were 11,147 patients with an appropriate period for treatment effect. Out of these patients, 69 % achieved (n = 7,696) the recommended HbA1c level (HbA1c < 53 mmol/mol). The greener the grid, the better the recommended level is achieved among type 2 diabetes patients.
Figure 14. Spatial distribution of age-adjusted type 2 diabetes prevalence (A), percentage of type 2 diabetes patients whose HbA1c was measured (B) and percentage of measured type 2 diabetes patients who achieved the recommended HbA1c level (C) in 2017. The data are presented by 2 km x 2 km grids.
Table 3. Comparison of pros and cons among different geospatial scales and areal classifications. Municipality level and postal code area level are based on administrative boundaries whereas urban-rural settlement type classification and 2 km x 2 km grid level are based on geocoded exact patient locations.

<table>
<thead>
<tr>
<th>Geospatial scale / areal classification</th>
<th>Pros</th>
<th>Cons</th>
<th>Usage</th>
<th>Level of geographical detail</th>
</tr>
</thead>
</table>
| Municipality level                     | • Gives good overview of the clinical work as many municipalities have only one health care centre  
• It is possible to link open access socioeconomic and statistical data at the municipality level | • Administrative units where areas may differ greatly in terms of population (size, socioeconomic) and area extent | • Valid for health care service providers and decision makers to support decision making | + |
| Postal code area level                 | • It is possible to link open access socioeconomic and statistical data at postal code level  
• Functional unit that shares some similarities of the population  
• Can be used as a representative of a neighbourhood | | • Valid for health care service providers and decision makers to support decision making | ++ |
| Geocoded exact patient locations       | • Enables the choice of a desired geospatial scale | • Requires caution with patient privacy | • Valid for exploration of accessibility (travel distance / time / cost / mode) by patient and by service unit  
• Valid for optimising care and health care services | ++++ |
| Urban-rural classification             | • Enables exploration without administrative boundaries  
• Reveals settlement type | | • Valid for exploration of differences by settlement type | + |
| 2 km x 2 km grid level                 | • Enables exploration without administrative boundaries | • Requires caution with patient privacy  
• Missing data  
• Restricted access to socioeconomic and statistical data at the grid level | • Valid for more accurate identification of geographic variation | +++ |
5 DISCUSSION

5.1 MANAGEMENT OF TYPE 2 DIABETES CARE COULD BENEFIT FROM UTILISING EHRS AND GEOSPATIAL DATA

Previous studies have highlighted that combining electronic health records (EHRs) of diabetes patients with geospatial data can assist disease management at a local level and provide evidence for policymaking to improve population health (Geraghty et al. 2010; Spratt et al. 2015; Gabert et al. 2016; Richardson et al. 2017; Bilal et al. 2018b; Hirsch et al. 2018; Lê-Scherban et al. 2019). The unquestionable benefit of patient EHRs is that they include a geographical component, such as municipality, postal code area and current residential address. With these references, EHRs become geospatial data, and this information can be visualised, analysed and optimised. I linked type 2 diabetes patient EHRs with geospatial data at various geospatial scales or areal classifications in articles I, II and IV. In this synopsis it was illustrated how the type 2 diabetes prevalence and care quality is manifested by using the scales, and classifications of municipality, postal code area, urban-rural settlement type classification and 2 km x 2 km grids. These illustrations (Figures 11–14) were done to visualise and explore the areal variation in the quality of care and to discuss the advantages and challenges of each areal classification (Table 3).

Health geography and geospatial analysis of disease—encompassing disease mapping—provide useful tools for visualizing health information (Dummer 2008; Nykiforuk & Flaman 2011; Lyseen et al. 2014). Maps can act as important tools to support decision making (Dummer 2008; Nykiforuk & Flaman 2011; Lyseen et al. 2014), but the choice among the different areal units and the way the data is aggregated must be considered carefully, especially when interpreting the results (Dummer 2008). In addition, it has been suggested that researchers should include multiple area levels (Meijer et al. 2012; Kirby et al. 2017) and think of ways to modify the spatial scale to consider rural and sparsely populated areas (Macintyre et al. 2002) in neighbourhood investigations.

In general, the higher the resolution, the more detailed analysis results are revealed (Figures 11–14, Table 3.). For instance, spatial distribution by postal code areas (Figure 12) provides a higher level of geographic detail than the municipality level (Figure 11). However, the suitable areal classification depends on the purpose of use (Table 3). For example, to improve health care quality or health care service planning, the municipality can be a relevant unit because primary health care services are currently mainly provided at the municipality level in Finland. In addition, higher levels of patient privacy can be achieved when the areal units are larger. When using EHRs, it is important to consider patient privacy. Researchers might have to determine the relevant scale for analysing and reporting sensitive EHR data. Better patient privacy can be achieved at the municipality and postal code area levels because address information is not necessary. Additionally, it is not possible to distinguish individual patients from thematic maps using municipality or postal code area as a spatial scale. When age adjusting the prevalence, areal units and different age classes in them must have a large enough population. This factor may become a challenge, especially in more sparsely populated areas, when using more detailed geographical scales.
Overall, all areal classifications produced very similar results with regard to the quality of care (Figures 11–14), but every classification had its own pros and cons (Table 3). Therefore, I propose that for health care professionals, health care providers and decision makers, the areal classifications based on administrative boundaries (municipality and postal code area) are the most workable when assessing the quality of care, especially in remote settings. At least in situations such as the health care district of Siun sote, where primary health care services are provided at the municipality level, municipality is currently a relevant unit. If one needs to study more detailed geographical variation, postal code areas are relevant units. The advantages of these administrative-based units are that it is reasonably easy to get and link socioeconomic and statistical data with the units. However, if one needs or wants to observe the quality of care without administrative boundaries, GIS-based urban-rural settlement type classification is a good option. Nevertheless, I propose that for research purposes and for more accurate exploration, it is more useful to use classifications, such as grid-based classifications, that utilise exact geocoded patient locations.

The availability and use of EHRs provide a good opportunity to evaluate chronic disease monitoring, management and evidence-based health care (Birkhead et al. 2015; Namulanda et al. 2018). EHRs provide large “real-world” datasets, with data available nearly real time that comprise large comprehensive patient groups. However, one must remember that EHRs include only patients who have been diagnosed with the disease or have been treated in primary or specialised care. Another challenge for using EHRs is related to data quality issues: the records might have missing data. For example, the body mass index is more probably recorded for patients who are obese. However, when EHRs are used for research purposes, when these kinds of data quality issues are raised and when health care professionals are notified, the quality of records might improve.

The geospatial variations in the quality of care can be considered in type 2 diabetes management and when planning health care services. For example, if the treatment targets for type 2 diabetes patients defined in the Current Care Guidelines are better fulfilled in some areas compared to others, perhaps good treatment practices in these better performing areas can be transferred to weaker performing areas. On the other hand, perhaps additional resources can be targeted to the weaker performing areas. To some extent, such work has already commenced in the health care district of Siun sote, especially related to diabetes diagnostics and treatment practices (Wikström et al. 2019). However, it is important to acknowledge that when creating or utilising indicators for disease monitoring and management, the processes and factors that affect the quality of patient care are complex. For example, caution is required if the quality of care indicators are inspected mechanically to find out how a certain health care unit has performed. Comparisons between health care units can be unfair if the demographic and socioeconomic structure of the areas are not considered. The differences in the population structure in some areas, such as an aged population with older patients, can lead to a situation where the health care unit is unable to perform as well as in some other area with younger patients. Caution is especially required if the performance is a basis for incentives or financing. Regardless, the use of patient EHRs in demonstrating areal variation in the quality of care can be valuable for quality improvements in health care (Spratt et al. 2015; Wikström et al. 2019). For example, the early detection of type 2 diabetes improved in Siun sote, and the regional differences decreased after mapping the prevalence of type 2 diabetes and providing the information for health care professionals (Wikström et al. 2019).
In summary, geospatial variation in the quality of type 2 diabetes care assessed by the process of care and treatment outcomes can be observed with the help of patient EHRs. Moreover, the variation in prevalence can be studied. Information is needed on whether the care follows the treatment targets and whether some areas have a higher disease burden. After basic knowledge is provided, development actions can be planned and initiated.

5.2 SMALL-AREA-BASED SOCIOECONOMICS PROVIDE COST-EFFICIENT FIRST-HAND INFORMATION FOR HEALTH CARE SERVICE PLANNING

Individual socioeconomic characteristics (see for example Espelt et al. 2008; Agardh et al. 2011; Sundquist et al. 2011; Gary-Webb et al. 2013; Bijlsma-Rutte et al. 2018; Ibáñez et al. 2018) as well as the area-level socioeconomic characteristics (see for example Cox et al. 2007; Maier et al. 2013; Grintsova et al. 2014; Bilal et al. 2018b) are associated with the type 2 diabetes prevalence, the risk of getting the disease and the outcomes of care. Information on a patient’s socioeconomic characteristics is not recorded in EHRs. Obtaining access to individual register-based socioeconomic information is often a long and expensive process that requires many different permits. Alternatively, gathering individual socioeconomic information would require surveys, which also consume time and money. Given that EHRs include geographic references, the records can be combined with contextual geospatial data, such as socioeconomic data on a population.

When assessing the associations of socioeconomic status (SES) on the quality of care, it is convenient to use small-area-based (postal code area) SES data rather than individual SES, as discovered in article III. In addition, the results from articles I and II indicate that socioeconomic factors at the small-area level are associated with type 2 diabetes care outcomes. In Finland, small-area SES data at the postal code area level is available as free, open access datasets.

The findings from articles I–III verify that postal code area level educational attainment is associated with treatment outcomes indicated by HbA1c values (article III) or the achievement of the recommended HbA1c treatment balance (articles I and II). In postal code areas with a lower level of educational attainment, patients had higher long-term blood glucose values or were worse with regard to achieving the recommended level. This information, for example, provides cost-efficient first-hand information for health care service planning and proper resourcing. Health care service providers and decision makers could identify low socioeconomic areas based on educational attainment. Subsequently, they could target interventions for these low SES neighbourhoods or areas. Thus, one would manage to avoid using individual SES information if it is not easily available.

Other small-area SES factors that were studied in articles I–IV gave more mixed results. One reason for the more evident association between educational attainment and treatment outcomes indicated by HbA1c might be that the management of blood sugar levels requires the patient’s devotion to the care. Previous research demonstrated that a higher educational level can lead to situations when the diabetes patient understands and applies information from GPs more effectively, and thus they are better able to take appropriate care of themselves (Ayyagari et al. 2011).
To conclude, the results from articles I–III indicate that from postal code level SES factors, education at the very least is a valid factor for anticipating type 2 diabetes care outcomes. In addition, SES data on finer geographic detail is available at the grid level, as utilised in article IV, but these data are not freely available at the moment, and thus their use requires further exploration.

5.3 SHOULD HEALTH BEHAVIOURS AND EXPERIENCES BE INCLUDED IN EHRS IN THE FUTURE?

It is possible to obtain individual and area-level socioeconomic information from registers and link it with patient EHRs. A researcher can use GIS data and methods and acquire information about built environment and accessibility and link that material with patient EHRs. EHRs enable one to use a lot of data, but not everything is possible. What is currently lacking in registers are variables related to patient’s lifestyles, health behaviours and experiences with his or her health and health care.

Some information on health behaviours can be obtained from EHRs, such as smoking status and information on alcohol use based on Alcohol Use Disorders Identification Test (AUDIT). However, information on physical activity or dietary habits are not collected or recorded in a structural format. From a research point of view, it would be interesting to study the effects of health behaviours on type 2 diabetes care. If the health records are not in a structured format, it can be time consuming to further utilise them, at least from a research point of view. As stated earlier, EHRs include only patients who have been diagnosed with the disease or treated by the health care professional. Thus, even if a structured way of gathering health behaviour information exists, some patients would be missing the information.

In addition, a patient’s experience about the care is not included in EHRs. Notably, the patient experience is positively associated with clinical effectiveness (Doyle et al. 2013). Therefore, it would also be important to assess how the patient experiences the care and reports their well-being and state of health by subjective indicators. An example of these indicators are Patient Reported Outcome Measures (PROM) and Patient Reported Experience Measures (PREM).

To conclude, type 2 diabetes patients and the management of type 2 diabetes care might benefit if EHRs included structured information on the patient’s physical activity, dietary patterns and experiences with care. One way of gathering this kind of information would be using mobile applications designed for diabetes patients that would automatically send the information to the health care systems. Other ways of gathering data might be allowing patients to self-report data on a patient portal that links data to their EHRs (Casey et al. 2016) or GPs could record the information during the clinical visit.
6 CONCLUSIONS

Many studies in the past have concentrated on the risk and development of type 2 diabetes. Less is known about the progression, and management of the disease. It is important to assess the evidence-based quality of care because this factor can have a direct effect on the patient’s health and well-being. In the case of type 2 diabetes, balanced care of the disease prevents complications, comorbidities and reduces costs in health care (Rossi et al. 2011; Zoungas et al. 2012; Keng et al. 2019). In many countries, including Finland, type 2 diabetes care is based on clinical guidelines. Despite the guidelines, in Finland the real outcomes of care at the patient level and in different geographical contexts remain poorly explored. However, increasing deployment and better availability of EHRs have enabled a more straightforward assessment of quality of care of chronic diseases, such as type 2 diabetes.

This study has created a conceptual model that describes the relationship between individual characteristics, socioeconomic factors, built environment characteristics and access to care with the quality of type 2 diabetes care (Figure 3). The model helps to assess the relationships of the socioeconomic and environmental influences in the patient’s residential neighbourhood on type 2 diabetes care. The quality of care was assessed through indicators related to the process of care and treatment outcomes at the individual patient level. First, whether several register-based individual and register- or GIS-based small-area factors in the patient’s neighbourhood were associated with the quality of type 2 diabetes care was analysed empirically (articles I–IV). Second, how various geospatial scales or areal classifications can be utilised to demonstrate the spatial distribution of type 2 diabetes prevalence and quality of type 2 diabetes care was evaluated. Finally, this thesis discussed how the information about factors associated with the quality of care, and using different geospatial areal classifications, can be utilised in the management of type 2 diabetes care and health care service planning.

The process of care and treatment outcomes are two aspects of studying the quality of type 2 diabetes care. The third aspect is the cost-effectiveness of care (see Figure 3). For example, the achievement of diabetes care targets leads to lower costs for the treatment of diabetes complications (Keng et al. 2019). What is the geospatial variability of the costs to achieving the quality of care according to treatment guidelines and targets? Is there variation and what sort of variation in healthcare costs across health care units? Do certain patients in certain areas cost more and why? These aspects were not covered in my thesis, and this research stream might be the natural way to proceed in the future.

The study design in this thesis is cross-sectional. In the future, a longitudinal study design should be considered in order to study how exposures throughout the life-course may influence health outcomes as well as the quality of care. Where did the type 2 diabetes patients live earlier in life? What was the SES of the individual, family or neighbourhood earlier in life? Is there a correlation between childhood SES and treatment outcomes in adulthood? Bilal and colleagues (2018a) suggest that future studies of neighbourhood characteristics and diabetes to measure and evaluate changes in neighbourhood characteristics. Previous research revealed that socioeconomic conditions in the place of residence during childhood are associated with health outcomes later in life (Curtis et al. 2004; Monden et al. 2006). Derks and
co-workers (2017) found that socioeconomic inequalities in early life are associated with diabetes-related outcomes in adulthood. The temporal dimension is one aspect in longitudinal analysis and the other one is residential mobility. The neighbourhood around an individual’s dwelling is not the only important spatial context for interactions (Kwan 2009; van Ham & Manley 2012). Focusing only on residential neighbourhoods at a certain time can introduce uncertainty into the research results (Park & Kwan 2017). Equally important might be the places for leisure, work and places people travel through during daily routines (van Ham & Manley 2012; Kwan 2018; Mennis & Yoo 2018). However, the average age for type 2 diabetes patients is nearly 70; thus, places like worksites, would not be as relevant to study. Then again, this thesis is a register-based study, and it was not possible to obtain information about where people spend time. In addition, it was not possible to obtain a patient’s residential history from EHRs. The temporal dimension and residential mobility should be considered in future studies.

