STAKES DISCUSSION PAPERS 1 / 2006

DISCUSSION PAPERS

KAJANTIE MIRA, MANDERBACKA KRISTIINA,
MCCALLUM ALISON, NOTKOLA IRMA-LEENA, ARFFMAN MARTTI,
FORSSAS ERJA, KARVONEN SAKARI, KORTTEINEN MATTI,
LEYLAND ALASTAIR, KESKIMÄKI ILMO

How to carry out register-based health services research in Finland?

Compiling complex study data in the REDD project

STAKES

Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskus

postimyynti: Stakes / Asiakaspalvelut PL 220, 00531 Helsinki
puhelin: (09) 3967 2190, (09) 3967 2308 (automaatti)
faksi: (09) 3967 2450 • Internet: www.stakes.fi

Kajantie Mira[1]
Manderbacka Kristiina[1]
McCallum Alison[2]
Notkola Irma-Leena[3]
Arffman Martti[1]
Forssas Erja[1]
Karvonen Sakari[4]
Kortteinen Matti[4]
Leyland Alastair[5]
Keskimäki Ilmo[1]

[1] STAKES, Outcomes and Equity Research
[2] NHS Lothian Health Board, Edinburgh
[3] Finnish Information Centre for Register Research
[4] STAKES, Welfare Research
[5] University of Glasgow, MRC Social and Public Health Sciences Unit

Correspondence:
Kristiina Manderbacka
STAKES
PO Box 220, 00531 Helsinki, Finland
E-mail: kristiina.manderbacka@stakes.fi

Pages: 82, price: 19 €

ISBN 951-33-1735-8 (nid.)
ISSN 1795-6897 (nid.)
ISBN 951-33-1865-6 (PDF)
ISSN 1795-8202 (PDF)

Valopaino Oy
Finland, Helsinki 2006

Contents

Abstract

This report uses the research project 'Regional disparities, social segregation and socioeconomic patterning: Where do inequities in access to health care arise?' (REDD) as an example of health services research that makes use of register data. The REDD project aims to evaluate the association between area variation, urbanisation, socioeconomic patterning, the development of social segregation in Finland and their contribution to the changing regional and socioeconomic gradients in health and access to health care. In this report the development and progress of the REDD project are described in order to share experiences on carrying out register-based research in health and health care. The report can be used as material for research training and teaching.

Register-based research differs from other ways of collecting data in that the only hypotheses that can be tested are those for which data already exists or can be constructed. Furthermore, the process of finalising the research questions is an iterative one, determined by the receipt of permission to use the registers selected, the process of data compilation, the extent to which data cleaning yields analysable variables and the technical limitations in developing the actual research datasets.

The report documents the planning and compilation of the REDD study data, problems encountered and solutions chosen and provides brief examples of potential uses of the data. It outlines the challenges and benefits of register-based health care research for researchers planning to use similar data. The report examines the background of the REDD project, its objectives for using register data and the impact of the data sources chosen on the planning of the study settings and the ability to establish the datasets necessary to analyse the questions posed in the research plan. The report includes a brief introduction to the legislation, the relevant statistical register authorities, the process of applying for data and the methods used to compile the data. It illustrates how the final research data sets were formed, provides examples of unexpected defects in the data and draws conclusions on the use of register data in health services research.

Tiivistelmä

Tässä raportissa käytetään ’Alueellinen ja sosiaalinen eriytyminen vai sosioekonomiset erot: Miksi terveyspalvelujen saatavuus ja käyttö on eriarvoista?’ (REDD) tutkimushanketta esimerkkinä terveydenhuoltotutkimuksesta, jossa aineistot pohjautuvat olemassa oleviin rekisteritietoihin. REDD hankkeen tavoitteena on tutkia alue-erojen, kaupungistumisen, sosioekonomisten erojen ja sosiaalisen eriytymisen sekä niiden välisten yhteyksien vaikutusta terveyden ja terveyspalvelujen käytön ja saatavuuden sosioekonomisiin ja alue-eroihin. Hankkeen eteneminen kuvataan raportissa kokemusten jakamiseksi rekisteripohjaisen terveys- ja terveydenhuoltotutkimuksen tekemisestä. Raporttia on tarkoitus käyttää materiaalina tutkijakoulutuksessa ja muussa opetuksessa.

Rekisteritietoihin pohjaava tutkimus eroaa muunlaisiin aineistoihin pohjaavasta tutkimuksesta siinä, että tutkimuksessa voidaan tarkastella vain sellaisia kysymyksiä, joiden tarkastelemiseen on jo olemassa aineistoja tai joihin aineistoja voidaan muokata. Lisäksi tutkimuskysymysten muotoilu on rekisteritutkimuksessa iteratiivinen prosessi, johon vaikuttaa lupien saaminen suunniteltujen rekisterien käyttöön, aineiston keruuprosessi, miten hyvin aineiston muokkaaminen tuottaa analysoitavissa olevia muuttujia sekä tekniset rajoitteet varsinaisten tutkimusaineistojen muodostamisessa.

Raportissa dokumentoidaan REDD aineiston suunnittelu- ja keruuprosessi, niissä kohdatut ongelmat ja ongelmien ratkaisu hankkeessa. Raportti antaa myös esimerkkejä aineistojen suunnitellusta käytöstä. Rekisteritutkimusta suunnitteleville tutkijoille kuvataan niitä haasteita ja etuja joita terveydenhuoltotutkimuksen tekeminen rekisteripohjaisten aineistojen avulla tarjoaa. Raportissa tarkastellaan REDD hankkeen taustaa, syitä rekisteripohjaisten aineistojen valintaan, valittujen aineistojen vaikutuksia tutkimusasetelmiin ja mahdollisuuksiin muodostaa suunnitelmassa hahmoteltujen kysymysten analysoimiseen sopivia tutkimusaineistoja. Raportti johdattelee rekisteritutkimukseen liittyvään lainsäädäntöön ja kuvaa terveydenhuoltotutkimuksen kannalta tärkeitä rekisteriviranomaisia ja rekistereitä. Lisäksi käydään läpi tutkimuslupien hakuprosessi ja aineistojen keruussa käytetyt menetelmät, annetaan esimerkkejä varsinaisten tutkimusaineistojen muodostamisesta sekä aineistoissa todetuista yllättävistä puutteista. Lopuksi tehdään vielä saatujen kokemusten pohjalta johtopäätöksiä rekisteriaineistojen käytöstä terveydenhuoltotutkimuksessa.

1 Introduction

Administrative registers are the primary source of population data on health service use, sources of variation in access to care and the outcomes of treatment. The combination of increasing evidence of the contribution of health services to improved health and concern about inequities in treatment among population groups have increased the role of administrative registers as data sources in health services research. Despite increased recognition of the importance of register-based research and quality standards for their use for administrative purposes, there are few textbooks on register-based research or practical handbooks that are suitable for health service researchers. Many aspects of register-based research differ from other methodological approaches; therefore specific expertise and know-how are required for planning data compilation in register research. It is essential to be well acquainted with registers used in research, their contents, data quality, definitions, data gathering process etc., but even when this is the case, unexpected problems and questions may arise. Often the problems that occur are specific and the course of action needs to be decided case by case. Nevertheless many of the steps in register-based research are generic and it may be useful to examine experiences of one research project as an example of how register-based research can be done.

Finland has exceptional resources for conducting register-based health care research, mainly for three reasons. First, the Finnish welfare and taxation systems have produced a large amount of reliable administrative data that dates back several decades. Second, a unified personal identification code system was adopted in Finland on April 1st 1964 and by 1968 it covered all citizens, while from 1969 it has been used in most administrative data registers. This has enabled researchers to link data from several databases easily and reliably. Third, Finnish legislation on the use of personal data for purposes other than the original purpose makes it permissible for register-based research to be conducted. In Finland personal data can be used for research purposes, but only under strict provisions defined by law. Research must aim at the common good and no harm can be inflicted on individuals during the process.

This report constitutes a co-operation between the research project entitled 'Regional disparities, social segregation and socioeconomic patterning: Where do inequities in access to health care arise?' (REDD) and the Finnish Information Centre for Register Research (ReTki). The REDD project is co-ordinated and mainly carried out at the National Research and Development Centre for Welfare and Health (STAKES). It is partly funded by the Academy of Finland through its health services research program (TERTTU, www.uta.fi/laitokset/tsph/terttu/). The REDD research team (Appendix 1) has a long experience in register-based health and welfare research and has extensively used the same registers previous to the REDD project. Altogether the project utilises data from twelve different registers and databases, such as registers on hospital admissions, causes of death, health insurance reimbursements, pensions, and Populations Censuses. The objective of the REDD project is to evaluate the association between area variation, urbanisation, socioeconomic patterning, the development of social segregation in Finland and their contribution to the changing regional and socioeconomic gradients in health and access to health care. REDD comprises altogether eight sub-studies and datasets examining different aspects of public health care and treatment results. It is a good example of how register data can enable us to address research questions that could not be studied otherwise or that would be very expensive to study.

The Finnish Information Centre for Register Research, ReTki (www.rekisteritutkimus.fi) is an independent collaboration unit that was established in 2003 and initiated by the Society for Social Medicine in Finland and the Finnish Epidemiological Society. The Academy of Finland supported the launch of ReTki by financing its working up to the end of 2005. At the beginning of

2006 ReTki received a more permanent status and its operation is now financed by the National Public Health Institute, the Institute of Occupational Health, the Social Insurance Institution and STAKES. The objectives of ReTki are to maintain a co-operation network between the register authorities on the one hand and research institutes and researchers conducting register research on the other hand, and to give training, consultation and advice for researchers planning or carrying out research projects utilising administrative registers.

This report documents the planning and compilation of the REDD study data and gives brief examples on some potential uses of the data. The objective is to share experiences of the challenges and benefits of register-based health care research with researchers planning to use similar source data in their studies, while also describing the progress of the REDD project and explaining issues arising due to the special features of register data. The aim is to describe how to plan a register-based health services research project, problems to be expected and how to access data. Questions about data quality and validity, and issues related to protection of sensitive personal data are also addressed. Besides documenting the formation of the REDD data, the report is meant to be used as material in teaching the methodology of register-based research.

In the following chapter the background of the REDD project is described. The objectives and aims of the project and the rationale for using register data to attain these goals are briefly outlined. Further, how the special nature of the data affected the planning stage of the project and its effects on the differentiation of the research questions are described. The third chapter depicts the authorities and data sources used in data compilation and describes in brief the history and the original purposes of these databases. The legislation regarding the use of administrative data in research is discussed in chapter four as well as the application process that was required to obtain permissions for using the REDD data. In the planning stage, the decisions on target populations and relevant available information are made, but defining the final study settings, data structure and variables for analyses is a continuous process that is made clear as the quality of the data and its exact content is revealed. This process and the problems encountered and the data compilation process are described in chapter five. The process of forming final research data sets and examples of unexpected defects in the data are portrayed in chapter six. And finally some conclusions from the process are presented.

2 Background and objectives of the REDD project

The REDD study project is a good example of how research data based on linked administrative registries allows for the investigation of topics that would otherwise not be feasible to study or the research would be very expensive to carry out. While the REDD project aims to explore the contribution of various area-based socioeconomic and spatial factors—such as the economic situation of municipalities, organisation of health care and social segregation, and the regional and socioeconomic inequities—in health and access to health care detailed information is needed both on variables concerning geographical units (hospital districts, municipalities) and on variables concerning individual persons. In Finland, the unique personal identification number assigned to all inhabitants and used in most administrative registers enables the compilation of complex secondary databases needed for this kind of research. In this chapter the background and study questions of the REDD project are briefly described. The purpose is to give context to the description that follows of the planning and compilation of the REDD research data.

2.1 Spatial and socioeconomic differences in access to health care in Finland

Until the 1980s, research on socioeconomic disparities in access to health care in Finland has mainly reflected urban–rural differences in service availability and affluence, with affluence concentrated in urban areas (Kalimo et al. 1992, Purola et al. 1967, 1971). Particularly since the recession and health service reforms of the early 1990s, a more "western European" pattern of social disadvantage has emerged. This is manifest as urban poverty, affluence in rural areas surrounding major cities, and continuing disadvantage in deindustrialised, depopulated and remote areas (Riihinen 1965, Siirilä 1990, Malinen et al. 1994, Kortteinen and Vaattovaara 2000, Kainulainen et al. 2001, Vaattovaara and Kortteinen 2003). Different area types may produce distinctive social and spatial patterning of health service use that does not necessarily reflect health need (Gardner and Chapple 1999, Tod 2001, 2002, Kennedy et al. 2002).

The Finnish health services aim to provide equal access to health care across the country. However, the system has retained a pluralist referral system. Access to specialist treatment may follow referral from a public health centre, workplace primary care or by accessing a private specialist directly. The availability of these different referral routes is likely to vary. Relevant factors include distance, organisation of health services, employment (availability of workplace services), income (ability to absorb higher costs of private care), education (knowledge and ability to navigate the health system) and level of health (presence or absence of co-morbidity). While the mechanisms by which these factors exert their effects are incompletely understood, previous studies have identified points in the treatment pathway that are worthy of more detailed study.

While primary care in Finland has been widely admired, decentralisation has led to concerns about geographical and social inequities (Koivusalo 1999). Current problems include waiting times for appointments, variations in list sizes and also staff shortages (Kokko 2003). Studies from other countries suggest that these features may result in avoidable admissions for ambulatory care sensitive conditions (Basu et al. 2002, Oster and Bindman 2003). In Finland, before the establishment of asthma treatment guidelines and targeted education, there were regional variations in treatment, avoidable admissions and lengths of stay (Haahtela and Klaukka 1998, Csonka et al. 2000, Haahtela et al. 2001). While treatment quality has improved and avoidable admissions diminished, the socioeconomic and geographical distribution of these improvements is unknown. Improvements in the care of patients with coronary heart disease have diminished the socioeconomic differences in treatment only slightly (Hetemaa et al. 2003). Socio-economic and spatial variation in admission rates for other ambulatory care sensitive conditions may exist but this has not been studied previously.

Social and spatial disparities in mental health care are also important. As in other countries, the risk of depression increases with lower socioeconomic status. Depression is also an important source of co-morbidity and poorer outcome following treatment for other conditions (Isometsä et al. 2000). The one-year prevalence of major depression in adults is approximately 9% in Finland (Lindeman et al. 2000). Mental health problems are also the commonest reason for receipt of a disability pension, with depression the most common diagnosis (Salminen et al. 1997). Despite the rapid increase in the use of antidepressants during the 1990s, social and spatial variations in the diagnosis and treatment of depression remain (Puumalainen and Helin-Salmivaara 2003).