The findings of this thesis increase the understanding about the complex setting and various factors that can be related to the quality of type 2 diabetes care. The research gap between the evidence and treatment is narrowed by utilising and linking EHRs of all diagnosed type 2 diabetes patients with geospatial and other register-based data (see Figures 1 and 3) from the study region. First, area-level socioeconomic factors at the postal code level, and in the patient’s immediate neighbourhood, are associated with treatment outcomes among type 2 diabetes patients. The results indicate that valid postal code area based socioeconomic variables, such as education, provide a useful way to predict the treatment outcomes rather than using individual based socioeconomic factors. Second, when exploring urban-rural inequalities in the quality of care, a more refined settlement type classification better reveals the differences compared with a simple urban-rural dichotomy. The results also show that accessibility measured as travel distance from patient’s home to the health care centre is not a barrier to balanced type 2 diabetes care. Third, the patient’s individual characteristics, such as age and gender, are factors that relate to the quality of care. In addition to factors that are currently available in structural format in the EHRs, information related to the patient’s lifestyle and health behaviours would enhance the research on type 2 diabetes care. Fourth, EHRs of the type 2 diabetes patients—including spatial reference—have enabled the exploration of geospatial variation in the quality of care at various areal classifications (municipality, postal code area, urban-rural settlement type classification and 2 km x 2 km grids). These above mentioned findings might be useful in decision making to identify small areas or settlement types where the disease burden is high and to see whether the care is being performed according to clinical guidelines. With this information, the management of type 2 diabetes care could be more effectively tailored and improved to small areas, sub-regions and settlement types in most need and to socioeconomic groups at risk.

The findings of this thesis reveal two main points. First, when various geospatial areal classifications are utilised, it is possible to explore broadly or in more detail the disease burden and where the care is according to treatment guidelines. Varying areal classifications could be used for different purposes to support decision making when planning and managing type 2 diabetes care. Second, when combining EHRs with register- and GIS- based data, it enables the possibility to target more effective disease management to certain areas with certain population structure, such as to areas with old or low educated people. This thesis provides valuable information about the primary health care quality in the treatment of type 2 diabetes for the entire
health care district of Siun sote, in North Karelia, Finland. Similar approaches that link EHRs with register and geospatial data could be used for other diseases and regions. The use of EHRs of type 2 diabetes patients or patients suffering from other chronic diseases and combining it with contextual geospatial data and areal classifications should be used more in the evaluation of chronic disease management and disease monitoring.
REFERENCES


ARTICLES

ARTICLE I

ARTICLE II

ARTICLE III

ARTICLE IV
Toivakka, M., Tykkyläinen, M. & Laatikainen, T. (2020). Association of built environment characteristics with type 2 diabetes care outcomes in North Karelia, Finland. (manuscript)

The articles were adapted with the permission of the copyright owners. Articles II and III are open access articles and are licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/+).

Author’s contribution in the articles:

<table>
<thead>
<tr>
<th>Article</th>
<th>Study design</th>
<th>Manuscript draft</th>
<th>Data preparations</th>
<th>Methods and analysis</th>
<th>Manuscript revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article I</td>
<td>MTo, TL, JD</td>
<td>MTo</td>
<td>MTo, HT, PK, TL</td>
<td>MTo, MTy</td>
<td>MTo, MTy, HT, PK, JD, TL</td>
</tr>
<tr>
<td>Article II</td>
<td>MTo</td>
<td>MTo</td>
<td>MTo</td>
<td>MTo, TL, MTy</td>
<td>MTo, TL, MTy, TK</td>
</tr>
<tr>
<td>Article III</td>
<td>MTo, MTy, TL</td>
<td>MTo</td>
<td>MTo, AP</td>
<td>AP, LM</td>
<td>MTo, AP, LM, MTy, TL</td>
</tr>
<tr>
<td>Article IV</td>
<td>MTo, MTy, TL</td>
<td>MTo</td>
<td>MTo</td>
<td>MTo</td>
<td>MTo, MTy, TL</td>
</tr>
</tbody>
</table>

AP = Aki Pihlapuro, HT = Hilkka Tirkkonen, JD = James Dunbar, LM = Lauri Mehtätalo, MTo = Maija Toivakka, MTy = Markku Tykkyläinen, PK = Päivi Kekäläinen, TK = Timo Kumpula, TL = Tiina Laatikainen
ARTICLE I
Type 2 diabetes care in North Karelia Finland: Do area-level socio-economic factors affect processes and outcomes?

Maija Sikiö a,*, Markku Tykkyläinen a, Hilkka Tirkkonen b, Päivi Kekäläinen c, James Dunbar d, Tiina Laatikainen d, e, f

a Department of Geographical and Historical Studies, University of Eastern Finland, Joensuu, Finland
b Department of Internal Medicine, Hospital District of North Karelia, Joensuu, Finland
c Greater Green Triangle University Department of Rural Health, Flinders and Deakin Universities, Warrnambool, Australia
d Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland
e Hospital District of North Karelia, Joensuu, Finland
f Department of Chronic Disease Prevention, National Institute for Health and Welfare (THL), Helsinki, Finland

Article history: Received 17 April 2014
Received in revised form 16 July 2014
Accepted 14 September 2014
Available online 5 October 2014

Keywords:
Type 2 diabetes
Area-level socio-economic status
Quality of care
Outcomes of care
Patient database

Abstract

Aims: This research assessed the impact of area-level socio-economic factors on the prevalence and outcomes of type 2 diabetes in North Karelia, Finland.

Methods: All type 2 diabetes patients (n = 10,204) were analyzed from the regional electronic patient database during the years 2011 and 2012. The patient’s individual laboratory data was used to assess whether hemoglobin A1c (HbA1c) was measured and whether the recommended level of HbA1c <7% (<53 mmol/l) was achieved. The variables describing socio-economic characteristics of postal code areas were retrieved from the database of Statistics Finland. Linear and logistic regression analyses were used to determine associations.

Results: HbA1c had been measured in 83% of patients. Over 70% of those with HbA1c measured reached the recommended level of HbA1c. The worse the area-level socio-economic status, the more probably HbA1c was not measured. Achieving the recommended HbA1c level was associated with being female and having a better area-level socio-economic status. The age-adjusted prevalence of type 2 diabetes was not linearly dependent on the socio-economic circumstances of the postal code areas.

Conclusions: This study shows that socio-economic factors at the small area-level are associated with treatment outcomes. The information from the regional electronic patient database linked with area-level socio-economic information could be effectively utilized to improve diabetes care.

© 2014 Elsevier Ireland Ltd. All rights reserved.
1. Introduction

The prevalence of diabetes mellitus is increasing worldwide. The International Diabetes Federation estimates that the global prevalence of diabetes will be 8.8% by 2035, which means that 592 million people will be living with diabetes [1]. Most of the patients with diabetes have type 2 diabetes [2] and the diabetes rates have recently increased in parallel with the growing obesity epidemic [3,4]. Clinical management falls largely on primary health care and type 2 diabetes is expected to generate increasing economic costs for society [5].

Previous research has shown that the prevalence and the risk of getting diabetes are associated with low socio-economic circumstances [6–11]. Most studies have concentrated on type 2 diabetes but in some studies it has not been possible to distinguish diabetes types [6,10]. Patients’ individual socio-economic status (SES) [9] and area-level SES [6,10–12], or a combination of the two [7,8] has been used to explain the impact of socio-economic status on diabetes.

Area-level socio-economic factors, along with other factors such as variations in clinical practices, ethnicity and poor environmental quality [11–13], are used for explaining the geographical variation of diseases. Area-level socio-economic factors are usually in the form of deprivation indices that identify the least and the most deprived or disadvantaged areas [7,8,14]. Mostly, the area-level SES is used to explain the prevalence of disease or risk of getting it [7,8,10,14]. Studies on area-level SES and their impacts on the outcomes of care are rare, though a Californian study has assessed the association between area-level SES and hemoglobin A1c (HbA1c) levels in patients with type 2 diabetes [15]. A similar research setting was used in a study on cardiovascular disease progression among adults with diabetes, which explored the association between neighborhood deprivation and HbA1c, among other cardiometabolic risk factors [16]. Individual-level SES and the association between the probability of reaching the glycemic goal defined in the form of the recommended levels of HbA1c has been examined slightly more frequently in patients with type 2 diabetes [17–19].

In Finland, Current Care Guidelines form the basis of the treatment and management of diseases and risk factors in health care [20]. For example, in the Finnish Current Care Guidelines for diabetes it is recommended that the HbA1c level should be lower than 7.0% (53 mmol/mol) and that it should be measured every 3–6 months. However, the treatment goals should be tailored for each patient based on his or her risk factors [21]. It is unknown how well the clinical guidelines are implemented in practice in different geographical contexts and what are the real outcomes of care.

As health outcomes vary spatially, the aim of this study was to find out if the socio-economic factors at the small area-level are associated with type 2 diabetes prevalence and whether they affect the outcomes of care.

2. Materials and methods

2.1. The area context and study setting

The region of North Karelia in Eastern Finland (Fig. 1, on top left) consists of 13 municipalities and by the end of the year 2012, the population of the region was 165,794, from which 49.7% were men and 50.3% were women [22]. All of the municipalities agreed to establish a common electronic patient database at the beginning of the year 2000. The regional database was implemented step-by-step beginning in 2009, and it has covered all of the municipalities in North Karelia from the beginning of the year 2011. The North Karelian IT-Center maintains the patient database, which is called Mediatri. The type 2 diabetes data used in this study are received from the database.

The IT-Center collected data from Mediatri for the years 2011 and 2012 on every patient who had a type 2 diabetes diagnosis (ICD10 code E11) with following patient information selected; the place of domicile (municipality, postal code area), date of birth, date of diagnosis, gender, laboratory data (different tests, date for tests) and all other permanent diagnoses for the patient. To guarantee the anonymity of the patients, personal identity numbers were not provided to us. The use of this data in this research was approved by the ethics committee of the North Savo Hospital District on 13.11.12.

2.2. Data

From the database, we obtained information on 10,204 patients who had type 2 diabetes and who were alive at the end of the 2012. Men comprised 52.9% (n = 5402) of the caseload and women 47.1% (n = 4802). The mean age for men was 66 years and for women, 70 years. The prevalence of type 2 diabetes in the population was 6.2% (for men 6.6% and for women 5.8%) in 2012.

The process of care was assessed by whether HbA1c was measured during the years 2011–2012 among type 2 diabetes patients after the initial diagnosis. The outcomes of care were categorized as good [HbA1c <7% (<53 mmol/mol)] and poor [HbA1c ≥7% (≥53 mmol/mol)] HbA1c levels among those patients whose HbA1c were measured. In order to specify if HbA1c was measured, the measurement had to be from the same date as the diabetes diagnosis or after the diagnosis. If the HbA1c measurement was only taken before the diabetes diagnosis, it was not included. When analyzing the attainment of the recommended HbA1c level, only patients, who had at least three months between their diabetes diagnosis and their last HbA1c measurement were included in the analyses to guarantee an appropriate period for treatment effect.

Because we had the information on each patient’s place of domicile, we joined the patients’ health and locational data to the SES information about municipalities and postal code areas. As postal code areas are more socio-economically homogenous areas than are municipalities, postal code areas were chosen as the main statistical unit. In total, only nine patients lacked the information about which postal code area they were living in.
The socio-economic attributes of the postal code areas were retrieved from the Statistics Finland [23] database. Some of the patients were living in very sparsely populated postal code areas where Statistics Finland does not provide socio-economic information because the area contains less than 100 inhabitants. Statistical information was available for 131 postal code areas that encompass 10,067 patients. This means that 137 patients were excluded when conducting the analyses of socio-economic variables. On average, there are 1246 inhabitants (median: 324) in each of the postal code areas. Ten variables were retrieved from the database to describe the socio-economic characteristics of the postal code areas (Table 1).

### Table 1 - Socio-economic variables depicting characteristics in postal code areas (N = 131).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age</td>
<td>45.5</td>
<td>5.2</td>
<td>30.5</td>
<td>55.7</td>
</tr>
<tr>
<td>Educated (%)</td>
<td>62.3</td>
<td>9.9</td>
<td>38.4</td>
<td>84.5</td>
</tr>
<tr>
<td>Mean income (thousands/€)</td>
<td>19.4</td>
<td>2.4</td>
<td>14.7</td>
<td>26.0</td>
</tr>
<tr>
<td>Median income (thousands/€)</td>
<td>16.1</td>
<td>3.1</td>
<td>10.3</td>
<td>24.4</td>
</tr>
<tr>
<td>Lower and upper clerical employee (%)</td>
<td>23.4</td>
<td>11.2</td>
<td>4.0</td>
<td>52.8</td>
</tr>
<tr>
<td>Unemployed (%)</td>
<td>7.2</td>
<td>2.3</td>
<td>3.0</td>
<td>14.4</td>
</tr>
<tr>
<td>Income &lt; 12,000 € (%)</td>
<td>35.2</td>
<td>7.9</td>
<td>17.4</td>
<td>59.2</td>
</tr>
<tr>
<td>Labor force in primary production (%)</td>
<td>19.6</td>
<td>14.2</td>
<td>0.6</td>
<td>63.4</td>
</tr>
<tr>
<td>Labor force in manufacturing (%)</td>
<td>24.1</td>
<td>8.1</td>
<td>7.3</td>
<td>46.6</td>
</tr>
<tr>
<td>Labor force in service sector (%)</td>
<td>55.1</td>
<td>13.4</td>
<td>23.5</td>
<td>84.4</td>
</tr>
</tbody>
</table>

The socio-economic attributes of the postal code areas were retrieved from the Statistics Finland [23] database. Some of the patients were living in very sparsely populated postal code areas where Statistics Finland does not provide socio-economic information because the area contains less than 100 inhabitants. Statistical information was available for 131 postal code areas that encompass 10,067 patients. This means that 137 patients were excluded when conducting the analyses of socio-economic variables. On average, there are 1246 inhabitants (median: 324) in each of the postal code areas. Ten variables were retrieved from the database to describe the socio-economic characteristics of the postal code areas (Table 1).

### 2.3. Statistical analyses

Principal component analysis was used to compress similar information of the area-level socio-economic variables into fewer factors. It was performed because socio-economic variables were highly correlated with each other and the principal component analysis is one method for handling multicollinearity [24]. The method has been used elsewhere to compress socio-economic variables into a smaller number of constructs and to create indices for health inequalities analysis [25].