2.2 Objectives of the REDD project

The REDD research project aims to study socioeconomic and spatial differences in health care use, and health service organisation. Changes in the relationships between these are analysed from the early 1990s to 2003. This period covers the Finnish recession, entry to the European Union and also more recent economic stabilisation. The objectives of the project focus on four broad topics: 1) the impact of the organisation of health services on regional and socioeconomic inequities in access to specialised hospital care, 2) ambulatory care sensitive conditions as an indicator of socioeconomic and spatial gradients in primary care, 3) socioeconomic and spatial variation in the use of mental health services, and 4) development of statistical models to analyse the contribution to inequities in health service utilisation of individual, socioeconomic, area and health service factors. In the following, these research topics in the REDD project and the major corresponding specific research questions are described.

Impact of the organisation of health services on regional and socioeconomic inequities in access to specialised hospital care

There is growing evidence of socioeconomic differences in access to specialist health care in Finland. Earlier studies have shown persistent and increasing disparities in elective surgery in the 1980s and early 1990s (Keskimäki et al. 1996, Keskimäki 2003) and invasive treatment for coronary heart disease (Keskimäki et al. 1997, Salomaa et al. 2001, Hetemaa et al. 2003, 2004). Several studies illustrate regional differences in hospital care and costs in Finland. Some argue that such variations result from differences in clinical practice and administrative decisions at hospital level, and are not an indication of systematic disparities in resources or supply of health services. However, after the 1993 central government subsidy reform, the rate and nature of specialist care development varied across the country.

Poor access to outpatient specialist care in the public sector has remained a problem, although supply increased in the 1990s. Approximately one third of specialist consultations occur in the private sector. These services are concentrated in cities, close to major hospitals. Patients also have to pay over 50% of the cost of private services. Use of private services, therefore, is concentrated in urban and more affluent populations. Socio-economic inequalities in hospital care, therefore, may reflect, at least in part, the impact of the private sector on public hospital referrals (Keskimäki 2003). Differences in access to treatment and care may also reflect individual and area effects of education and material disadvantage. While the relationship between education and income varies by occupational position and industry (Naderi and Mace 2003) in many countries, the availability of individual register-based data on income and education in Finland means that these factors can be modelled separately. It is possible, therefore, to explore the relative importance of education and material disadvantage and thus assist policymakers and practitioners in prioritising strategies for improving equity.

In order to evaluate the contribution of the organisation of health services on regional and socioeconomic inequities in access to specialised hospital care, trends in the socioeconomic and spatial patterning of hospital care in Finland from the early 1990s to the early 2000s are explored in the project. More specifically, access to elective interventions is examined in the sub-study. The main focus is on the following research questions:

- How is the overall socioeconomic and geographical patterning of hospital use modified by spatial and socioeconomic differentiation across the country and in the capital region? Specifically, what impact do spatial differentiation and deprivation have on socioeconomic differences in hospital care?
- Have aggravated levels of individual deprivation (long-term or repeated unemployment and income inequalities) resulted in an underclass with poor access to health services?
- To what extent do organisational factors in health services, such as the specialisation of local services, and public provision of specialist ambulatory care contribute to the changing patterns of socioeconomic inequalities in access to and utilisation of hospital care?
- What is the impact of private service provision and hospital referrals from private practitioners on socioeconomic and social inequities in access to hospital care and on changing patterns of hospital use?
- What impact does co-morbidity have on socioeconomic and spatial differences in health service use?

Ambulatory care sensitive conditions and amenable deaths as indicators of socioeconomic and spatial gradients in health service

The quality of individual and area based register data in Finland means that trends in the proportion of admissions for ambulatory care sensitive conditions and in causes of death that are amenable to health service intervention and variations between areas and socioeconomic groups can provide reliable indicators of the health care system. The conditions that fall into these categories are distinguished by clear, agreed and objective diagnostic criteria and broad agreement on what constitutes best practice. The proportion of admissions for ambulatory care sensitive conditions, and the variation between population groups, provides a measure of the effectiveness of primary care, by measuring where prevention and chronic disease management in the community has been unsuccessful. Similarly, detailed analysis of the nature, rate and variation in the levels of amenable causes of death provides reliable data on imbalances in the level and appropriateness of health service provision compared with the ability to benefit from interventions in the population.

The economic recession of the early 1990s affected primary care expenditure more severely than specialist health care. Over four years, spending on primary health care reduced by 13% compared with 5% for hospital care (Linna and Häkkinen 1996, Social Insurance Institution 1996). Since then, the allocation of new resources has favoured specialist services. Hospitals, for example, established 1300 new medical posts in 1999–2000. As the overall pool of doctors did not increase at the same pace (the increase was 300–400 per year), this reduced the availability of doctors in health centres. The slower pace of professional development in primary care and limited incentives for staff in rural areas have exacerbated this situation with rural, northern and eastern Finland suffering disproportionately. By comparison, opportunities for non-specialist doctors in private and workplace care in the capital region, combined with lower public sector pay, have contributed to larger nominal patient lists than in similar areas in Norway (Norway average 1300, Jepson, 2001) and the UK (London average 1985, Office for National Statistics 2003). Waiting lists for non-urgent appointments, therefore, may reach four to five weeks (McCallum, unpublished thesis data).

Studying socioeconomic and regional differences in the use of primary care in Finland is difficult because of limited primary health care data. This sub-study addresses regional and socioeconomic equity in primary care by exploring trends in socioeconomic and spatial patterning of hospital admissions due to ambulatory care sensitive conditions. This is a reliable indicator of the quality of anticipatory care and chronic disease management (Bindman et al. 1995, Basu et al. 2002). It is plausible that poor access to services, or differential improvements in quality, would first become apparent in increasing rates of indicator conditions among disadvantaged populations, producing steeper socioeconomic gradients in areas with deteriorating services. Earlier studies have identified widening socioeconomic disparities in diabetes-related mortality in Finland from the 1980s to the 1990s (Forssas et al. 2003). This sub-study will focus on socioeconomic and geographical differences in complications and co-morbidity in people with diabetes, specifically:

- What is the relative influence of age, gender, socioeconomic status, area and health service factors (municipality, hospital, hospital district) on variation in admission rates for ambulatory care sensitive conditions?
- How has socioeconomic and regional patterning in the rates of ambulatory care sensitive conditions, including diabetic complications, changed in the 1990s?

The sub-study on socioeconomic and spatial variation in amenable deaths updates analyses undertaken in the 1980s (Poikolainen and Eskola 1986, 1995) and reflects renewed interest in this area of research. As data quality increases and evidence of the relative effectiveness of different methods of organising, delivering and integrating care as well as the level of benefit expected from specific treatments accumulates, policymakers and researchers are becoming increasingly interested in the relationship between socioeconomic and area variation in amenable deaths, and in intra and inter country comparisons of variations in the organisation and delivery of health care. The quality of data in the Causes of death statistics and the ability to link cause-of-death data with hospital discharge registers is possible in many countries where other forms of register-based research are not possible, either because other forms of data are unreliable or because privacy laws limit the possibility of linking data from more than one register. At international level, therefore, policymakers and researchers are interested in how to undertake this kind of research, as well as comparing the findings with their own. Researchers analysing data from Finnish registers have a responsibility, therefore, to demonstrate the robustness of their data and methods so that they can provide an example to those in other countries who have fewer opportunities to gain experience in this type of research. The sub-study on socioeconomic and spatial variation on amenable deaths will focus specifically on research questions, such as the following:

- What are the changes in socioeconomic and regional patterning in amenable mortality in the 1990s?
- What is the relative influence of individual sociodemographic, area and health service factors (municipality, health centre area, hospital district) on variation in amenable mortality?