Eight out of ten socio-economic variables were used in the principal component analysis. Instead of mean income, median income was used to describe the income level of the postal code area. The mean age of the population in the postal code area was not included as we used each patient's
Table 2 – Univariate logistic regression models in explaining the dependence of HbA1c measurement and recommended HbA1c level on patient’s gender, age and area-level socio-economics. Odds ratios (OR) with 95% confidence intervals (CI) are presented in the table. Statistical significance with a significance level of p < 0.05 is marked in bold.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Is HbA1c measured? (0 = no, 1 = yes)</th>
<th>HbA1c level (0 = 7% and over, 1 = less 7%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gender in total:</td>
<td>Male</td>
</tr>
<tr>
<td>Gender (0 = male, 1 = female)</td>
<td>1.15 (1.04–1.28)</td>
<td>N/A</td>
</tr>
<tr>
<td>Age</td>
<td>1.02 (1.02–1.03)</td>
<td>1.01 (1.02–1.03)</td>
</tr>
<tr>
<td>Educated (%)</td>
<td>1.00 (0.99–1.01)</td>
<td>1.00 (0.99–1.01)</td>
</tr>
<tr>
<td>Mean income (thousands/€)</td>
<td>1.00 (0.97–1.03)</td>
<td>1.00 (0.97–1.04)</td>
</tr>
<tr>
<td>Median income (thousands/€)</td>
<td>1.00 (0.98–1.02)</td>
<td>1.01 (0.98–1.04)</td>
</tr>
<tr>
<td>Lower and upper clerical employee (%)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00 (0.99–1.01)</td>
</tr>
<tr>
<td>Unemployed (%)</td>
<td>0.94 (0.91–0.97)</td>
<td>0.95 (0.91–0.99)</td>
</tr>
<tr>
<td>Income &lt; 12,000 € (%)</td>
<td>1.00 (0.99–1.01)</td>
<td>1.00 (0.99–1.01)</td>
</tr>
<tr>
<td>Labor force in primary</td>
<td>1.00 (1.00–1.01)</td>
<td>1.00 (1.00–1.01)</td>
</tr>
<tr>
<td>Labor force in manufacturing (%)</td>
<td>1.01 (1.00–1.02)</td>
<td>1.01 (1.00–1.02)</td>
</tr>
<tr>
<td>Labor force in service sector (%)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.99 (0.99–1.00)</td>
</tr>
<tr>
<td>PC 1 (well-paid knowledge communities)</td>
<td>0.98 (0.93–1.03)</td>
<td>0.99 (0.92–1.06)</td>
</tr>
<tr>
<td>PC 2 (industrial communities)</td>
<td>0.95 (0.90–1.00)</td>
<td>0.91 (0.85–0.98)</td>
</tr>
<tr>
<td>PC 3 (idle unemployment communities)</td>
<td>0.92 (0.87–0.96)</td>
<td>0.91 (0.85–0.98)</td>
</tr>
</tbody>
</table>

age in the later logistic regression models. The analysis resulted in three principal components. Variables depicting the proportion of lower and upper clerical employees, median income of individuals, the proportion of people with at least a high school diploma or vocational training certificate and the proportion of labor force in the service sector loaded positively for the first component, whereas the proportion of labor force in primary production and the proportion of persons with annual income less than 12,000 euros had negative loadings.

The second component is comprised of the proportion of the labor force in manufacturing with a negative loading, whereas the proportion of the labor force in the service sector and the proportion of lower and upper clerical employees loaded oppositely for the second component. Finally, the proportions of the median income of individuals and persons with annual income less than 12,000 euros loaded positively for the third component, while the proportion of unemployed loaded negatively. The exact variable loadings for the components are presented in the Supplementary Table 1. The components were named after a varimax rotation based on their socio-economic attributes as defined by their variables and the types of communities that emerged were obtained by principal component scores. Both the components and the area clusters can be called well-paid knowledge communities, industrial communities and idle unemployment communities (Fig. 1, on right). All three components explain 88.1% of the variance of the variables.

To test how the SES factors at the small area-level affect the prevalence of type 2 diabetes, linear regression analysis was applied. The dependent variable was the percentage of type 2 diabetes patients, both with and without age-adjustment in the population of each postal code area. The independent variables were the three principal components.

The process of diabetes care was assessed with a logistic regression analysis. It was performed to analyze the factors that explained if HbA1c was measured or not and if HbA1c was less than 7% after three months. The former variable describes the success of glycemic monitoring in reaching the patients and the latter estimates the quality of care. Odds ratios (OR) and their confidence levels (CI) are used as a measure of the possible association. First, 14 univariate logistic regression analyses were conducted for all of the patients and genders separately (Table 2). Next the patients’ genders and ages, as well as the three principal components, were put into the same logistic model using a forward conditional method (Table 3).

3. Results

The crude prevalence of type 2 diabetes varied from 2.7% to 12.4% and the age-adjusted prevalence, as depicted in Fig.1 on middle left, varied from 2.4% to 11.5% between postal code areas in North Karelia. The crude prevalence was linearly dependent on the socio-economic circumstances of the postal code areas, as the components of well-paid knowledge communities (β = −0.865, p < 0.001) and idle unemployment communities (β = 0.405, p = 0.002) were statistically significant variables in the linear regression model. These two principal components explained 27.6% of the prevalence of type 2 diabetes at the small area-level. The components did not significantly explain the variation in the age-adjusted prevalence. The different results of the regression models revealed that type 2 diabetes is frequent primarily in aging populations that usually live in such rural areas and small communities where many inhabitants are unemployed or pensioners (Fig. 1, on right). Young age structure and advanced socio-economic conditions in the regional center and in its commuter belt were especially favorable for low prevalence.

The monitoring of HbA1c values is central in careful diagnosing diabetes and evaluating the outcomes of treatment. The HbA1c levels had been measured from eight out of ten patients (n = 8461, 82.9%) after the type 2 diabetes diagnosis. The measurement rate was slightly higher among
women (84.0%) than among men (82.0%). There were 7996 patients who had at least three months from their diagnosis to their last HbA1c measurement. Out of these patients, 71.5% (n = 5714) reached the recommended HbA1c level. The HbA1c was less than 7% for 73.1% of women and for 69.9% of men.

Table 2 presents the results of the gender differences for the two categorized variables (HbA1c measured or not, HbA1c less than 7% or higher) we were interested in. Female gender was associated with higher HbA1c follow-up rates and a higher proportion of achieving the recommended HbA1c level. From area-level SES factors, the unemployment rate in the area was negatively associated with rates of HbA1c measurement both in men and women. The neighborhood influences the reach to male patients only, as the high proportion of labor force in the manufacturing sector increased and the proportion in the service sector decreased the rate of measurement.

The SES factors affected the quality of care more than they did the rates of glycemic monitoring that reached the patients (Table 2). Neighborhood variables (better education, higher income, higher proportion of lower and upper clerical employees, higher proportion of labor force in the service sector) were associated with the better attainment of the recommended level of HbA1c. The regional center and industrial communities performed better that the periphery.

A simplified logistic regression model showed the impacts of patient’s gender, patient’s age and area-level SES components on the follow-up rate and the targeted HbA1c levels (Table 3). When including both genders and analyzing HbA1c follow-up rates, three variables remained in the model; the older the patient was, the more probable it was that the HbA1c had been measured, whereas it was more probable in industrial communities and idle unemployment communities that HbA1c was not measured. The main reasons for follow-up were factors other than those in the model, as these variables explained only 2% of the variation of the success of glycemic monitoring in the region. When both genders were analyzed separately, the situation was similar, except that the association between women and the low frequency of HbA1c measurements in industrial communities was absent.

The attainment of the recommended HbA1c level was impacted by being female, by being a younger woman, and by having better area-level socio-economic status. In general, the patients in the areas with higher rates of work participation were more likely to achieve good HbA1c values (Table 3). Again, the impacts of the five variables were very minor as these variables explain only 0.9% of the variation of the targeted HbA1c level. When both genders were analyzed separately, the patient population bifurcates. In men, only well-paid knowledge communities showed a significantly higher probability of reaching the recommended HbA1c level. However, the HbA1c among women in industrial communities was more likely to be <7%. Women become marginalized in poor conditions, as in areas characterized by idleness and unemployment, their HbA1c is significantly more likely to be 7% or over.

4. Discussion

This study reveals that the prevalence of diagnosed type 2 diabetes in North Karelia is 6.2%. Even though the National Diabetes Prevention Programme has improved the screening for undiagnosed type 2 diabetes patients [26], there still may be many undiagnosed cases, as found in an earlier Finnish study. It was discovered that in 2004, there were at least as many undiagnosed type 2 diabetes cases as there were diagnosed ones [27]. For example, among women, the prevalence of diagnosed type 2 diabetes was 4.3%, but in the study, they found an additional that 6.9% had type 2 diabetes that was not diagnosed before. This meant that the total prevalence for type 2 diabetes in women was 11.2%.

Our study reveals that high area-level prevalence of type 2 diabetes is strongly attributed to age structures in areas as selective-out migration has prevailed in North Karelia. Thus, the spatial differences in prevalence mainly result from spatial differences in demography. However, the co-variation still holds geographically, although it must be interpreted, for instance, through demographic properties. With one reservation, age, our findings of correlations between area-level prevalence and SES characteristics are in line with previous research at a larger area-level [7,10].

In Germany [7], it was found out that the socio-economic status of municipalities plays a significant part in the explanation of diabetes prevalence. The German study used the German Index of Multiple Deprivation, which included indicators on income, employment, education, municipal revenue, social capital, environment and security. Also, a French study [10] found a positive association between area-level deprivation and diabetes. The prevalence of diabetes was higher in French cantons that had a lower socio-economic status. They used demographic and socio-economic data about population density, unemployment rate and mean annual household income. Even when taking into account the
ethnic profile of the population in countries with high immigration rates, the SES differences in the prevalence of type 2 diabetes are recognized [12].

The outcomes of care among all type 2 diabetes patients are moderate in North Karelia. The levels of HbA1c had been measured in over 80% of the patients. Many of the patients (71.5%) who had at least three months from their diagnosis to their last HbA1c were at or below the recommended HbA1c level (<7%). Still there are type 2 diabetes patients whose HbA1c level has not been measured at all, even though HbA1c measurement should be performed yearly for diabetes patients [21]. It is important to monitor HbA1c levels because it has been proven that with good glycemic control, the complications among diabetes patients are rarer [28,29], resulting in lower health care costs [30].

The results in North Karelia indicate a slightly lower measurement rate than did the findings in a Swedish study where approximately 90% of the type 2 diabetes patients had HbA1c levels measured. In Sweden, less than half of the patients reached the recommended levels of HbA1c, but it needs to be acknowledged that in the Swedish study, the goal for HbA1c was ≤5% (<42 mmol/mol) [27]. If we would have used the same cut-off, taking into account the three months, only 36.8% would have reached the level of HbA1c ≤5%. Also in Scotland, over 90% of type 2 diabetes patients had a record of HbA1c, but only 44.2% attained HbA1c less than 7% [31], which means that in North Karelia the recommended levels of HbA1c are achieved more often than in Scotland.

The recommended HbA1c level of ≤7% was selected for this study based on the recommendations of Finnish guidelines. According to the ADA standards of medical care, HbA1c ≤7% is reasonable for many adults [32]. However, providers might suggest more stringent (<6.5%) or less stringent (<8%) goals for different patient groups, depending on the history of hyperglycemia, comorbidities or advanced complications. These variations in clinical practice were not taken into account in this study and even if applied, they would not change our understanding about the impacts of area-level SES differences on diabetes care.

We have demonstrated that socio-economic disparities at a small area-level have an effect on the outcomes of care, but if the local demographics are not considered, the effect is not very substantial. Also, this is in line with the Greater Sacramento area study, which found out that the association between area-level SES and glucose control was small [15], although the association was even smaller in our study. One explanation might be that the postal code areas in North Karelia are too heterogeneous according to socio-economic characteristics. Such populations, when cleaned from the impact of the age structure, are still relatively mixed without the strong spatial segregation of health, access to services and well-being. This might be the reason why area-level SES factors do not explain more of the differences.

Local unemployment was one of the major factors leading to a lower activity in the follow-up of HbA1c in North Karelia. The most probable explanation is that this reflects the very active role of occupational health care in the Finnish health service system, where the employed have easier access to health care and they are more actively followed up on through regular health checks [33]. These kinds of differences in care can easily lead to increasing health inequalities [34]. Part of the occupational health care is private in Finland and this patient data is not included in the regional patient database. Non-existence of the data from private occupational health care might dilute some of the differences. If the data would have been included, most likely, the employed patients would have even better screening rates and quality of care as those factors influence the patients’ ability to work. However, the relatively small share of private occupational health care in North Karelia most likely does not markedly influence the results.

As in Canada [35], the patient’s individual socio-economic status may affect the self-management of diabetes. For example, patients with a higher educational level can understand and apply information more efficiently, and in that way, they are better able to take appropriate care of themselves [36]. A positive association between educational level and care outcomes also existed in North Karelia. The higher the proportion of educated people in a small area, the more likely the recommended HbA1c level was achieved. We used the area-level mean values of education instead of the duration of education by individuals. The lack of data on the patients’ individual socio-economic statuses can be considered to be a limitation in this study. Thus, the results predominantly depict the role of the neighborhood in diabetes health. Recently, it has been noted that when examining the association between area-level socio-economic circumstances and diabetes or health in general, it would be important to include the patient’s individual socio-economic status into the analyses [7,8].

A strength of this study was that it included all diagnosed cases of type 2 diabetes in North Karelia, eliminating selection bias. However, there might be some deficiency in registering the diagnoses, and in that way some patients with type 2 diabetes might be missing from the data. Also, all of the municipalities use the same regional laboratory, so the HbA1c measurements were standardized. This study used statistical information on postal code areas, which are much smaller geographical units than are municipalities with some neighborhood and functional area features. The validity of the selected area unit should be kept in mind when generalizing and interpreting the results [37]. A significant benefit of the small-area analysis is that it is much cheaper to carry out than is a large scale survey.

5. Conclusions

If local socio-economic factors, and especially local demographics, are taken into account in the planning, resourcing and targeting the organizing of local health services, the productivity of that health system can be raised and the evidence-treatment gap could shrink geographically. At the practical level, feeding this information about spatially related outcomes of care back to healthcare staff would raise their awareness of the ‘at risk’ SES groups and encourage interventions at the health center level. With the patient data from an IT-Center, routine reporting could be developed for more comprehensive glycemic monitoring of
the patient population with diabetes. Patients not attending the follow-ups or those that have poor treatment levels could be identified and reminders could be sent to both the service providers and the patients. By combining the regional patient database with geospatial modeling it makes it possible to develop systems for the assessment of the quality of care with different levels of granularity.

**Conflict of interest**

The authors declare that they have no conflict of interest.

**Acknowledgments**

This work was partly funded by the Finnish Foundation for Cardiovascular Research and the rest of the funding came from the home institutions of the authors.

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.diabres.2014.09.023.

**References**


ARTICLE II
Do the classification of areas and distance matter to the assessment results of achieving the treatment targets among type 2 diabetes patients?

Maija Toivakka\textsuperscript{1*}, Tiina Laatikainen\textsuperscript{2,3,4}, Timo Kumpula\textsuperscript{1} and Markku Tykkyläinen\textsuperscript{1}

Abstract

**Background:** Type 2 diabetes is a major health concern all over the world. The prevention of diabetes is important but so is well-balanced diabetes care. Diabetes care can be influenced by individual and neighborhood socio-economic factors and geographical accessibility to health care services. The aim of the study is to find out whether two different area classifications of urban and rural areas give different area-level results of achieving the targets of control and treatment among type 2 diabetes patients exemplified by a Finnish region. The study exploits geo-referenced patient data from a regional primary health care patient database combined with postal code area-level socio-economic variables, digital road data and two grid based classifications of areas: an urban–rural dichotomy and a classification with seven area types.

**Methods:** The achievement of control and treatment targets were assessed using the patient’s individual laboratory data among 9606 type 2 diabetes patients. It was assessed whether hemoglobin A1c (HbA1c) was controlled and whether the recommended level of HbA1c was achieved in patients by different area classes and as a function of distance. Chi square test and logistic regression analysis were used for testing.

**Results:** The study reveals that area-level inequalities exist in the care of type 2 diabetes in a detailed 7-class area classification but if the simple dichotomy of urban and rural is applied differences vanish. The patient’s gender and age, area-level education and the area class they belonged to were associated with achievements of control and treatment targets. Longer distance to health care services was not a barrier to good achievements of control or treatment targets.

**Conclusions:** A more detailed grid-based area classification is better for showing spatial differences in the care of type 2 diabetes patients. Inequalities exist but it would be misleading to state that the differences are simply due to urban or rural location or due to distance. From a planning point of view findings suggest that detailed geo-coded patient information could be utilized more in resourcing and targeting the health care services to find the area-level needs of care and to improve the cost-efficient allocation of resources.

**Keywords:** Area classifications, Accessibility, Rural health, Urban health, Care outcomes, Type 2 diabetes
Introduction

Type 2 diabetes continues to be a major health burden globally [1]. The changes in lifestyle and in particular the increasing rates of obesity are affecting the increase of diabetes prevalence across the world [1–3]. The prevention of diabetes is important but so is well-balanced diabetes care. Good management of type 2 diabetes improves the quality of life of the patients, reduces complications among patients [4], decreases the risk of comorbidities [5] and reduces the economic burden [6, 7].

Socio-economic inequalities in diabetes care do exist [8]. Achievement of the treatments targets in the diabetes care are affected by individual [9, 10] and neighborhood [11–13] socio-economic status (SES). It is commonly believed that poor geographical accessibility to health care services may lead to delayed care and underuse of health care and this is believed to be the case especially among residents living in rural areas [14]. However, it should be kept in mind that rural health and health in general are interrelated with broader social, economic, political, cultural, historical [15, 16] and spatial structures.

In Finland, primary care is available to all residents and is delivered mainly in public health care centers by general practitioners (GPs). Most of the population lives reasonably close to the nearest health service provider, but in rural areas there are some long distances. In some areas, the availability of public transport is inadequate. However, some of the chronic disease patients are entitled to reimbursements for transportation to be able to attend the regular check-ups.

The aim of the study is to find out whether two different area classifications give significantly different area-level results of achieving the targets of control and treatment among type 2 diabetes patients exemplified by a Finnish region of North Karelia, equivalent in area to New Jersey or 7/10 of Belgium. The focus is to reveal and compare the possible spatial health care diversities by using 2-class (less detailed) and 7-class (detailed) grid based classifications of urban and rural areas. The first hypothesis is that the 7-class classification is better for showing differences in urban and rural areas in the care of type 2 diabetes patients. The second hypothesis claims that the longer the distance to the health center is the more it deteriorates the achievement of control and treatment targets. The study exploits individual geo-referenced type 2 diabetes patient record data from a regional primary health care patient database combined with postal code area-level socio-economic variables, digital road data and grid based classifications of areas.

Classifications of urban and rural areas

The absence of a generally accepted definition of urban and rural area types makes it difficult to examine spatial health and health care inequalities in a valid way and in particular to compare the results between different countries [17, 18]. This might be one reason for varying results on health and health care inequalities in and between urban and rural areas. It has been suggested that these inequalities should be examined across different settlement types and should not just rely on an urban–rural dichotomy [16, 19, 20].

Often the definitions of urban and rural are based on population density and distance to certain functions such as services [21, 22]. Definitions are usually developed for a certain purpose and generalization can lead to lack of explanatory variation [22]. Traditionally, countries provide the classifications of urban and rural areas based on different indicators, and usually these areas are consistent with the administrative borders such as counties, municipalities, census blocks or census tracts.

However, grid based classifications also exist. In England and Wales urban and rural areas (RUC: Rural–Urban Definition for Small Area Geographies) have been classified for policy purposes by using 1 hectare grid cells [23]. The Organization for Economic Cooperation and Development (OECD) and the European Commission have developed a grid based harmonized definition of cities in Europe which improves cross-country analysis of cities [24]. Grid based and comparable on-task tailored classifications absorb more variation than conventional classifications and thus could be more useful in health related studies.

Helminen et al. [25] have developed a grid based 7-class classification of urban and rural areas for Finland in 2014. Before that, the multiclass classifications of urban and rural areas for various policy purposes were based on municipal borders. Old classification methods became problematic and outdated as many municipalities merged in 2009–2012 creating commuter belts where both urban and rural characteristics could be identified. The new classification procedure is well documented and could be produced for other countries as well.

The 2014 Finnish area classification divides urban areas into three (inner, outer, peri-urban) classes and rural areas into four (local centers in rural areas, rural areas close to urban areas, rural heartland areas and sparsely populated rural areas) classes [26]. It depicts settlement structures focusing on population density, relative location, land use and economic structures. This classification system uses geospatial data represented by a 250 × 250 m grid of cells. Data on population, labor, commuting, buildings, roads and land use have been used. Based on the data, variables describing the amount,
density, efficiency, accessibility, intensity, versatility and orientation of the areas have been calculated. Each cell is classified into one of the seven classes according to the defined criteria. All seven area classes are found in the study region described later (Fig. 1).

The Finnish Environment Institute maintains a classification on population centers (known also as statistical locality), which is provided by Statistics Finland. All clusters of buildings with at least 200 inhabitants are defined as population centers [27]. The definition utilizes the building and population data of Statistics Finland’s 250 × 250 m grid data. The definition takes into account the population size, number of buildings and their floor area. The distance between buildings included in population centers is 200 m at maximum with certain exceptions. This categorization was also used in this study to include a simple urban–rural dichotomy (Fig. 1). The patients living in population centers reside in urban areas and the patients living outside the population centers reside in rural areas. The study region of North Karelia is more rural as 70.3 % of the population lived in population centers compared with the Finnish total urban population of 83.7 % in 2012 [28].

Accessibility of health care services
The accessibility of health care services is affected by the locations of both the health care provider and the patient. According to Penchansky and Thomas [29] accessibility (distance, transportation, travel time and cost) is one of the five dimensions of access among availability (the supply of services), accommodation (hours of operation, waiting times), affordability (price of services) and acceptability (clients’ satisfaction). The poorer the accessibility is the larger the disadvantage is made up by the friction of distance. Further on, accessibility can mean either the potential or revealed accessibility [30–32]. Potential accessibility consists of estimated values often based on surrogate variables, whereas revealed accessibility means the actual use of health care services, thus the health care service utilization [30]. Much of the research is focused on the methodology of measuring potential accessibility [31, 33–37] but less on revealed accessibility and its effects on the outcomes of care.

Transportation options, transportation costs and distance to a health care provider differ depending on the domicile of each patient. Poorer accessibility to health care services is believed to lead to poorer health outcomes [14]. Commonly it is thought that utilization decreases as distance increases. However, the effects of distance vary depending on the health service under consideration [30, 32].

Although distance may influence health care utilization, it is not a barrier to chronic care [38], and patients will travel longer distances for check-ups or for chronic conditions [39]. Mixed evidence is found in the care of diabetes. Driving distance has not been associated with care outcomes within urban settings in Canada [40], and differences in care outcomes have not been found between rural and urban patients in Australia and the USA [41, 42]. In some cases driving distance has been associated with care outcomes in rural areas in the USA [43, 44]. These diabetes related studies used very different definitions of urban and rural or did not define them at all.

Methods
Study region and data
The region of North Karelia (13 municipalities, 165,800 inhabitants, population density of 9.3/km²) in Eastern Finland is characterized by a regional center (75,000 inhabitants) providing both primary and specialized health care services and there are 24 health care centers in the region. Every patient belongs to a service area of a certain health care center based on the postal code area in which they live.

At the end of the year 2012 approximately 6 % (n = 10,294) of the population in North Karelia had a diagnosed type 2 diabetes. All patient documents have been filed to the regional patient database and individual level data (on e.g. the place of domicile, gender, age, laboratory analyses) was retrieved from this database for every patient who had type 2 diabetes diagnosis (ICD-10 code E11). The data acquisition from the care provider, approved by the ethics committee of the North Savo Hospital District, is described in detail elsewhere [13]. Individual level socio-economic information was minimal, whereupon the postal code area-level data of a patient’s domicile were utilized for describing the socio-economic environment of each patient to see in what kinds of neighborhoods they live. The socio-economic characteristics (Table 1) of the postal code areas were retrieved from the Statistics Finland’s database [45].

From the original patient group 94.1 % (n = 9606) were geocoded by address-matching in ArcGIS 10.2.1 software by using Digiroad. The Finnish Transport Agency maintains the national road and street database, Digiroad, which contains precise and accurate data on the location of all roads and streets and address information in Finland. The road distance from a patient’s home to the health care center was calculated using origin-destination cost matrix analysis provided by the software. The travel distance of each patient by road was used because patient record data does not provide information about how the patients usually travel (by car, walking, by public transport). Therefore neither travel time, the way of moving nor the real costs were available.
Achievement of the control and treatment targets

Finnish Current Care Guidelines form the basis of the treatment and management of diseases and risk factors in health care [46]. Guidelines for diabetes recommends that the hemoglobin A1c (HbA1c) level should be lower than 7.0 % (53 mmol/mol) and that it should be measured every 3–6 months in diabetes patients [47]. HbA1c provides long-term blood sugar levels and it is a good
indicator for the good quality of care. It is widely used to measure the outcomes of care \[9–13, 48\].

In this study, the achievement of the control and treatment targets were assessed by the realization of a control measurement and the achievement of the recommended HbA1c level. As the HbA1c should be measured regularly, it was assessed whether HbA1c was measured during the years 2011–2012 among the type 2 diabetes patients. Only the latest measurement was used. The clinical outcomes of care were categorized as good \([\text{HbA1c } <7 \% (<53 \text{ mmol/mol})]\) and poor \([\text{HbA1c } \geq 7 \% (\geq 53 \text{ mmol/mol})]\) HbA1c levels among those patients whose HbA1c was measured. Only patients, who had at least 3 months between their diabetes diagnosis and their last HbA1c measurement were included in the analyses to guarantee an appropriate period for treatment effect.

### Statistical analyses

First, a Chi square test \((\chi^2)\) of independence was used to compare differences in the achievement of the control and treatment targets between different area classes. At very first, it was tested whether HbA1c was measured equally often in the seven area classes. Next, we tested the differences of the measurement frequencies of patients living in and outside a population center. A similar test of independence was performed to investigate whether HbA1c was under the recommended level of less than 7 % in patients in the seven area classes and in patients living in or outside a population center.

Second, logistic regression analysis was used to test which variables affect the probability that HbA1c is measured and the probability that HbA1c is less than 7 % (dependent variables). Both the patient’s individual characteristics (patient’s age, gender and road distance from a patient’s home to the health care center) and the patient’s neighborhood characteristics (the percentage of educated people, the percentage of the unemployed and the median income) were set as independent variables. The 7-class classifications of urban and rural areas were used as independent variables so that sparsely populated rural areas was the reference category. Additionally, the dichotomous variable of living in a population center was used as an independent variable.

### Results

The average road distance from a patient’s home to the health care center was 2.1 km in population centers and 14.9 km outside them. The longest distance was 92 km. The majority (approx. 70 %) of patients were living within a 5-km radius from the health care center. Table 2 presents the results of the dependence of the achievement of the control and treatments target by the seven area classes and Table 3 presents the results by the simple dichotomous variable of urban and rural.

The best control measurement rates were found in peri-urban areas, rural heartland areas and rural areas close to urban areas (Fig. 1). In all of these classes approximately 85 % of the patients had gone through the HbA1c measurement. The weakest situation was in local centers in rural areas where HbA1c was measured in 79.9 % of the patients. The best results for HbA1c level lower than 7.0 %, were found in outer and inner urban areas and peri-urban areas. The worse outcomes of care were found again in local centers in rural areas where HbA1c was measured in 79.9 % of the patients. The best results for HbA1c level lower than 7.0 %, were found in outer and inner urban areas and peri-urban areas. The worse outcomes of care were found again in local centers in rural areas and especially in sparsely populated rural areas. The differences between the seven area classes related to the achieved treatment targets were statistically significant. However, there were no statistically significant differences in existence of the control measurement or achieving the recommended level of HbA1c when the dichotomous variable of urban and rural was under investigation (Table 3). The control measurement results (83 %) were the same for both urban and rural patients but urban patients achieved the

<table>
<thead>
<tr>
<th>Area classes</th>
<th>Patients’ mean age</th>
<th>Mean age by area</th>
<th>Educated (%) by area(^a)</th>
<th>Unemployed (%) by area</th>
<th>Median income (thousands/€) by area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner urban area</td>
<td>71.7</td>
<td>43.6</td>
<td>76.6</td>
<td>5.9</td>
<td>16.854</td>
</tr>
<tr>
<td>Outer urban area</td>
<td>65.7</td>
<td>39.1</td>
<td>76.3</td>
<td>8.2</td>
<td>18.398</td>
</tr>
<tr>
<td>Peri-urban area</td>
<td>64.9</td>
<td>36.0</td>
<td>79.3</td>
<td>5.0</td>
<td>22.740</td>
</tr>
<tr>
<td>Local centers in rural areas</td>
<td>69.9</td>
<td>48.2</td>
<td>60.6</td>
<td>7.5</td>
<td>15.817</td>
</tr>
<tr>
<td>Rural areas close to urban areas</td>
<td>67.8</td>
<td>40.7</td>
<td>70.8</td>
<td>6.3</td>
<td>18.299</td>
</tr>
<tr>
<td>Rural heartland areas</td>
<td>68.2</td>
<td>47.0</td>
<td>60.5</td>
<td>6.9</td>
<td>15.265</td>
</tr>
<tr>
<td>Sparsely populated rural areas</td>
<td>67.4</td>
<td>48.4</td>
<td>56.4</td>
<td>7.2</td>
<td>14.936</td>
</tr>
<tr>
<td>Population center = urban</td>
<td>68.7</td>
<td>44.3</td>
<td>67.1</td>
<td>7.0</td>
<td>16.941</td>
</tr>
<tr>
<td>Outside population center = rural</td>
<td>66.1</td>
<td>45.9</td>
<td>61.6</td>
<td>7.0</td>
<td>15.861</td>
</tr>
</tbody>
</table>

\(^a\) At least high school graduate or vocational training

Table 1 Patients’ mean age and area-level characteristics in different area classes
recommended level of HbA1c a little more often, but as mentioned, the result was not statistically significant.

Table 4 presents which variables affect the probability that HbA1c is measured and that HbA1c is less than 7 % tested by the logistic regression models. The variables that remained statistically significant at the level of p value below 0.05 are included in the table. The probability of HbA1c measurements increased with age-ing. The level of education in the neighborhood also increased the probability of attendance at HbA1c screenings. Compared with patients in inner urban areas, outer urban areas and local centers in rural areas, patients in sparsely populated rural areas had their HbA1c measured more often. Female gender and younger age increased the probability of achieving the recommended HbA1c level of 7 %. Surprisingly, the model suggests that when the road distance from a patient’s home to the health care center increases it is more probable that HbA1c is less than 7 %. Additionally, when sparsely populated rural areas are compared with other area classes, all the other areas perform better in achieving the recommended level of HbA1c. The dichotomous variable of population centers did not remain in the model.