Socioeconomic and spatial variation in the use of mental health services

Mental health care in Finland experienced marked structural changes in the 1990s. As in other countries, policy shifts emphasised ambulatory care. The recession accelerated the reduction in beds and related resources but most hospital districts, which took over management of mental health services in the early 1990s, did not replace inpatient care with comprehensive community services. Consequently, resources and facilities for patients with mental health problems decreased (Taipale 1996). During the 1990s, municipal mental health clinics in several regions were reorganised, becoming units within health centres. The extent to which the reorganised services retained their specialist care focus, or became primary health care centres for patients with mental health problems, has varied over time and between regions. The frequency with which symptoms of mental health problems are reported in the population, the observed associations between symptoms of mental health problems, particularly depression, and variations in access to somatic health care and in the outcomes of care highlight the requirement for systematic analysis of service use and uptake of treatment. This can only be undertaken by exploitation of comprehensive health service register data supported by socioeconomic and spatial data of similar quality.

Care for patients with depression, particularly the use of anti-depressant medication, also changed markedly during the 1990s. In terms of defined daily doses (DDD) per population, the use of antidepressant drugs quadrupled between 1991 and 2001 (Puumalainen and Helin-Salmivaara 2003). This may partly reflect improved diagnosis of depressive disorders (Sorvaniemi et al. 1998). However, this increase also occurred in other countries, and largely concerned newer, more expensive drugs, for example, Selective Serotonin Reductase Inhibitors (SSRIs). The sub-study aims to analyse the impact of the changes in psychiatric care in Finland in the 1990s on socioeconomic and geographical equity by exploring the following research questions:

- What impact did the rapid decrease in the availability of inpatient care in the 1990s have on socioeconomic and regional differences in hospital care for mental health problems?
- How did socioeconomic and geographical factors affect the diffusion of the use of antidepressant drugs?
- Does socioeconomic and spatial patterning of the use of inpatient and outpatient mental health services in the capital area reflect known patterns of need or illustrate likely inequities in access to care?
- How does depression influence socioeconomic and spatial patterning of access to somatic health care?

Modelling the contribution of individual, socioeconomic, area and health service factors to inequities in health service utilisation

To understand how spatial and socioeconomic differences in health service use arise, it is essential to be able to separate factors acting at individual and ecological level. These include, for example, factors associated with the population being treated, those related to the context in which individuals live and access services, and features of the services themselves. During recent decades, statistical methods to analyse these hierarchical study settings, typical of health services research, have been developed. These methods, developed in other research fields, are used increasingly in the analysis of socioeconomic and spatial influences on health and health care (Rice and Leyland 1996, Keskimäki et al. 2001, Leyland and Goldstein 2001, Leyland and Groenewegen 2003). Such multilevel analyses allow the contribution of variables measured at the individual, area, hospital and hospital district level and the interactions between them to be examined. They also enable the relative contribution of factors that are compositional (attributes of the population of individuals that make up an area) and contextual (features of the area itself) to be assessed using cross-sectional or longitudinal data. Most countries, however, lack the individual data on health care required to exploit the full potential of these methods. The comprehensive administrative registries in Finland enable individual linkages and patient follow-up. It is feasible, therefore, to develop and apply robust methods to analyse changes in socioeconomic and spatial patterning of health and health care use. The methodological objective in the REDD project is to develop multilevel models to analyse the following questions:

- What is the relative contribution of individual and area level disadvantage on spatial and socioeconomic differences in access to treatment and care?
- How do spatial factors and the organisation of health services interact to produce socioeconomic differences in treatment rates?
- Do spatial factors (municipal and environmental) influence the impact of individual education or material disadvantage on access to treatment and care?
- How do socioeconomic and spatial factors affect access to treatment and care?
- How have the relationships between socioeconomic and spatial factors and access to treatment and care changed since 1990?

2.3 Operationalisation of the study objectives in defining study data

The REDD research team considered it important to study socioeconomic and spatial patterning of health care use across different aspects of the health care system and among different population groups. Taken together, for example analysis of socioeconomic and spatial variation in mental health services, in somatic specialist hospital care and in primary care can provide a more comprehensive picture of the influence of these factors on access to health care and on treatment received. Selection of populations for the sub-studies was based on these principles.

Elective surgery was selected as a case study for specialised hospital treatment, since it exhibits a strong element of discretion on the part of health service providers as to how and when treatment is offered. Seven types of procedures were selected for analyses: coronary revascularisation (coronary angioplasty and bypass operation), hip and knee replacement, lumbar disc operation, cataract extraction, prostatectomy, and hysterectomy.

The study group also decided to examine socioeconomic and regional patterning of two common serious diseases, namely diabetes and coronary heart disease. Members of the study group have examined variations in receipt of treatment and in the outcomes of care in these patient groups previously and REDD data allows a longer time period and trends to be analysed and enables further explanation of the differences reported earlier.

In many countries the way that primary care data are collected and stored makes it difficult to create and sustain registers of patient populations that are large enough for meaningful analysis. This is also the case in Finland. Internationally, however, the utility and validity of studying hospital admissions for conditions determined to be sensitive to receipt of high quality primary care has been widely demonstrated. This reflects broadly consistent findings that the balance between urgent, demand-led care, and chronic disease management reflects the extent to which primary care services provide ongoing care for all population groups. To study socioeconomic and spatial variation in primary care in Finland, therefore, a sub-study population comprising hospital admissions due to ambulatory care sensitive conditions was defined. Explanations for variations in ambulatory care sensitive admissions for which other authors have produced supporting evidence include poor access to services, or differential improvements in quality between health care providers and social groups. It is plausible that such variation would first become apparent in increasing rates of indicator conditions among disadvantaged groups, producing steeper socioeconomic gradients in areas with deteriorating services. In line with the categories employed in other countries, ambulatory care sensitive conditions were defined as asthma, COPD, hypertension, angina pectoris and heart failure.

Another perspective for evaluating the performance of the health care system was also selected as a case study, namely avoidable mortality, or 'mortality amenable by health care' (e.g. Nolte and McKee 2004). Although avoidable mortality should only be seen as an indicator of potential weaknesses of the health care system, Nolte and McKee suggest that avoidable mortality could be used to gain new insights into inequalities in access to care. The study group defined avoidable mortality according to Nolte and McKee's list of 34 conditions, but added asthma and COPD to the list as other authors have highlighted the relative lack of attention paid to the contribution of respiratory disease to the burden of amenable premature death. Further, it was not possible to include the category of deaths due to misadventures to patients during surgical and medical care that appears in Nolte's and McKee's list in the REDD project data set, since these causes cannot be determined from the Finnish Causes of death statistics. The proportion of deaths associated with failures of the quality of medical or surgical treatment, and the underlying causes, are broadly comparable in developed western countries, although the relative importance of different factors varies. Inclusion of such data in the REDD project would have required access to another series of register data and the application of different methods of data extraction and analysis. For these reasons, the risk of losing the focus of this sub-study and diluting the availability of the expertise required to develop the datasets necessary for analysis was considered to outweigh the potential benefit of including the additional data.

As stated in the objectives, two types of mental health care were selected for the analyses, namely psychiatric hospital inpatient care and use of antidepressants. Additionally, residents of Helsinki City were selected to study mental health care, since in this setting data can be obtained for use of private care, specialist ambulatory care and hospital inpatient care.

The population of Helsinki City was defined as an interesting study population for two reasons. First, traditional socioeconomic disparities in access to health care in Finland have reflected urban-rural differences in service availability and affluence, with affluence concentrated in urban areas. However, earlier research by the study group and others have suggested that particularly since the recession and health service reforms of the early 1990s a more "western European" pattern of social disadvantage has emerged, manifest in Helsinki as urban poverty (Kortteinen and Vaattovaara 2000, Vaattovaara and Kortteinen 2003). Second, Helsinki is one of the few municipalities in Finland that maintains a comprehensive computerised register of ambulatory health centre use. This allows the study of public and private hospital and ambulatory care. Where patients attend workplace primary care (occupational health care), however, ambulatory care data are missing, although, as with other areas, it is possible to examine variation in referral to hospital and in ambulatory care sensitive admissions.