The findings from logistic regression models confirm the findings from Tables 2 and 3 that a more refined area classification reveals spatial differences in the achievement of the control and treatment targets. If the patient group is merely divided into the two categories of urban and rural, no differences are found in the achievement of the control or treatment targets. If the broader, more detailed classification of urban and rural is used, differences between specific area types are observed.

Discussion
The aim of the study was to find out whether the area classifications significantly influence the area-level results of achieving the targets of control and treatment among type 2 diabetes patients. The results have been statistically tested to understand the risk in the interpretation of results in the research area. As most phenomena are geographically contingent [49], we do not aim at making generalizations about the likely transferability of findings to other regions although statistically significant results indicate that similar differences are possible to exist elsewhere.

---

**Table 2 Realization of the control measurement and achieving the recommended level of HbA1c by 7-class area classification**

<table>
<thead>
<tr>
<th>7-class classification of areas</th>
<th>Numbers of patients and their areal percentage distribution</th>
<th>Proportions of HbA1c measured patients a to the diagnosed (%)</th>
<th>Proportions of HbA1c &lt;7 % patients a to the measured (%)</th>
<th>Patients’ mean driving distances and the ranges b in km</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner urban area</td>
<td>849 (8.8 %)</td>
<td>82.8</td>
<td>74.8</td>
<td>2.0 (0–4.0)</td>
</tr>
<tr>
<td>Outer urban area</td>
<td>1433 (14.9 %)</td>
<td>80.5</td>
<td>75.6</td>
<td>2.1 (0–9.5)</td>
</tr>
<tr>
<td>Peri-urban area</td>
<td>644 (6.7 %)</td>
<td>85.6</td>
<td>74.8</td>
<td>5.0 (0.1–27.1)</td>
</tr>
<tr>
<td>Local centers in rural areas</td>
<td>1414 (14.7 %)</td>
<td>79.9</td>
<td>69.2</td>
<td>1.8 (0–5.7)</td>
</tr>
<tr>
<td>Rural areas close to urban areas</td>
<td>725 (7.5 %)</td>
<td>84.6</td>
<td>71.8</td>
<td>7.8 (0–27.9)</td>
</tr>
<tr>
<td>Rural heartland areas</td>
<td>2376 (24.7 %)</td>
<td>84.9</td>
<td>73.1</td>
<td>6.0 (0–36.0)</td>
</tr>
<tr>
<td>Sparsely populated rural areas</td>
<td>2165 (22.5 %)</td>
<td>83.5</td>
<td>66.7</td>
<td>12.1 (0–91.8)</td>
</tr>
<tr>
<td>Total</td>
<td>9606 (100 %)</td>
<td></td>
<td></td>
<td>5.9 (0–91.8)</td>
</tr>
</tbody>
</table>

* a \( p \text{ value}<0.05 

b Minimum and maximum values in brackets

**Table 3 Realization of the control measurement and achieving the recommended level of HbA1c by 2-class area classification**

<table>
<thead>
<tr>
<th>2-class classification of areas</th>
<th>Numbers of patients and their areal percentage distribution</th>
<th>Proportions of HbA1c measured patients to the diagnosed (%)</th>
<th>Proportions of HbA1c &lt;7 % patients to the measured (%)</th>
<th>Patients’ mean driving distances and the ranges a in km</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population center = urban</td>
<td>6754 (70.3 %)</td>
<td>83.0</td>
<td>72.0</td>
<td>2.1 (0–22.5)</td>
</tr>
<tr>
<td>Outside population center = rural</td>
<td>2852 (29.7 %)</td>
<td>83.0</td>
<td>70.7</td>
<td>14.9 (1–91.8)</td>
</tr>
<tr>
<td>Total</td>
<td>9606 (100 %)</td>
<td></td>
<td></td>
<td>5.9 (0–91.8)</td>
</tr>
</tbody>
</table>

* Minimum and maximum values in brackets
Our first research task was to clarify whether differences in the achievement of the control and treatment targets among type 2 diabetes patients exist in different area classes. We applied 2-class and 7-class grid based classifications of urban and rural areas. When the simple dichotomy of urban and rural was used, no differences were found in the achievement targets assessed by the realization of the control measurement and the achievement of the recommended HbA1c level. The detailed classification with seven different area classes revealed statistically significant spatial differences in the achievement of control and treatment targets. These results strongly indicate that it is more informative to apply a more refined area classification than a simple urban–rural dichotomy, as has also been suggested earlier [16, 19, 20]. The comparison of achievement of control and treatment targets in diabetes care within a detailed area classification can help to identify areas at risk in a finer scale. Different results from 2-class and 7-class classifications of urban and rural areas indicate that the classification methods and classification principles chosen can easily affect the results and conclusions. Classifications can even be contradictory. For instance, several population centers in the 2-class classification of urban areas belong to rural heartland areas or sparsely populated rural areas in the 7-class classification (Fig. 1). The different choices of areal units (whether it is based on administrative borders, grids or something else) affect the results, which should be kept in mind especially when comparing studies and results between countries.

The analyses for the testing of the second hypothesis revealed that differences in the existence of control measurement between urban and rural areas were not due to the remote location of the rural patients as the road distance from a patient’s home to the health care center was not a significant factor in explaining the control measurement rate. For the achievement of recommended HbA1c level, the model suggested that when the distance increases it is more probable that the recommended HbA1c level is achieved. This clearly states that the distance is not a barrier to good control or to achieve treatment targets.

Even though rural patients have to travel from longer distances to health care services than their urban counterparts, the distances are not that hindering in the care of type 2 diabetes in the study region. In Australia, for example, the distances can be much longer, producing a bigger barrier to the accessibility of health care [41]. In case of a need for control measurement the visits to the health care services can be planned beforehand and combined with other errands run in local centers. Moreover, the National Health Insurance scheme, which is part of the Finnish social security system, provides the reimbursement of health care related travel costs and accepts travel by taxi if a patient is unable to use a less expensive mode of transport for health reasons or if public transport is not available [50].

The analyses revealed some additional findings and issues. Firstly, regardless of the distance, inequalities in diabetes care exist, and these are partly due to a patient’s

### Table 4 Effect of patient characteristics, area-level factors and area classes on achieved treatment targets

<table>
<thead>
<tr>
<th>Variable</th>
<th>Is HbA1c measured? (0 = no, 1 = yes)</th>
<th>HbA1c level (0 = 7% and over, 1 = less 7%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (0 = male, 1 = female)</td>
<td>1.22 (1.10–1.35)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.02 (1.02–1.03)</td>
<td>0.99 (0.99–1.00)</td>
</tr>
<tr>
<td>Educated (%)</td>
<td>1.02 (1.01–1.04)</td>
<td></td>
</tr>
<tr>
<td>Unemployed (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median income (thousands €)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance (km)</td>
<td>1.01 (1.00–1.02)</td>
<td></td>
</tr>
<tr>
<td>Inner urban area</td>
<td>0.57 (0.39–0.83)</td>
<td>1.63 (1.32–2.03)</td>
</tr>
<tr>
<td>Outer urban area</td>
<td>0.56 (0.40–0.79)</td>
<td>1.64 (1.36–1.97)</td>
</tr>
<tr>
<td>Peri-urban area</td>
<td>1.53 (1.22–1.91)</td>
<td></td>
</tr>
<tr>
<td>Local centers in rural areas</td>
<td>0.67 (0.56–0.81)</td>
<td>1.23 (1.02–1.46)</td>
</tr>
<tr>
<td>Rural areas close to urban areas</td>
<td>1.33 (1.08–1.65)</td>
<td></td>
</tr>
<tr>
<td>Rural heartland areas</td>
<td>1.42 (1.23–1.65)</td>
<td></td>
</tr>
<tr>
<td>Sparsely populated rural areas</td>
<td>Reference category</td>
<td>Reference category</td>
</tr>
<tr>
<td>Pop. center (0 = outside, 1 = inside)</td>
<td>0.022</td>
<td>0.014</td>
</tr>
</tbody>
</table>

The logistic regression models revealing the effects of patient characteristics, neighborhood characteristics, area classes or the dichotomy of urban and rural on the HbA1c control measurement and the achievement of the recommended HbA1c level. The odds ratios (OR) with confidence intervals (CI) of the variables that remained statistically significant (p < 0.05) in the models are presented in the table
individual characteristics such as gender and age. Also urban and rural neighborhoods seem to matter which reflect individual characteristics and status to a high degree. However, these individual and area-level variables explain only a small part of the variation. Clearly, local differences exist but it would be misleading to state that these differences are simply due to the urban or rural location. Smith et al. in their review [51] indicate that rurality per se does not necessarily lead to rural–urban disparities, but for example it may exacerbate the effects of socio-economic disadvantages. This can be observed for example when economic growth takes place predominantly in cities leading to selective migration of the healthier. Inequalities by area types in health care mainly stem from differences in their socio-economic and demographic characteristics [16, 19, 21], originated from their socio-economic legacy and selective migration [52–54]. This can be seen to be the case in our study region as well. One of the limitations of this study was that we were not able to analyze socio-economic characteristics and the ways of life (values, norms, nutrition etc.) on an individual level as these particulars are not available in patient records.

Secondly, the service structure of the health care system and the processes of care are important factors to achieve efficient care results. Even though clinical guidelines are a national standard and largely implemented, the performance in health care is hardly ever homogeneous. Performance gaps evolve when society develops. The study indicates that combining information from different databases is cost-effective in comparison with surveys and it can be useful in planning, resourcing and targeting primary health care services. In this study, we were neither able to assess service structures or different processes of care nor to get individual socio-economic characteristics of the patients. These factors might explain one part of the differences found in the achievement of control and treatment targets in diabetes care. Also patient’s motivation to his/hers own care can play an important role in achieving the treatment targets. If such data could be inserted into patient record databases, future research could examine these factors cost-efficiently.

Conclusions

In conclusion, individual and area-level differences exist in the achievement of control and treatment targets of care of chronic conditions such as type 2 diabetes even in an area with a relatively homogenous public primary health care system. Geographical accessibility seems not to be a deteriorating factor in the care of type 2 diabetes. The patient’s gender and age, area-level education as a surrogate and the area class they belonged to were associated with achievement of control and treatment targets, and thus such information could be utilized much more in planning, resourcing and targeting the health care services. However, these factors explained only a small part of the variation. More information on the impact of processes and resources in health care and individual level characteristics are needed to obtain a comprehensive picture of factors predicting variation in the outcomes of care, but even so area-level information seems to be suggestive, at least for small-area health care planning.

Authors’ contributions

Maija Toivakka performed the analyses and wrote the manuscript. TL and MT (Markku Tykkyläinen) contributed to the analyses and their statistical interpretation. TL, MT and TK edited the manuscript. All authors read and approved the final manuscript.

Author details

1 Department of Geographical and Historical Studies, University of Eastern Finland, PO. Box 111, 80101 Joensuu, Finland. 2 Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. 3 Hospital District of North Karelia, Joensuu, Finland. 4 Department of Chronic Disease Prevention, National Institute for Health and Welfare (THL), Helsinki, Finland.

Acknowledgements

This work was partly funded by the Finnish Foundation for Cardiovascular Research, Juho Vainio Foundation and Funding from the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding. The rest of the funding came from the home institutions of the authors.

Compliance with ethical guidelines

Competing interests

The authors declare that they have no competing interests.

Received: 9 September 2015 Accepted: 15 September 2015 Published online: 30 September 2015

References


ARTICLE III
The usefulness of small-area-based socioeconomic characteristics in assessing the treatment outcomes of type 2 diabetes patients: a register-based mixed-effect study

Maija Toivakka 1*, Aki Pihlapuro 2, Markku Tykkyläinen 1, Lauri Mehtätalo 3 and Tiina Laatikainen 2,4,5

Abstract

Background: Assessment of the differences in the outcomes of care by socioeconomic status (SES) is beneficial for both the efficient targeting of health care services and to decrease health inequalities. This study compares the effects of three patient-based SES predictors (earned income, educational attainment, employment status) with three small-area-based SES predictors (median income, educational attainment, proportion of the unemployed) on the treatment outcomes of type 2 diabetes patients.

Methods: Mixed-effect modeling was applied to analyse how SES factors affect the treatment outcomes of type 2 diabetes patients. The treatment outcomes were assessed by the patients’ latest available glycated hemoglobin A1C (HbA1c) value. We used electronic health records of type 2 diabetes patients from the regional electronic patient database, the patients’ individual register-based SES information from Statistics Finland, and the SES information about the population of the postal code area of the patients from Statistics Finland.

Results: The effects of attained education on the treatment outcomes, both at the patient-level and the small-area-level are quite similar. Age and male gender were associated with higher HbA1c values and lower education indicated higher HbA1c values. Unemployment was not associated with HbA1c values at either the patient-level or the area-level. Income gave divergent results: high values of HbA1c were associated with low patient incomes but the median income of the postal code area did not predict the treatment outcomes of patients.

Conclusions: Our comparative study of three SES factors shows that the effects of attained education on the treatment outcomes are rather similar, regardless of whether patient-based or small-area-based predictors are used. Small-area-based SES variables can be a good way to overcome the absence of individual SES information, but further research is needed to find the valid small-area factors by disease. This possibility of using more small-area-based data would be valuable in health service research and first-hand planning of health care services.

Keywords: Individual-level socioeconomic status, Small-area-based socioeconomic status, Care outcomes, Type 2 diabetes mellitus, Electronic health records

* Correspondence: maija.toivakka@uef.fi
1 Department of Geographical and Historical Studies, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland
Full list of author information is available at the end of the article

© The Author(s). 2018 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background

Individual-level and area-based socioeconomic status (SES), such as income, education and occupation, have been used to examine the associations between SES and health risks in chronic disease patients. For example, previous research has shown that low individual or neighbourhood SES is associated with the risk of getting diabetes [1–3], the increased prevalence of chronic obstructive airway diseases [4], all-cause mortality in adults with atrial fibrillation [5] and increased risk of coronary heart disease [6–8]. In addition, the care of diabetes can be influenced by individual and neighbourhood SES [9, 10].

The patient’s SES information is rarely linked to public health databases or patient medical records. Thus, if the impacts of individual SES factors on care outcomes are to be assessed, then it is necessary to conduct surveys or combine information from other databases (e.g., census, educational, occupational, housing and tax records), which may not be easily accessed. Access to individual SES information often requires a cumbersome permission process due to the need to ensure information security, which consumes time and money. Socioeconomic variables by area are widely used in health research [2, 3, 5, 6, 8] and this has been suggested as a sufficiently valid and easy approach to overcome the absence of individual SES information [11, 12].

The aim of this study is to compare the predictive values of patients’ individual SES variables with the respective SES variables of postal code areas on the treatment outcomes of type 2 diabetes patients. The treatment outcomes were assessed by the patients’ latest available glycated hemoglobin A1C (HbA1c) value, which was used as an indicator of good glycemic control. We investigated whether the socioeconomic characteristics of patients are overwhelmingly more meaningful than respective SES variables of postal code areas or if they both provide similar predictive results about the influence of SES on the treatment outcomes. If the small-area-based average of SES has a predictive value, then it could be used in first-hand planning and targeting of health care services.

Methods

Patient group and glycemic control

In this study, the data consists of all diagnosed type 2 diabetes (ICD10 code E11) patients (10,204) at the end of 2012 in the region of North Karelia (13 municipalities, 165,800 inhabitants), Finland. The prevalence of type 2 diabetes in the population was 6.2% in 2012. The patient data is retrieved from the regional electronic patient database and the use of the data was approved by the ethics committee of the North Savo Hospital District. The data have a nested grouping structure with 13 municipalities, 131 postal code areas (4–33 postal code areas per municipality) and 10,204 patients, out of which 10,067 patients were able to have their postal code of residence identified (5–623 patients per postal code area).