3 Original source registers for the REDD study data

The data for the REDD project is gathered from ten different registers or databases. These are maintained by five different register authorities, namely STAKES, Statistics Finland, the Social Insurance Institution, the Finnish Centre for Pensions, and the City of Helsinki. Furthermore collaboration with two ongoing research projects carried out at STAKES provided the possibility to exploit already existing research datasets. In this chapter the original source registers and databases used in compiling the REDD study data are described. The descriptions of the data sources are organised according to the agencies maintaining the databases.

3.1 National Research and Development Centre for Welfare and Health (STAKES)

The Finnish Care Register HILMO

The Finnish Care Register HILMO (http://info.stakes.fi/hilmo/) is probably the most important register used in health care research in Finland. It is maintained by STAKES Information Division for statistical purposes, but the national legislation allows its utilisation for research purposes. HILMO has been fully operational since 1967 (and to a lesser extent from 1960), and contains information on all inpatient hospital admissions from 1969 onwards. The register was originally named the Hospital Discharge Register since the entry to the register was made as the patient was discharged from the ward. In 1994 the register was replaced with HILMO and at the same time it was diversified to also cover day surgery and information on long-term patients collected through patient counts at the end of each year. Data collection on social care started in 1996 and HILMO was divided into two parts; hospital care and social care registers. At present, the HILMO register annually gathers data on about 1.2 million hospital discharges and 1.3 million care periods in social care institutions (STAKES 2006). The REDD project uses information only on hospital care.

Since 1969 the registration practices and variables have gone through many changes and therefore some modifications are required if extended time series are used in research. Nevertheless, the main content has remained the same: data that identifies the care provider and the patient, information about the admission and discharge as well as on diagnosis and received treatment. Nowadays HILMO contains some 100 variables.

Care data are reported to STAKES by the service provider. A specific handbook is issued yearly to guide informants on how the data is reported and its required contents (STAKES 2004).

The reliability of the data is the responsibility of the register authority. Most register authorities do regular checks and calibrations on their registers. HILMO is followed and checked by cross-tabulating specific diagnoses against age and gender. Additionally dates of admissions, discharges and procedures should appear in sensible and chronological order. In case of abnormalities or inconsistencies in the data, the hospital in question is asked to revise the information. Free software for checking the data before sending it to STAKES is available for the informants. The most recent version is from 2001, HILTA-01 (http://info.stakes.fi/nettihilmo/hilta). Comprehensive evaluations on the coverage or accuracy of HILMO have not been published but according to validation studies in the 1980s the hospital discharge register contained about 95% of all discharges in the hospitals (Salmela and Koistinen 1987) and most central information (diagnoses, surgical procedures) were recorded correctly in at least 95% of the discharges compared to the corresponding medical records (Keskimäki and Aro 1991). Several studies have assessed the validity of HILMO in terms of reporting hospital treatment due to coronary heart disease (Mähönen et al. 1997, Pajunen et al. 2005, Rapola et al. 1997). According to these studies, register diagnoses were reasonably valid when compared with diagnoses made with standard criteria from the FINMONICA/FINAMI study. Also diagnoses of some psychiatric disorders, such as schizophrenia (Mäkikyrö et al. 1998) and bipolar disorder (Kieseppä et al. 2000) have been found to be reasonably valid in HILMO.

Data from STAKES research projects

Hospital Benchmarking System

A project to monitor the productivity and efficiency of the hospitals in Finland was launched in 1997 by the Health Services Research Unit at STAKES. The project aims to develop tools to monitor and follow the development in cost-effectiveness in health care and enable comparisons between hospitals and regions. The hospital benchmarking system combines cost data with output data, thus allowing productivity comparisons between hospitals. All patient contacts within participating hospitals since 1998 are recorded in the benchmarking database. In 2005 the database covered 95% of all hospital care in Finland. The database is updated annually. The hospitals send their records of hospital care electronically to STAKES, where measurements and indicators are produced from the data and recorded in the database.

Standard reports are produced from the database regularly. All hospitals have access to the information via the internet. Currently, the reports are not public and a password is required to enter the database, but the future plan is to incorporate the Hospital Benchmarking System reports into the official health care statistics that are subsequently released publicly. The benchmarking data is collected by a research team at CHESS at STAKES. The management team includes members from all hospital districts. The Hospital Benchmarking System, its methodological background, and data collecting system are described in more detail at the website of the project (www.stakes.fi/benchm/english/index.html). REDD uses only data on ambulatory visits from the benchmarking database.

Diabetes in Finland

Diabetes in Finland (Niemi and Winell 2005) is a research and development collaboration between STAKES, the Diabetes Association and the Social Insurance Institution. The project focuses on the prevention of diabetes and co-morbidities among diabetic patients and it aims to give a real time description of the diabetic population and the prevalence of diabetes related co-morbidities and to evaluate the reliability of register data in these tasks. The project compiles data from several registers containing information on the diagnosing, treatment, co-morbidities, complications and outcomes of the diabetic patient. These registers include among others the Causes of death statistics from Statistics Finland, registers on entitlements to special reimbursement and reimbursed prescription costs from the Social Insurance Institution, HILMO and the Medical Birth Register from STAKES, and the Finnish Register for Visual Impairment from the Finnish Federation of the Visually Impaired.

3.2 Statistics Finland

For the study data, the REDD project exploits several statistical registers maintained by Statistics Finland in order to obtain data on sociodemographic factors, mortality and the population at risk. These registers are the 2000 Populations Census and Longitudinal Database of the 1970–1995 Populations Censuses, Employment statistics in 1991–2002, and the Causes of death statistics.

Populations Census in 2000 and Longitudinal Database of Population Censuses

The first Populations Census in Finland was executed in 1950, as the 1940 Census was cancelled because of the war. The 1950 and 1960 Censuses were carried out entirely through questionnaires. Data collection was time consuming and expensive, so when the personal identification system and first personal registries were created in Finland in the early 1960s it became sensible to use registers to collect Census data. Administrative data were for the first time used in the 1970 Census and their role continuously increased after that. Since 1990, Census data have been collected exclusively from administrative registers. A parallel survey was collected in 1990 to test the reliability of the register data which proved to be high while the expenses were only one-tenth compared to the questionnaire surveys.

Populations Censuses (Statistics Finland 2004) have been conducted every fifth year by Statistics Finland since 1970. The latest one available at the moment was collected in 2000 and the information was publicly released December 2002. Census data contain individual level information on several socioeconomic variables, such as profession, domicile, employment, and income. Information on family structures, living conditions, housing stock and recreational dwelling is produced as well.

Additionally, Census data is compiled in a longitudinal database, the latest up-to-date database includes data from the 1970–2000 Censuses. The longitudinal database was created to make easy comparisons in time possible by harmonising the changing classifications and decision rules. For instance, when studying occupational status, the longitudinal data set proves to be useful. The basic Census data only includes social class information of the economically active working population. For persons outside the labour force the longitudinal Census data looks back to each previous Census up to the 1970 Census in order to find information about occupation. At the time when the REDD data was compiled the longitudinal data covered the years 1970–1995.

Employment statistics

Statistics Finland has produced annual employment statistics since 1987 (Statistics Finland 2004). The system was created to monitor the population's economic activity. Data are collected from some 30 administrative register or statistical sources. Before adoption of the statistics, the compilation of the database was tested by comparing register data to information collected through questionnaires. The register data again proved to be reliable.

The main sources for employment statistics data include data from the Population Information System, the Ministry of Labour, the tax administration and several registers of Statistics Finland, such as registers on educational degree and workplace. Employment statistics describes the year-end situation for every resident in relation to income, education, professional status and unemployment. Information describes the individual's situation at the turn of the year but also data on annual income and employment and unemployment time in the preceding year are presented. These kinds of data are considered confidential at the individual level, but at the area level they produce valuable regional information on economy and employment.

For health services research, it is often necessary to use data from Employment statistics at an individual level and to link these data to individual level information from other data sources, such as hospital discharge records. When Statistics Finland releases confidential data for outside use, the linked data is first made anonymous and then processed in a way that also makes indirect identification impossible.

As with the Populations Censuses, the Employment statistics are compiled in a longitudinal database. The latest up-to-date version of the longitudinal employment statistics database covers the years 1987–2002.

Causes of death statistics

The Causes of death statistics (www.stat.fi/meta/til/ksyyt_en.html) are used to monitor mortality in Finland and its development over time. Causes of death statistics have been produced since 1936, when the death certificates were issued by medical doctors systematically across the whole country. The register is complemented and compared with the information in the Population Information System that is maintained by the Population Register Centre.

The Causes of death statistics are updated every year and include every Finnish resident that has died during the preceding calendar year in Finland or abroad. The Causes of death statistics cover also deaths of non-residents. The classification of causes of death has changed several times during the years. For the first fifteen years the register followed national coding that was validated in 1936. In 1951 the International Classification of Diseases (ICD) coding by the World Health Organization was adopted. At that time the sixth version of the ICD-classifications was in use. Since then coding has been changed and updated four times, while the latest version ICD-10 has been in use since 1996.

A time series on the Causes of death statistics has been available electronically since 1969 and was used to compile the REDD data. There are two variables in the register that describe the underlying cause of death; the selected underlying cause and the statistical underlying cause of death. The selected cause of death agrees in most cases with the one defined by a medical doctor. International rules and guidelines are adapted to obtain international comparability. This modified cause of death is used in published statistics and is therefore called the statistical cause of death. Both definitions for causes of death may need to be considered when causes of death are investigated in health services research. Statistics on causes of death published by Statistics Finland use the statistical underlying cause of death only. Additionally the statistics include information about all the conditions mentioned in the death certificate as well as the external causes on the circumstances about death and sociodemographic information for the deceased.

The Finnish Causes of death statistics—based on the underlying cause of death as stated on the death certificate—has been found to be valid and reliable in general, although some causes of death, such as alcohol poisonings have been found to be underestimated in the statistics (Lahti 2005). While the autopsy rate in Finland is high in international comparison: 30% for all deaths and about 60% among those of working age, the number of deaths with undetermined or unspecified causes of deaths remain small. Out of nearly 50 000 deceased in 2003 only about a hundred cases remained without a cause of death at the time of the publication of the statistics. These are partly deaths that occurred abroad where only a notice on death is received or the death certificate had not been given by the deadline for the statistics.

3.3 Social Insurance Institution

A third important contributor to register-based health research in Finland is the Social Insurance Institution. When a reimbursement for medical expenses, pension or other social benefit is applied for from the Social Insurance Institution, all information that is required for the handling process, is recorded in the data management system of the Institution. The Social Insurance Institution maintains several registers and databases for different health insurance and social security benefits. Three of these databases are exploited in the REDD project: the Register on entitlements for special reimbursement for medicine costs, the Register on prescription reimbursements and the Register on pensions.

The Social Insurance Institution registers were originally not intended for statistical or research purposes but primarily for an administrative function. However, these registers are reasonably accurate and reliable so they form a valuable source of information for health services research. The information is gathered from forms and applications completed by customers or service providers. Logically the reliability of the data depends heavily on the motivation of the informant. When financial benefits determined by the law are sought and granted after a close evaluation process, only a few inaccuracies can be expected to be recorded in the final databases. However, the fact that the data itself are accurate and reliable does not indicate the total coverage of the disease or condition. For example, of the 200 000 diabetic people in Finland only two thirds require medical treatment, so the remaining third have no record in the registers of the Social Insurance Institution. To receive a social benefit usually requires action on the recipient's part, and if the person does not seek it for some reason, she does not appear in the registers. Additionally diagnostics may vary regionally or depending on medical practices.

In the following, the short descriptions on the Social Insurance Institution registers used in the REDD project are given with some notes on issues relevant to research purposes. More detailed description of all registers and databases maintained by the Social Insurance Institution and their contents are described in the publication issued by the Institution, which can be obtained at the Institution's website (Kelan tilastoryhmä 2005; for the content of the registers see also www.kela.fi/in/internet/suomi.nsf/Docs/240901083700TL?OpenDocument).

Register on entitlements for special reimbursement

Of all Social Insurance Institution registers, the Register on entitlements for special reimbursement is probably most often used in health services research. A right for special reimbursement is granted by the Social Insurance Institution to a person suffering from some of 44 specific chronic diseases or conditions decided by the Council of State. The persons with a special reimbursement right are entitled to an elevated reimbursement for their medicine expenditure for the prescriptions due to the special reimbursement right conditions. The regulations on how the elevated reimbursement is calculated have changed several times. For instance, between 1.4.1994 and 31.12.2005 the special reimbursement was 100% for most of the 44 conditions or 75% for some instead of the basic 50% of the costs of a one-time purchase above the fixed minimum prescription payment. In the beginning of 2006, the minimum prescription payment was abolished in the two lower reimbursement categories and the percentages were reduced to 72% and 42% of the costs. In the higher special reimbursement category, the minimum prescription payment was changed into 3 €/medicine (Laki sairausvakuutuslain muuttamisesta 885/2005, Valtioneuvoston asetus lääketieteellisin... 2005). A detailed description of the changes in the system is available in the Statistical Yearbooks of the Social Insurance Institution (for changes before 1990 see Social Insurance Institution 2001).

To receive a special reimbursement right the patient has to be examined, diagnosed and given a certificate by a medical doctor, usually a specialist. The application is then reviewed by the Social Insurance Institution as to whether the required criteria for the special reimbursement are met before granting the right for the elevated reimbursement level. Each of the 44 conditions or diseases has uniformly defined criteria for obtaining the right of reimbursement. For example, the diagnostic criteria for the right for special reimbursement of medicine costs for coronary heart disease are (1) chronic angina pectoris symptoms responding to anti-angina medication, and with unequivocal ECG changes (if QS waves are not detected in resting ECG, typical ischaemic changes are required in exercise test), or (2) diagnosed myocardial infarction, or (3) a revascularisation operation (coronary angioplasty or coronary artery bypass grafting), or (4) coronary heart disease verified in angiography. Because of the strict evaluation process the register provides reliable information on common and serious illnesses and is therefore often used to define study populations or co-morbidities in health care research. The criteria of granting the special reimbursement rights for the 44 diseases and conditions are prepared in the Social Insurance Institution in co-operation with clinical experts and decided by the Social Insurance Institution.

The public mandatory health insurance system was adopted in Finland in 1964 and since then records of reimbursements have been entered into the register. The register holds information on starting and possible ending dates for the reimbursement right and codes for the disease or condition that is the basis for the reimbursement right. The list of conditions entitling to special reimbursement has changed during the running time of the system. For instance, coronary heart disease was included in 1986 (Valtioneuvoston päätös vaikeista…1986). While the register is based on administrative decisions, any changes introduced in entitlement criteria or conditions affect the register only gradually. When using data from the Register on Entitlements for Special Reimbursement, the researcher should be aware that the diagnosis coding used in the register does not follow the ICD classification and criteria to grant reimbursement are not always equal to diagnostic criteria used in clinical research. In addition, regional variations exist in medical diagnostic and treatment practices, which have some noticeable impact on the regional numbers of special reimbursement rights (e.g. Klaukka 2001).