The treatment outcomes were assessed by the patients’ latest available glycated hemoglobin A1C (HbA1c) value in the time period from 3.1.2011–16.1.2013. HbA1c provides a long-term blood sugar value and it was used as an indicator of good glycemic control. The recommended HbA1c level for good treatment balance is <7% (53 mmol/mol) based on Finnish guidelines but also according to the American Diabetes Association (ADA) standards of medical care HbA1c <7% is a reasonable goal for many adults [13]. Altogether, HbA1c measurement was found for 89.9% (n = 9172) of the patients. Out of these patients, 72.5% (n = 6652) reached the recommended HbA1c level. The average HbA1c value was 6.6 (Table 1).

Patient-based predictors

Each patient’s age, gender, earned income (€), educational attainment and employment status were used in the analysis (Table 1). The patient’s age and gender were obtained from the electronic patient database and the socioeconomic characteristics of each patient were provided by Statistics Finland via its protected remote access service, confidentially according to the Personal Data Act. Individual socioeconomic characteristics from Statistics Finland are from the end of the year 2012. Education was based on the patient’s latest highest degree and it was classified into six classes: no degree, upper secondary level education, lowest level tertiary education, lower-degree level tertiary education, higher-degree level tertiary education, and doctorate or equivalent level tertiary education. The information on whether the patient is unemployed was retrieved from Statistics Finland’s main type of activity variable. ‘Main type of activity’ describes the nature of a person’s economic activity during a year.

Small-area predictors

To measure the role of neighbourhood in the treatment outcomes, small-area-based socioeconomic variables were gathered from the 2011 Statistics Finland postal code area database. Three variables were used to describe the socioeconomic characteristics of the postal code areas: median income, the proportion of people with at least a high school diploma or vocational training, and the proportion of people unemployed (Table 1). These three variables were selected to test the predictive value of small-area-based variables for the treatment outcomes because we had patient-based corresponding variables for comparison.

Analyses

To analyse how the SES variables at the level of single patient, postal code area and municipality affect the
treatment outcome of the type 2 diabetes patients, we used the following mixed-effect model with a random intercept:

$$y_{ijk} = \beta_0 x_{ij} + \beta_1 x_{ik} + b_M^j + b_P^{ij} + e_{ijk}$$

where $y_{ijk}$ is the HbA1c value of the patient $k$ of postal code area $j$ within municipality $i$, $x_{ij}$ includes the postal code area predictors and $\beta_0 x_{ij}$ the corresponding regression coefficients, $x_{ik}$ includes the patient-based predictors and $\beta_1 x_{ik}$ the patient-based regression coefficients, $b_M^j$ is the random effect for municipality, $i$, $b_P^{ij}$ is the random effect for postal code area $j$ within municipality $i$, and $e_{ijk}$ is the residual error of patient $k$ in postal code area $j$ of municipality $i$. The random effects and residuals are assumed to be independent and normally distributed with zero means and variances $\sigma_M^2$, $\sigma_P^2$, and $\sigma^2$. The random effect is used to take into account the grouped, nested structure of the data [1-4]. More specifically, parameter $\sigma_M^2$ describes the unexplained variability in the municipality-level means of HbA1c, $\sigma_P^2$ correspondingly describes the unexplained variability of postal code area-based means around the municipality-level mean, and residual variance $\sigma^2$ describes the unexplained variability of individual observations around the postal code area-based mean. At the same time, they model the dependence of observations that belong to the same postal code area or municipality, thus allowing hypothesis testing on the fixed effects that takes into account the lack of independence among the observations from the same groups. Because the variance components are independent, the variances can be directly summed to obtain unexplained area-based variance as $\sigma_M^2 + \sigma_P^2$ and total unexplained variance as $\sigma_M^2 + \sigma_P^2 + \sigma^2$, and the corresponding standard errors as a square root of the variance. We also considered more advanced mixed-effect models with random intercept and slope, but the model with random intercept was deemed sufficient.

Several models were fitted to the dataset. The first model, the simple model (SM) included only a fixed intercept, age, gender and the random effects and residuals, providing estimates of the total variability among municipalities, postal code areas, and patients within postal code areas. The other models included additional patient-based fixed predictors (patient-based model, PBM), small-area-based predictors (area-based model, ABM) and both (combined model, CM). By comparing the estimated variances of random effects among these models, we analysed the potential of the small-area-based and patient-based predictors in explaining the variability in HbA1c. We were especially interested in whether the patient-based models or combined models had much lower total unexplained variance (i.e., the sum of the unexplained variability between municipalities, postal code areas, and patients) than the area-based model.

Results

Adding the small-area-based or patient-based socioeconomic variables to the simple model reduces the total unexplained variability (Table 2, Random part column), which confirms that there is such a component in the

<table>
<thead>
<tr>
<th>Table 1 Statistical characteristics for HbA1c value, patient-based and small-area-based data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>HbA1c</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Age (patient)</td>
</tr>
<tr>
<td>Income (patient)</td>
</tr>
<tr>
<td>Education (patient)</td>
</tr>
<tr>
<td>No degree</td>
</tr>
<tr>
<td>Upper secondary level education</td>
</tr>
<tr>
<td>Lowest level tertiary education</td>
</tr>
<tr>
<td>Lower-degree level tertiary education</td>
</tr>
<tr>
<td>Higher-degree level tertiary education</td>
</tr>
<tr>
<td>Doctorate level tertiary education</td>
</tr>
<tr>
<td>Unemployed (patient)</td>
</tr>
<tr>
<td>Median income (small-area)</td>
</tr>
<tr>
<td>Educated, % (small-area)</td>
</tr>
<tr>
<td>Unemployed, % (small-area)</td>
</tr>
</tbody>
</table>
### Table 2

Parameter estimates for simple model (SM), patient-based model (PBM), area-based model (ABM), and combined model (CM)

| Variable                  | Fixed part | | | | | Random part |
|---------------------------|------------|---|---|---|---|---|---|
|                           | SM         | PBM | ABM | CM | SM      | PBM | ABM | CM |
| **Fixed part**            |            |    |    |    |         |    |    |    |
| Age                       | 4.86 (2.68–7.04) | 2.98 (0.29–5.67) | 4.97 (2.78–7.15) | 3.06 (0.37–5.74) |
| Gender (2 = male)         | 83.00 (51.95–134.06) | 101.24 (49.16–153.32) | 83.18 (51.95–134.40) | 100.84 (48.80–152.88) |
| Income (patient)          | −5.24 (−8.01−2.46) | 0.00 | −5.05 (−7.82−2.28) | 0.00 |
| Education (patient)       | −40.18 (−71.80−8.56) | 0.00 | −35.24 (−66.93−3.56) | 0.03 |
| Unemployed (patient)      | 77.54 (−71.57−226.65) | 0.08 | 74.54 (−74.53−223.62) | 0.33 |
| Income (small-area)       | 0.02 (0.00–0.04) | 0.08 | 0.02 (0.00–0.04) | 0.10 |
| Education (small-area)    | −14.33 (−20.27−8.39) | 0.00 | −12.75 (−18.74−6.77) | 0.00 |
| Unemployed (small-area)   | 7.25 (−12.39−26.89) | 0.47 | 4.40 (−15.31−24.12) | 0.66 |

| **Random part**           |            |    |    |    |         |    |    |    |
| Municipality (M)          | 0.0859 | 0.0774 | 0.0961 | 0.0944 | 0.0761 | 0.0807 | 0.0863 | 0.0936 |
| Postal code area (Pa)     | 0.1196 | 0.1115 | 0.0483 | 0.0496 | 0.1157 | 0.1186 | 0.0467 | 0.0488 |
| Patient (P)               | 1.2237 | 1.2177 | 1.2052 | 1.2185 | 1.2177 | 1.2185 | 1.2052 | 1.2185 |
| Areas (M + Pa)            | 0.1473 | 0.1357 | 0.1076 | 0.1066 | 0.1473 | 0.1357 | 0.1076 | 0.1066 |
| Total (M + Pa + P)        | 1.2325 | 1.2252 | 1.2252 | 1.2252 | 1.2325 | 1.2252 | 1.2252 | 1.2252 |

*The fixed part presents estimated regression coefficients of the linear mixed-effect models and their 95% Confidence Intervals (95% CI) and p-values. The estimates and 95% Confidence Intervals have been multiplied by 1000 to help the interpretation process*

*The random part presents unexplained variance at different levels (in the form $se^2$)
unexplained variability of the simple model that can be explained by the socioeconomic variables. However, the component is small, only 1.2% \( \frac{(1.2325^2 - 1.2252^2) / 1.2325^2 \times 100\%}{1.2252^2 \times 100\%} = 1.2\% \) compared with the total unexplained variability in the simple model but 47\% \( \frac{(0.1473^2 - 0.1076^2) / 0.1473^2 \times 100\%}{0.1473^2 \times 100\%} = 47\% \) compared with the total unexplained variability at the area-level. The small-area predictors in the area-based model reduce the area-based unexplained variability compared with the simple model, whereas the patient-based predictors in the patient-based model explain both patient-based variability and area-based variability. Interestingly, adding the patient-based predictors to the area-based model (combined model) provides only very slight (0.3\% \( \frac{(1.2232^2 - 1.2252^2) / 1.2252^2 \times 100\%}{1.2252^2 \times 100\%} = 0.3\% \)) reduction to the total unexplained variability compared with the area-based model. This confirms that the small-area predictors alone can explain a major part of such variability in the HbA1c that is associated with the socioeconomic factors, while in comparison, patient-based information provides only a slight improvement.

The Table 2 fixed-part column describes the estimated regression coefficients of a simple mixed-effect model (SM) on age and gender, patient-based model (PBM) for patient-based predictors, area-based model (ABM) for postal code area predictors, and a combined model (CM) for both. In addition to the patient's age and male gender, which both increase the HbA1c level, less educated people have a higher HbA1c value. This effect can also be rather well explained by the proportion of people with at least a high school diploma or vocational training by area. When patient-based information on education is not used (ABM), the coefficient of the education at the level of the postal code area increases and models at least part of the variation, which is modeled through patient-based education in PBM and CM. A comparison of the coefficient of small-area-based education 14.33\times10^{-2} to the minimum and maximum education proportions in the data (0.384–0.845 Table 1) shows that it can at most explain about 0.007 unit differences in the mean HbA1c value between postal code areas, which is about 8\% (0.007/0.08318*100\% = 8\%) of the difference between the genders. The conclusion on the effects of educational factors is that either patient-based or small-area-based factors have quite similar impacts. The patient's income is also a significant predictor in PBM and CM, showing that high values of HbA1c are associated with low incomes, but this association is not present at the ABM. Unemployment does not have an effect on the HbA1c value of either the patient-level or area-level.

Discussion

In this study, we used electronic health records about type 2 diabetes patients from the regional electronic patient database, the patient's individual register-based SES information and register-based SES information by postal code area to compare the effect of patient-based and small-area-based factors of SES on the treatment outcomes. Patients' glycemic control was used as an example of treatment outcome. We tested how the patient's HbA1c value is associated with different patient-based and postal code area SES factors.

In these analyses, age and male gender were associated with higher HbA1c values and less educated patients had a higher HbA1c value, as did those living in low-educated areas. Unemployment did not have an effect on the HbA1c value of either the patient-level or small-area-level. Income was the only predictor that gave divergent results: high values of HbA1c were associated with patients' low incomes, but these associations were not present at the small-area-level.

Multilevel analysis revealed that the educational attainment of a neighbourhood amidst the area-based socioeconomic variables can explain a major part of such variability in the HbA1c that is associated with socioeconomic characteristics of a neighbourhood, while in comparison patient-based information on SES provides only a slight improvement. This means that the small-area-based information on educational attainment can be almost as useful as patient-based information when assessing the socioeconomic differences in the treatment outcomes.

There has been previous research with similar and conflicting results on the agreement between individual-level and area-based SES factors [11, 12, 15, 16]. However, this previous research has focused on health outcomes, health inequalities, or health risk factors but not on the treatment outcomes. For example, Krieger [11] compared the association of individual-level and census-based socioeconomic variables with hypertension, height, smoking, and number of full-term pregnancies. He concludes that the methodology provides a valid and useful approach to overcoming the absence of individual socioeconomic data. Domínguez-Berjón et al. [12] investigated the association between health outcomes (perceived health status, the presence of at least one chronic condition, smoking) and small-area-based socioeconomic measures, and also the association with individual socioeconomic measures. Both yielded similar results and they conclude that area-based measures can be applied to monitor health inequalities when individual information is not available. Marra et al. [15] determined the agreement between aggregate-level and individual SES factors among asthma, diabetes, and rheumatoid patients. They found that agreement between individual-level and aggregate-level SES variables may depend on patient group and in their study, individual-level variables were assumed to be better than aggregate-level variables. Pardo-Crespo et al. [16] studied the agreement...
between individual and area-level SES measures and compared the association of individual- and area-level SES measures with health outcomes (low birth weight, childhood obesity, and smoking household members) among children. They found that there was a significant disagreement between individual-level and area-level SES measures. However, these previous studies have been mainly correlative and they have not used mixed-effect models to test the explanatory power of SES variables.

In our study, we used mixed-effects models to take into account the nested grouped structure of the data into municipalities and postal code areas within municipalities. This allowed us to analyse which components of the total variability were explained by the small-area-based and patient-based predictors. It also took the dependence of the data into account in the tests of the fixed predictors. Ignoring the dependence by treating each patient as an independent observation would have led to an anti-conservative test (too small \( p \)-values) in this situation.

A strength of this study was that it included all diagnosed cases of type 2 diabetes in the region, eliminating selection bias. In addition, we used objective register-based socioeconomic information both at the patient-level and area-level gathered from Statistics Finland. One limitation of the data is that the regional patient database does not include patient data from private occupational health care. This can actually mitigate the SES differences, as employed patients, most likely, would have even better treatment outcomes. The study did not analyse lifestyles (e.g., nutrition, physical activity) or health care processes. However, these factors are not available in electronic health registers and this can be seen as one serious limitation of register-based studies.

Based on our results, when assessing the treatment outcomes of type 2 diabetes patients, small-area-based SES variables (such as education) can provide a useful way to predict the treatment outcomes by area. We could assume that this assessment method also applies to the care of other chronic conditions, but this would need more research with different patient groups and with different outcome measures. Small-area-based variables can be a good way to overcome the absence of individual SES information, as suggested previously [11, 12], but further research is needed to find more valid area-based factors. Given that individual-level data on socioeconomic characteristics are not easily available and require lengthy and expensive permission processes due to the need to ensure information security, small-area-based SES variables could be more widely used at a low cost.

Conclusions

In summary, our comparative study of three SES factors shows that the effects of attained education on the treatment outcomes are rather similar, regardless of whether individual or area predictors are used. If it is possible to target health care services on demand by area, then the use of internally valid small-area-based SES factors provides cost-efficient first-hand information for improving quality and equity in health care. This possibility of using more small-area-based data would be valuable in health service research and in planning where large diagnostic-focused patient materials are used, and access to individual-level information on socioeconomic characteristics is complicated and expensive.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABM</td>
<td>Area-based model; CM: Combined model; HbA1c: Glycated hemoglobin A1c; PBM: Patient-based model; SES: Socioeconomic status; SM: Simple model</td>
</tr>
</tbody>
</table>

Acknowledgements

Not applicable.