Register on prescription reimbursements

Data on filled prescriptions can be used as an indirect indicator for sickness rates, although pharmaceutical prescriptions are not as definite morbidity indicators as special reimbursement rights. However, in some cases the prescriptions may reflect the history and development of the disease or its severity and complications. Information on prescriptions also provides a tool to monitor treatment practices and their changes.

The Register on prescription reimbursements maintained by the Social Insurance Institution contains information on the reimbursements of the outpatient costs of prescribed medicines from 1994 onwards. The register includes entries for pharmaceutical preparation according to the ATC classification, date of purchase, date of prescription, cost, reimbursement, dosage, as well as the patient's age and gender and place of residence. The prescribing doctor is also recorded (the SII code of the prescribing doctor, the doctor's speciality). Inpatient medication, purchases under deduction limit and purchases compensated by the employer are not recorded in the register (Kelan Tilastoryhmä 2005). The register covers a large amount of purchases, for example in 2004 the number of purchases was ca 20.5 million (National Agency for… 2005).

Register on disability pensions and early retirement

The Social Insurance Institution maintains a register on all its decisions on granting pensions and corresponding benefits. Records on disability pensions, rehabilitation grants and individual early retirements include ICD codes about the medical conditions that inflicted on the person's capacity to work and therefore entitled her to the benefit. In research use it is important to notice that the disability pension records of the Social Insurance Institution are not congruent with the registers of the Finnish Centre for Pensions. The main reason for the discrepancy of these registers is that the pension information for those receiving larger employment pensions (more than ca 1000 €/month) are not registered in the SII database and those with only basic pension or basic pension and smaller employment pension are registered in the SII register only. Therefore for the comprehensive information on disability pensions, data should be applied from both agencies.

3.4 Finnish Centre for Pensions

In Finland the administration of the earnings-related pension scheme is decentralized between different authorised pension institutions. The Finnish Centre for Pensions (www.etk.fi) is their central organisation. The Finnish Centre for Pensions gathers data from all pension institutions on self-employment and on every employment contract in Finland. It produces monthly, quarterly and yearly based statistics on pension receivers and pension expenses from the private sector.

A yearly publication on all pension receivers, including public sector and national pensions, is published in co-operation with the Social Insurance Institution.

The disability pension data released by the Finnish Centre for Pensions includes the personal identification number, the type of pension (full-time pension, part-time pension, early retirement, rehabilitation grant), date of beginning and end of the pension, the law the pension is based on, and ICD codes for the main cause of disability. In addition to the register on pension decisions the Finnish Centre for Pensions maintains registers on pension accruals and applications. More information on the registers of the Finnish Centre for Pensions can be obtained at the web site of the Centre (www.etk.fi).

3.5 City of Helsinki

In the REDD project, the administrative registers of the City of Helsinki are used for the sub-study on social segregation, and socioeconomic and small area variations in use of health care in the primary administrative areas of the city in order to define the study population and to gain information on public health care utilisation. The base population of the sub-study was planned to cover persons that have been the residents of the city at the end of each year between 1998 and 2003. The base population is collected from the population register system of Helsinki. The database of the system is updated weekly with the official data on population changes at the Population Register Centre.

The municipal health care registers in the City of Helsinki cover all ambulatory primary and specialist care visits to the doctor in Helsinki municipal health stations, hospitals and psychiatric outpatient services. Information concerning occupational health care is not available, since the city only provides part of the occupational health care. The database includes the date for the visit, several diagnostic and procedure codes, and the attending doctor's speciality. Also recorded are the sector and small area code of the service provider and some demographic background information about the patient. The structure of the database was modified last in 2001 and at the moment the database includes more than 70 variables. Up until 2001, the register included about 40 variables. The database also contains information on the patients' telephone consultations to some extent. Because these data are reportedly coded variably in different health stations, these data are not used in the REDD project.

4 Applying permissions to use register data for research in the REDD project

For a research project like REDD, which is based mainly on register data, it is important to gain straightforward access to the data and have an opportunity to link data coming from different registers. However, a large part of the data essential for studying people's socioeconomic circumstances and health is considered confidential and understandably protected by legislation. It is also a benefit for research to preserve effectively the confidentiality of administrative register data, since otherwise the public would not be willing to disclose their data related to personal sensitive issues. This chapter presents how confidential data can be received for research purposes. We start with an overview on the legislation regulating the data protection and releasing of data for research purposes. Then we go through the practical steps that are needed or advisable to take as part of any register research: first, the ethical committee's review; second, the application for permissions from register authorities; third, the execution of the formal file description. The chapter ends with a description of how this process was conducted in the REDD project.

4.1 Legislation for the protection of privacy and for delivering confidential data

The Finnish Constitution states that the right to privacy is a basic right. Detailed provisions on the protection of personal data are provided by an act. Everyone has the right to know and have control over what information is collected about him or her, regular sources and the purpose of processing the data and the use and regular destination of data. On the other hand the Finnish welfare system's use of the unified personal identification number has led to a large number of data files controlled by the competent authorities that have been found to be very useful for research purposes that benefit society as a whole. As the quantity of data available and the interest in making use of it has grown, it has become necessary to establish common regulations, while at the same time preventing possible misuse of the data and ensuring that no harm is inflicted on individuals at any stage. The Finnish legislation on data protection is in international terms relatively permissive and allows the use of administrative data for scientific, historical, and statistical research purposes. However, the legislation appoints enforcing authorities to ensure that individual rights are not violated when administrative data sources are exploited for research. In Finland the enforcing authority is the Data Protection Ombudsman (www.tietosuoja.fi).

The use of administrative registers for other than administrative purposes is regulated by, among others, three focal acts; the Personal Data Act, the Act on the Openness of the Government Activity and the Statistics Act. In addition there are three special acts concerning STAKES and its registers. All Finnish legislation can be electronically obtained from the Finlex database (www.finlex.fi).

Personal Data Act 1999 (523/1999)

The constitutional reform in the early 1990s and the EU Data Protection Directive in 1995 led to the renewal of the first Personal Data File Act that entered into force in 1988. In June 1999, the old Act was replaced with the new Personal Data Act that emphasized even more rights of the individual and duties of the controller. The Personal Data Act states that in principle the processing of personal data is prohibited without an informed consent from the individual. Processing means inter alia collection, recording, transfer, storage, manipulation, combination, protection and other measures directed at personal data. Furthermore, the Personal Data Act states that the purpose, sources and recipients of the data collection have to be defined before any personal data are collected and recorded. In general the recording and processing of sensitive data, such as race, religion, political affiliation as well as data on health and illness is prohibited. The Act lists, however, exceptions for this injunction. These include the processing of data on the basis of a specific act or using data to scientific research.

Register-based studies can normally be conducted without subjects' informed consents because there are special regulations for the processing of personal data in scientific research in the Personal Data Act (14§). The Act states that personal data may be used in research without the subject's informed consent if the following requirements are met:

1) when research cannot otherwise be carried out without identifying the person, and the consent of the data subjects cannot be collected because of the quantity of the data, age or other justifiable reason;
2) appropriate research plan where persons responsible for the research are named as well as everyone who had to have access to the data;
3) data that pertains to a given individual may not be disclosed to outsiders at any stage, and
4) the data file is archived, destroyed or made unidentifiable after ending the project.