Funding

MiT is supported by a grant by the Finnish Cultural Foundation and the North Karelia Regional fund. TL has received funding for the research team from the Juho Vainio Foundation, the Finnish Foundation for Cardiovascular Research and the Research Committee of the Kuopio University Hospital Catchment Area for State Research Funding. The Strategic Research Council at the Academy of Finland (consortium:312703, WP4312704) funded the final stage of this research. The funders were not involved in the article preparation process.

Availability of data and materials

The datasets (patient data from health records and individual socioeconomic data) generated and/or analysed during the current study are not publicly available to protect the privacy of the patients. Socioeconomic data by postal code area is open data and accessible on: https://www.stat.fi/tup/paavo/index_en.html.

Authors' contributions

MiT contributed to the conception and design of the study, acquisition of the data and drafted the manuscript. AP contributed to the acquisition of the data and performed the data analyses. LM contributed to the data analysis and interpretation of the data. MT and TL designed the study and interpreted the data. AP, LM, MT and TL revised the manuscript. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

Ethics approval and consent to participate

The use of the data in this research was approved by the Ethics Committee of the North Savo Hospital District on 13.11.2012. This was a register-based study and consent from the patients is not needed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

1 Department of Geographical and Historical Studies, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland. 2 Joint municipal authority for North Karelia social and health services (Sisä sote), Tiikamäentie 16, FI-80210 Joensuu, Finland. 3 School of Computing, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland. 4 Institute of Public Health and Clinical Nutrition, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland. 5 Department of Public Health Solutions, National Institute for Health and Welfare (THL), P.O. Box 30, FI-00271 Helsinki, Finland.
Received: 22 May 2018 Accepted: 30 October 2018
Published online: 14 November 2018

References


ARTICLE IV
Toivakka, M., Tykkyläinen, M. & Laatikainen, T. (2020). Association of built environment characteristics with type 2 diabetes care outcomes in North Karelia, Finland. (manuscript)
Association of built environment characteristics with type 2 diabetes care outcomes in North Karelia, Finland

Toivakka, M., Tykkyläinen, M., & Laatikainen, T.

1Department of Geographical and Historical Studies, University of Eastern Finland, P.O. Box 111, FI-80101, Joensuu, Finland
2Joint municipal authority for North Karelia social and health services (Siun sote), Tikkaamäentie 16, FI-80210, Joensuu, Finland
3Institute of Public Health and Clinical Nutrition, University of Eastern Finland, P.O. Box 1627, FI-70211, Kuopio, Finland
4Department of Public Health Solutions, National Institute for Health and Welfare (THL), P.O. Box 30, FI-00271, Helsinki, Finland
*Corresponding author

Keywords: electronic health record, built environment, neighbourhood characteristics, green space, type 2 diabetes, outcomes of care, geographical information systems

1 Introduction

Interest in residential area or neighbourhood effects on health has been a popular field of study since the beginning of the 1990s in health geography, epidemiology, and public health (Macintyre et al. 1993; Macintyre et al. 2002; Diez Roux & Mair 2010; Oakes et al. 2015). The physical surroundings of neighbourhoods are referred to as built environment. Built environment is defined as the human-made built environment or environment modified by humans including features of land use, public spaces, transportation, and access to resources (Schulz & Northridge 2004; Diez Roux & Mair 2010; Piccolo et al. 2015). Studying neighbourhood greenness and how it affects human health has been of particular interest in recent years. The mechanisms behind positive associations between greenness and health are not entirely clear. Green space affects health through four possible pathways (Hartig et al. 2014; James et al. 2015; Markevych et al. 2017): opportunity for physical activity, social cohesion, stress reduction, and reduction of environmental exposures, such as air pollution, and noise. These pathways most likely intertwine (Hartig et al. 2014), and some pathways might affect more in some health conditions than the others.

Obesity is a major risk factor in the development of type 2 diabetes (Hu 2011). Green space promoting physical activity, and an active lifestyle can especially prevent diabetes (Urban green spaces and health 2016). Review by Dendup and others (2018) suggests that higher level of green space and walkability are associated with a lower risk of type 2 diabetes. Similarly, den Braver and others (2018) conclude in their review that built environment including access to green space is associated with reduced diabetes risk and prevalence. Studies on green space exposure and diabetes have found that higher levels of green space are associated with reduced risk of diabetes (Astell-Burt et al. 2014; Brown et al. 2016; Dalton et al. 2016) and lower type 2 diabetes prevalence (Bodicoat et al. 2014; Lee et al. 2017; Müller et al. 2018). However, Ihlebæk et al. (2018) found no association for type 2 diabetes and urban greenness, and Maas et al. (2009) found less strong relation between diabetes and greenness compared with other studied diseases, such as anxiety disorder and depression. In addition, Tamosiuñas and others (2014) found little or no influence on the prevalence of diabetes, coronary heart disease, and on cardiovascular risk factors, such as blood pressure, and body mass index with access to green space. These previous studies on green space and diabetes associations have bypassed to examine the possible associations of green space on the quality of diabetes care. Noteworthy is, that some recent studies has raised the
concern that very little is known about whether where a person lives is associated with diabetes control (Hirsch et al. 2018; Tabaei et al. 2018; Lê-Scherban et al. 2019).

Examination of green space within neighbourhoods in diabetes-related studies involve calculating the proportion of greenness by using either land-use data or land-cover data (Astell-Burt et al. 2014; Bodicoat et al. 2014; Dalton et al. 2016; Ihlebæk et al. 2018; Müller et al. 2018) or a Normalized Difference Vegetation Index (NDVI) derived from satellite imagery (Brown et al. 2016). NDVI have been utilized more often in studies focusing on cardiovascular disease (Pereira et al. 2012; Paquet et al. 2013). In addition, accessibility measures, such as minimum distance from the participant’s residential address to the closest park or forest, have been utilized (Müller et al. 2018).

Previous studies have varying neighbourhood definitions. The calculation of green space within neighbourhoods has within 800m (Bodicoat et al. 2014; Dalton et al. 2016), 1km (Maas et al. 2009; Astell-Burt et al. 2014), 3km (Maas et al. 2009; Bodicoat et al. 2014; Dalton et al. 2016), and 5km (Bodicoat et al. 2014; Dalton et al. 2016), search radius around participants’ home locations. Land-use and land-cover classes, such as parks, grassland, water, cemeteries, gardens and different types of forest have been used for depicting green space. The size of minimum green space (e.g. 19m2 (Müller et al. 2018), 1 ha (Tamosiunas et al. 2014)) and spatial resolution (e.g. 10x10m (Ihlebæk et al. 2018), 15x15m (Brown et al. 2016)) for calculating green space from satellite imagery have been varying or has not been reported at all. The associations have mainly been studied in an urban context (Brown et al. 2016; Ihlebæk et al. 2018; Müller et al. 2018); at times the models are stratified by urban-rural divide (Bodicoat et al. 2014; Dalton et al. 2016).

With vast differences in climatic, vegetative, and cultural factors, green space and health associations may vary across the globe (Markevych et al. 2017). Built environment, and especially green space promoting physical activity has been proved to be associated with reduced diabetes risk and prevalence. But is this the case with the quality of diabetes care? Thus, we investigated the associations of built environment factors (green land use, neighbourhood socioeconomics, and urban-rural status) with process and treatment outcomes of care among type 2 diabetes patients.

2 Data

2.1 Patient group, process and treatment outcome measures

The data consists of all diagnosed type 2 diabetes (ICD10 code E11) patients (13,545) alive at the end of 2017 in the Siun sote region (Joint municipal authority for North Karelia social and health services). The study region of Siun sote in Finland is rather green by nature as forests cover 89 % of the land area (Environment 2018). This region comprises 14 municipalities, and at the end of 2017, the population was 166,441. In 2017, the prevalence of type 2 diabetes in the population was 8.1 %. Geocoding of patients’ home addresses had a success rate of 98.4 % (n = 13,322). These geocoded patients were utilized in the analyses. Electronic health records of type 2 diabetes patients are retrieved from the regional electronic patient database. Patient information comprised age, gender, date of birth, the place of domicile (municipality, address) laboratory data, and clinical visit data. We assessed the process of care and treatment outcomes of care by divergent laboratory results and information recorded in clinical visits; glycated hemoglobin A1c (HbA1c), low-density lipoprotein (LDL) cholesterol, and blood pressure (BP). Current Care Guidelines by the Finnish Medical Society Duodecim form the basis of the treatment and management of diseases and risk factors in health care (Current Care Guidelines 2015). Current Care Guidelines for type 2 diabetes provides treatment guidelines for HbA1c, LDL, and blood pressure.
The process of care was assessed by follow-up rates. During the years 2016 and 2017, HbA1c was measured from 85.8% (n = 11,430), LDL from 80.8% (n = 10,758), and BP from 74.3% (n = 9,898) of type 2 diabetes patients. Achievement of treatment outcomes was assessed among those patients who had process indicators measured. Treatment outcomes were assessed through achieving a certain cut-off value of the indicator considering an appropriate period for treatment effect. When analysing the attainment of the recommended HbA1c level, only patients, who had at least three months between their diabetes diagnosis and their last HbA1c measurement (n = 11,162) were included in the analyses. Out of these patients, 69.0% reached the recommended HbA1c level (HbA1c < 53 mmol/mol). When analysing the attainment of the recommended LDL level, only patients, who had at least one month between the diabetes diagnosis and their last LDL measurement (n = 10,659) were included in the analyses. Out of these patients, 58.1% reached the recommended LDL level (LDL < 2.5 mmol/l). For blood pressure no period for treatment effect was applied. From the measured patients, 27.5% achieved the recommended BP level (BP < 140/80 mmHg). Overall, for 89.8% (n = 11,968) of the patients, a record for body mass index (BMI) was found or possible to count. Out of those who had BMI record, 52.7% (n = 6311) were obese (BMI ≥ 30.0 kg/m²).

### 2.2 Green space measures

We extracted green space information from the CORINE Land Cover 2018 dataset and the topographic database of National Land Survey of Finland from 2018. The former provides information on Finnish land cover and land use in raster or vector formats (Finnish Environment Institute.) and the latter depicts the terrain of all of Finland (National Land Survey of Finland.). Meadows, agricultural fields, parks, sport, and recreational areas, and forests were used for describing the land use and greenness.

### 2.3 Other built environment measures

We extracted variables depicting socioeconomic environment of patients’ neighbourhood (1 km network buffer) from Grid Database of Statistics Finland. This database contains statistical data calculated by map grid of 250 m x 250 m. We extracted the following variables: percentage of people with at least an upper secondary qualification; average age of inhabitants; and percentage of unemployed from the labour force. Data are protected in the data groups if the population in the grid is less than ten. The percentage of educated people was able to count for 1 km buffers of 11,037 patients. Average age of inhabitants in 1 km buffers was counted for 12,999 patients, and the percentage of unemployed in 1 km buffers for 9,310 patients.

We also utilized urban-rural settlement type classification. The settlement type classification divides urban areas into three classes (inner, outer, peri-urban), and rural areas into four classes (local centers in rural areas, rural areas close to urban areas, rural heartland areas, and sparsely populated rural areas) (Environmental Administration 2018).

### 3 Methods

Road network-based buffers of 1 km around patient home locations were calculated by using Network Analyst tool in ArcGIS. A shortcoming of a circular buffer is that it may not accurately represent the area that is accessible by walking, driving bicycle or car. For this reason, we used road network-based buffers. Then the proportion of green space and other neighbourhood measures within each 1 km network buffer was calculated. Further on, 1 km network-based greenness was divided into tertiles (low green, medium green, high green) among all patients (Table 1). In addition, table 1 illustrates how green space tertiles are characterized by other neighbourhood measures.
Table 1. Characteristics of green space tertiles for all patients (n = 13322) in the study region.

<table>
<thead>
<tr>
<th></th>
<th>Low green</th>
<th>Medium green</th>
<th>High green</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green space</td>
<td>&lt; 17 %</td>
<td>17-49 %</td>
<td>&gt; 49 %</td>
</tr>
<tr>
<td>Meadow, mean %</td>
<td>0.1</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>Agricultural field, mean %</td>
<td>1.1</td>
<td>6.2</td>
<td>14.8</td>
</tr>
<tr>
<td>Park, mean %</td>
<td>1.8</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Sport / recreational area, mean %</td>
<td>1.4</td>
<td>1.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Forests, mean %</td>
<td>4.0</td>
<td>20.6</td>
<td>69.9</td>
</tr>
</tbody>
</table>

Other neighbourhood measures

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Educated (%)</td>
<td>71.2</td>
<td>63.3</td>
<td>20.8</td>
</tr>
<tr>
<td>Residents’ average age</td>
<td>46.7</td>
<td>48.5</td>
<td>49.7</td>
</tr>
<tr>
<td>Unemployed residents (%)</td>
<td>20.6</td>
<td>18.5</td>
<td>13.2</td>
</tr>
</tbody>
</table>

Patient characteristics

<table>
<thead>
<tr>
<th>Gender (female)</th>
<th>Low green</th>
<th>Medium green</th>
<th>High green</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.8</td>
<td>47.2</td>
<td>38.4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age, mean</th>
<th>Low green</th>
<th>Medium green</th>
<th>High green</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>69.7</td>
<td>68.9</td>
<td>67.4</td>
</tr>
</tbody>
</table>

We used χ²-test to study the initial associations between greenness variables and type 2 diabetes process and treatment outcome indicators (Table 2). Then logistic regression analyses were performed to ascertain the effects of age, gender, obesity, green space tertiles, the percentages of educated and unemployed in patient’s 1 km neighbourhood, average age of residents in patient’s 1 km neighbourhood, and urban-rural status on the likelihood that afore-mentioned process indicators were followed-up, and treatment outcomes were achieved.

Table 2. Initial associations between greenness tertiles and type 2 diabetes process and treatment outcome indicators.

<table>
<thead>
<tr>
<th></th>
<th>Low green</th>
<th>Medium green</th>
<th>High green</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>4483</td>
<td>4433</td>
<td>4406</td>
<td></td>
</tr>
<tr>
<td>HbA1c measured</td>
<td>85.3</td>
<td>86.2</td>
<td>86.0</td>
<td>0.452</td>
</tr>
<tr>
<td>HbA1c &lt; 53mmol/mol</td>
<td>70.9</td>
<td>68.7</td>
<td>67.5</td>
<td>0.006</td>
</tr>
<tr>
<td>LDL measured</td>
<td>79.5</td>
<td>80.4</td>
<td>82.4</td>
<td>0.002</td>
</tr>
<tr>
<td>LDL &lt; 2.5mmol/l</td>
<td>58.5</td>
<td>59.4</td>
<td>56.4</td>
<td>0.036</td>
</tr>
<tr>
<td>Blood pressure measured</td>
<td>72.0</td>
<td>74.4</td>
<td>76.5</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>BP &lt; 140/80 (mmHg)</td>
<td>27.8</td>
<td>29.3</td>
<td>25.4</td>
<td>0.002</td>
</tr>
<tr>
<td>BMI measured</td>
<td>90.4</td>
<td>88.8</td>
<td>90.4</td>
<td>0.018</td>
</tr>
<tr>
<td>Obesity (BMI &gt;= 30 kg/m2)</td>
<td>50.9</td>
<td>52.8</td>
<td>54.6</td>
<td>0.004</td>
</tr>
</tbody>
</table>

*Proportion of green space in %

4 Results

Tables 3, 4 and 5 illustrate the results of the logistic regression models on the associations between built environment characteristics and process and treatment outcomes of care. For each indicator two models were fitted. Model 1 includes individual level variables (patient’s age group, gender, and information whether the patient is obese) and green space tertiles. Model 2 includes the same variables as model 1, and additionally the percentages of educated and unemployed in patient’s 1 km neighbourhood, average age of residents in patient’s 1 km neighbourhood, and urban-rural status.

HbA1c, LDL, and BP are more probably measured from patients in the oldest age group (80 years and over). Patients aged 40–79 achieve the recommended HbA1c level ($\text{HbA1c} < 53\text{mmol/mol}$) better than
the very old (80 years and over). With LDL and BP, the situation is different: patients in the oldest age group achieve the recommended levels of LDL < 2.5mmol/l and BP < 140/80 mmHg better than patients belonging to other age groups. In sparsely populated rural areas, HbA1c and LDL are better measured from patients compared with patients in inner and outer urban areas. The greenness in the patient’s 1 km neighbourhood is not associated with neither the LDL measurement activity nor the LDL level. For BP measurement activity, the association exists (model 1) but when the model is adjusted with other neighbourhood characteristics and urban-rural status (model 2) the association falls out. However, with HbA1c the negative association between greenness and achieving the recommended levels of HbA1c remain even after controlling for other neighbourhood characteristics. Next, the results from Model’s 2 are observed in more detail for HbA1c, LDL, and BP separately.

Increasing age is associated with an increased likelihood of having HbA1c measured. The beta-coefficient for age groups younger than 40 years and 40–59 years are significant and negative. From old patients (80 years and over) HbA1c is measured 1.9 times (1/0.54) more often than from patients younger than 40 years, and 2.0 times (1/0.49) more often than from patients aged 40–59 years. Gender is not related to HbA1c measurement activity. HbA1c is measured 1.6 times (1/0.63) more likely from patients in green space tertile 1 (T1) compared with patients in tertile 3 (T3). HbA1c is better measured from patients in sparsely populated rural areas than from patients in inner and outer urban areas. When assessing the achievement of recommended HbA1c level, patients aged 40–59 years are 1.32 times or 32 %, and patients aged 60–79 years are 1.54 times or 54 % more likely to achieve the recommended HbA1c level than patients 80 years and older. The odds of having HbA1c less than 53 mmol/mol is 30 % greater for females as opposed to males. If the patient is obese, HbA1c is more likely above the recommended HbA1c level. When assessing the achievement of recommended HbA1c level, the b coefficient for green space tertile 2 (T2) and green space tertile 3 (T3) are significant and negative. This indicates that increasing greenness is associated with increased odds of worse treatment balance (HbA1c >= 53mmol/mol). Patients in tertile 1 are 1.2 times (1/0.85) more likely than patients in tertile 2, and 1.6 times more likely than patients in tertile 3, going to achieve the recommended HbA1c level. When the residents’ average age increases, and when the percentage of unemployed from the labor force increase in the 1 km buffer around the patient, it is more likely that the patient does not achieve the recommended level of HbA1c. Patients in rural heartland areas are 1.26 times (or 26 %) more likely to achieve the recommended HbA1c level than patients in sparsely populated rural areas (reference category). Controlling for patient’s age group, gender, and obesity, and neighbourhood SES-factors the negative association between green space and achievement of recommended HbA1c level remain.

For LDL measurement activity, the b coefficient for age groups of younger than 40 years, and 40–59 years are significant and negative, and 60–79 years is significant and positive. This means that LDL is measured more often from patients of 80 years and older compared with patients younger than 40 years and 40–59 years. Instead, LDL is measured 72 % more likely from patients aged 60–79 years compared with old patients (80 years and over). LDL is measured more likely from patients living in local centres in rural areas or rural heartland areas compared with patients in sparsely populated rural areas. However, LDL is better measured from patients in sparsely populated rural areas than from patients in inner and outer urban areas. Gender is not related to LDL measurement activity. Patients in the oldest age group achieve the recommended LDL level better than patients belonging to other age groups. Males achieve the recommended LDL level more likely than females.

Blood pressure is more often measured from patients of 80 years and older compared with younger age groups. Same tendency exists for the achievement of the recommended blood pressure level: patients 80 years and older have blood pressure more probably less than 140/80 compared to younger patients. In addition, when the resident’s average age increases in the 1 km buffer around the patients, it is more probable that blood pressure is measured from the patients, and it is more probable that patients
achieve the recommended blood pressure level. Blood pressure is 1.65 times more likely measured from patients in peri-urban areas compared with patients in sparsely populated rural areas.

Table 3. Odds ratios (OR) and 95% confidence intervals (CI) for HbA1c measurement activity and for achieving the recommended levels of HbA1c.

<table>
<thead>
<tr>
<th>Variable</th>
<th>HbA1c measured (1 = yes)</th>
<th>HbA1c level (1 = less than 53mmol/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>N</td>
<td>11968</td>
<td>8327</td>
</tr>
<tr>
<td>Patient's agegroup 0-39</td>
<td>0.40 (0.28-0.58)</td>
<td>0.54 (0.35-0.82)</td>
</tr>
<tr>
<td>Patient's agegroup 40-59</td>
<td>0.43 (0.36-0.52)</td>
<td>0.49 (0.40-0.61)</td>
</tr>
<tr>
<td>Patient's agegroup 60-79</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Patient's agegroup 80-</td>
<td>ref</td>
<td>ref</td>
</tr>
<tr>
<td>Patient's gender (1=female)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Obesity (1=obese)</td>
<td>1.15 (1.02-1.29)</td>
<td>x</td>
</tr>
<tr>
<td>Characteristics in 1km buffer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green space T1</td>
<td>ref</td>
<td>ref</td>
</tr>
<tr>
<td>Green space T2</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Green space T3</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Educated (%)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Residents' average age</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Unemployed residents (%)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Inner urban area (UB1)</td>
<td>0.65 (0.49-0.86)</td>
<td>x</td>
</tr>
<tr>
<td>Outer urban area (UB2)</td>
<td>0.64 (0.51-0.81)</td>
<td>x</td>
</tr>
<tr>
<td>Peri-urban area (UB3)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Local centres in rural areas (UB4)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Rural areas close to urban areas (UB5)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Rural heartland areas (UB6)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Sparsely populated rural areas (UB7)</td>
<td>ref</td>
<td>ref</td>
</tr>
<tr>
<td>R2</td>
<td>0.029</td>
<td>0.040</td>
</tr>
</tbody>
</table>
Table 4. Odds ratios (OR) and 95 % confidence intervals (CI) for LDL measurement activity and for achieving the recommended levels of LDL.

<table>
<thead>
<tr>
<th>Variable</th>
<th>LDL measured (1 = yes)</th>
<th>LDL level (1 = less 2.5 mmol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>N</td>
<td>11968</td>
<td>8327</td>
</tr>
<tr>
<td>Patient’s agegroup 0-39</td>
<td>0.53 (0.39-0.73)</td>
<td>0.62 (0.44-0.89)</td>
</tr>
<tr>
<td>Patient’s agegroup 40-59</td>
<td>0.76 (0.66-0.88)</td>
<td>0.80 (0.67-0.95)</td>
</tr>
<tr>
<td>Patient’s agegroup 60-79</td>
<td>1.65 (1.45-1.88)</td>
<td>1.72 (1.48-2.01)</td>
</tr>
<tr>
<td>Patient’s agegroup 80-</td>
<td>ref</td>
<td>ref</td>
</tr>
<tr>
<td>Patient’s gender (1=female)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Obesity (1=obese)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Characteristics in 1km buffer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green space T1</td>
<td>ref</td>
<td>ref</td>
</tr>
<tr>
<td>Green space T2</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Green space T3</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Educated (%)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Residents’ average age</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Unemployed residents (%)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Inner urban area (UB1)</td>
<td>0.79 (0.64-0.99)</td>
<td>x</td>
</tr>
<tr>
<td>Outer urban area (UB2)</td>
<td>0.77 (0.63-0.94)</td>
<td>x</td>
</tr>
<tr>
<td>Peri-urban area (UB3)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Local centres in rural areas (UB4)</td>
<td>1.53 (1.23-1.90)</td>
<td>x</td>
</tr>
<tr>
<td>Rural areas close to urban areas (UB5)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Rural heartland areas (UB6)</td>
<td>1.42 (1.14-1.78)</td>
<td>x</td>
</tr>
<tr>
<td>Sparsely populated rural areas (UB7)</td>
<td>ref</td>
<td>ref</td>
</tr>
<tr>
<td>R2</td>
<td>0.028</td>
<td>0.045</td>
</tr>
</tbody>
</table>
Table 5. Odds ratios (OR) and 95 % confidence intervals (CI) for BP measurement activity and for achieving the recommended levels of BP.

<table>
<thead>
<tr>
<th>Variable</th>
<th>BP measured (1 = yes) Model 1</th>
<th>BP level (1 = less 140/80) Model 1</th>
<th>Patient's agegroup 0-39</th>
<th>Patient's agegroup 40-59</th>
<th>Patient's agegroup 60-79</th>
<th>Patient's gender (1=female)</th>
<th>Obesity (1=obese)</th>
<th>Characteristics in 1km buffer</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td></td>
<td>11968</td>
<td>8327</td>
<td>9370</td>
<td>6401</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
</tr>
<tr>
<td></td>
<td>Patient's agegroup 0-39</td>
<td>0.56 (0.40-0.78)</td>
<td>0.58 (0.40-0.84)</td>
<td>0.43 (0.28-0.64)</td>
<td>0.39 (0.25-0.62)</td>
<td></td>
<td>0.56 (0.40-0.78)</td>
<td>0.58 (0.40-0.84)</td>
<td>0.43 (0.28-0.64)</td>
</tr>
<tr>
<td></td>
<td>Patient's agegroup 40-59</td>
<td>0.41 (0.35-0.47)</td>
<td>0.42 (0.35-0.50)</td>
<td>0.41 (0.35-0.48)</td>
<td>0.41 (0.34-0.50)</td>
<td></td>
<td>0.41 (0.35-0.47)</td>
<td>0.42 (0.35-0.50)</td>
<td>0.41 (0.35-0.48)</td>
</tr>
<tr>
<td></td>
<td>Patient's agegroup 60-79</td>
<td>0.64 (0.56-0.73)</td>
<td>0.65 (0.56-0.76)</td>
<td>0.64 (0.57-0.72)</td>
<td>0.65 (0.57-0.74)</td>
<td></td>
<td>0.64 (0.56-0.73)</td>
<td>0.65 (0.56-0.76)</td>
<td>0.64 (0.57-0.72)</td>
</tr>
<tr>
<td>Patient's agegroup 80-</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
</tr>
<tr>
<td>Patient's gender (1=female)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Obesity (1=obese)</td>
<td>x</td>
<td>x</td>
<td>0.89 (0.81-0.98)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.89 (0.81-0.98)</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

5 Discussion

Evidences exist that built environment, and especially higher level of green space is associated with lower diabetes risk and prevalence (den Braver et al. 2018; Dendup et al. 2018). However, it is unclear, whether built environments around people with diabetes are related to diabetes control. Green space could affect diabetes care outcomes through providing opportunities for physical activity and through reducing stress. This study examined whether neighbourhood built environment factors (green land use, neighbourhood socioeconomics, and urban-rural status) were associated with quality of type 2 diabetes care at the individual level. Quality of care was assessed with process of care measures (HbA1c, LDL, and BP measurement activity), and with treatment outcomes of care measures (achievement of recommended HbA1c, LDL, and BP levels).

The analyses revealed some similarities related to the measurement activity and achieving the treatment outcomes of care. HbA1c, LDL, and BP are more probably measured from patients in the oldest age group (80 years and over). Patients aged 40–79 years achieve the recommended HbA1c level (HbA1c < 53mmol/mol) better than the very old (80 years and over). With LDL and BP, the situation is different: patients in the oldest age group achieve the recommended levels of LDL < 2.5mmol/l and BP < 140/80 mmHg better than patients belonging to other age groups. In sparsely populated rural areas, HbA1c and LDL are better measured from patients compared with patients in inner and outer urban areas. The greenness in the patient’s 1 km neighbourhood is not associated with neither the LDL or BP
measurement activity nor the LDL or BP level. However, associations exist with greenness and achievement of the recommended HbA1c level. Increasing greenness, measured by tertiles, is associated with increased odds of worse treatment balance (Hba1c \geq 53\text{mmol/mol}). However, it has to be noticed that the most green tertile mainly consist of forest and agricultural land use, and actually the least green tertile includes more parks and recreational areas.

A recent study has examined whether residential (zip-code level) socioeconomic, food and built environment factors (residential walkability, proportion of park and green space are, open space and recreational areas per zip code) were associated with glycemic control (HbA1c < 7 %) in a population of urban adults with diabetes in New York, US (Tabaei et al. 2018). The study found that residential socioeconomic and food environment factors and walkability were related to better glycemic control. However, no association were found for proportion of residential area that was covered by parks, and proportion of open space and outdoor recreational area. In Pennsylvania and New Jersey, US it was studied whether built environment factors (food environment, physical activity environment, and community socioeconomic deprivation) were associated with HbA1c levels over time in three community types (township, borough, city census tract) (Hirsch et al. 2018). The study found that socioeconomic deprivation, food availability and physical activity favorability were associated with HbA1c level in certain community types. However, physical activity environment including count of outdoor public parks showed no clear association. Results from previous studies indicate that greenness as such is not associated with glycemic control. We found a negative association between the greenness and achievement of recommended HbA1c value. However, our patient group comprises both urban and rural patients.

A strength of our study was that it included all diagnosed type 2 diabetes cases in the entire health care district of Siun sote, eliminating selection bias. We used type 2 diabetes electronic health record data, and objective GIS-based built environment data avoiding time-consuming surveys. We defined patient’s neighbourhood based on 1 km road network buffers around patient’s residential address. Road network buffers represents better the area that is accessible by walking, driving bicycle or car than circular buffers. In addition, our definition of neighbourhood is more truthful in describing the residential environment than using administrative units as has been done in previous research (Hirsch et al. 2018; Tabaei et al. 2018) when studying the associations of built environments and glycemic control.

Our study had some limitations. First, in order to examine the associations of greenness and quality of type 2 diabetes care more carefully, we should have information on patients’ physical activity, how they utilize the green space (forests, recreational areas, parks) near their residential areas, and how they perceive their neighbourhood environments. We lacked this information as type 2 diabetes patient electronic health record data does not include this kind of information. Second, residential self-selection may bias the results, as certain characteristics of patients may induce patients to live in certain residential areas. Finally, we studied only the current residential built environment, and cannot say anything about cumulative exposure to greenness based on residential history.

6 Conclusions

The associations of residential built environment on the quality of type 2 diabetes care should be examined by using extensive measures. Even though, evidence exist that higher level of green space is associated with lower diabetes risk and prevalence, this seems not to be the case for quality of diabetes care. Our results do not support the idea that increasing greenness in the residential neighbourhoods would enhance the quality of care among type 2 diabetes patients in the non-metropolitan type of regional structure. However, we found that certain settlement types were associated with process and treatment outcomes. Our study provides additional information about the geospatial variation of
quality of type 2 diabetes care. This information may help management of type 2 diabetes care by identifying settlement types that needs actions to improve the quality of care.

REFERENCES


Type 2 diabetes is a major health challenge globally. The quality of type 2 diabetes care can be evaluated using indicators that are based on clinical guidelines. This study links and analyses electronic health records of all diagnosed type 2 diabetes patients with geospatial and other register-based data from the health care district of Siun sote, in eastern Finland. This dissertation provides valuable information about the quality of type 2 diabetes care at different area-levels.