Act on the Openness of Government Activities 1999 (621/1999)

The objective of the Act on the Openness of Government Activities is, inter alia, to promote openness on information management in government and to improve the possibilities of exploiting administrative data systems when it does not conflict with individual rights. According to the Act, every individual has the right to access information pertaining to him or her in an official document as a principal rule. According to the Act, an authority may, on a case-by-case basis, grant permission to gain access to a secret personal document for purposes of scientific research. When the personal data needed in the study is based on the registers of more than one authority, the permission has to be received from all of the authorities in question. However, for data belonging to different authorities who are under the same Ministry, the Ministry shall decide on the permission. In general, the grounds for applying, the purpose of use, quantity and quality of the data must follow the confidentiality and data protection regulations.

Statistics Act 2004 (280/2004)

The Statistics Act regulates the principles and procedures applied to planning and compilation of statistics by the Government agencies. In comparison to other official personal data, the Statistics Act restricts the individual's rights in regard to data pertaining to him or her and states that the requirement to offer access to the individuals own records does not apply to the files kept by statistical authorities. The Statistics Act also places restrictions on administrative authorities in their access to data that have been linked or collected by the statistical authorities. Data collected by the statistical authority may not be used for administrative, control or other corresponding purposes.

According to the Statistics Act, statistical primary data should be collected and filed without identifiers if that is feasible concerning the requirements of the compilation of statistics. For instance, data linkages of statistical registries are considered as an adequate reason to collect personal identification numbers. The Statistics Act allows the statistical authorities to release the confidential information, which they have collected for compiling statistics, on the basis of the consent from the person to whom the data are pertaining or by act. For research purposes, individual data can also be released without individual consent but personal identification numbers or other identifying information cannot be released in this case. The Act states, however, a couple of exceptions to this rule: information on age, gender, education, occupation, and cause of death can be given with the personal identification number if the receiver has the right to receive personal data on the basis of the Personal Data Act.

The laws regulating the action of STAKES

There are special Acts which regulate the collection of register data in the STAKES and also state how the register data can be released. Most important of these are the following: Act on the National Research and Development Centre for Welfare and Health (1073/1992), Act on the Statistical Action of STAKES (409/2001) and Act on the Nation-wide Registers on Health care (A 556/1989 and its change (1088/1992), D 774/1989). In general, these Acts state that STAKES can deliver confidential information with personal identification numbers from health registers for scientific research if the regulations of the Personal Data Act and other Acts are fulfilled. However, information on social benefits such as supplementary benefits and child custody can be released only in anonymous form.

4.2 Ethical review of the study plan

An ethical review of the study plan is not in general requested by the register authorities for permissions to use their data. However, in complex research projects an ethical review may facilitate the decision-making of the register authorities. Much of the ethical evaluation in register-based research deals with the protection of privacy of the individuals. In register research, the individuals can seldom be asked for their consent to use their personal data for research purposes. Therefore it is even more essential to take every precaution in insuring the anonymity of the data and preventing any possibility of indirect identification. Another reason why it is advisable to bring the study plan to the Ethics Committee is because scientific journals demand more and more that these kinds of studies have also been reviewed by an Ethics Committee.

The reviewing of study plans by an ethics committee is statutory in medical studies. These studies are normally reviewed by ethics committees of hospital districts. These committees may also take up other studies involving human studies. Many research institutes and universities also have ethical review boards which review other than medical human studies. Further, all register-based studies that are not medical studies requiring statutory ethical reviewing can be reviewed at the STAKES Ethical Review Board. Although it mainly reviews study plans that will be carried out at STAKES it also carries out ethical reviews for outside research projects involving human subjects.

4.3 Applications for permissions to receive register data

Application procedure in general

The research project intending to use register data that are managed by several authorities has to apply for permission to use register data from each authority concerned separately and to make an agreement on arranging the control of the data file. When the study addresses data from several health care organisations (for instance hospitals and health centres), the permission for data access may be applied from the Ministry of Social Affairs and Health. Each authority has its own procedure to deal with the applications. There are, however, some common features in the requirements of the applications. For instance, permission to use individual level register data can usually be granted only for research purposes. Moreover, the authorities are not allowed to give researchers a general authorisation for using their data for research, but the permissions are given for specific, clearly defined studies. Accordingly, to receive authorisation to use sensitive personal data from the register controllers and authorities, the research team has to provide a sufficiently detailed and appropriate study plan that includes the following items, along with the free format application or the filled application form:

- Names and responsibilities of the persons involved in the project and the project leader. Persons that are not named in the plan should not have access to the data.
- Clarification on how privacy of the study subjects in the personal data file is secured and how the process will be ongoing.
- Description of the data content and what other sources are used in the compilation process. When the data collection is completed a file description should be available to anyone.
- Description of the objectives and scientifically valid methods and guidelines used in the study.
- A plan for archiving or destroying the data after the project is completed.

Most register authorities normally require that all members of the research team that use the data set sign a statement that they will agree not to disclose the sensitive personal data, use these data for other purposes or harm the subjects of the study.

Application procedures to specific authorities

In the following we present, in short, the practical procedures of applying for permission to use register data, separately for each register authority whose data were planned to be used in this study.

a) STAKES and Helsinki health care register

If only registers of STAKES will be used in a study, the permission to use data is applied for from STAKES. There is a standardised application form available at the STAKES website (http://www.stakes.fi/FI/Tilastot/tausta/tutkimus) for this purpose. The form requests information on the following items for the study: the name of the research project; the register keeper, i.e. the organisation or research team that uses the register data and is responsible for it; the responsible leader; the members of the team that uses the database; the purpose of the study; the information content of the research register; the principles of data protection that are in use; and the plan for archiving or destroying the data after the project is completed. Additionally the timeframe of the study needs to be defined.

Similarly if only health care register data of one health care district or community will be used, the permission is applied for from the health authority itself. The application can be written in free format or the model application form available from the web page of the Ministry of Justice (http://www.om.fi/23957.htm) can be used.

Since both STAKES and health care centres work under the Ministry of Social Affairs and Health, it is possible to apply to the Ministry for permissions for admission to both registers' data by means of one application instead of sending an application separately to the respective register authorities. The application form to the Ministry of Social Affairs and Health is the same as that to STAKES, and it is available at the above mentioned STAKES web page. Before the Ministry grants authorisation for the project, the Data Protection Ombudsman (www.tietosuoja.fi), STAKES and the other register authorities are given a say on the matter.

b) Statistics Finland

Statistics Finland utilises a standardised application form (www.stat.fi/meta/tietosuoja/kayttolupa.html). Statistics Finland does not usually grant permissions for data on the whole population of the country or a whole region. Due to restrictions on releasing data with personal identifiers, Statistics Finland has often acted as the authority that carries out the linkages of identified data sets and removes the personal identification numbers, which is useful to know beforehand when applying for permission to use their data.

c) Social Insurance Institution

When permission to use data is applied for from the Social Insurance Institution, no standardised application form is available, but a free format application along with the project plan is sent to the Social Insurance Institution. In the application, the applicant should clearly denote what data are applied for, including information on classifications, timelines, age restrictions, regions and subject types.

d) Finnish Centre for Pensions (FCP)

Data on retirement and disability pensions can be applied for from the Finnish Centre for Pensions. Before granting permission, the Finnish Centre for Pensions consults all the relevant pension institutions, which may prolong the application process. The Centre does not utilise a standardised application form. Instead, a free format application is used.

e) Research projects

If you want to use the data of other research projects you normally need to agree with the study group about using the data. In addition, since most studies are conducted in research institutes or universities, their data is owned by the institute or university. Thus, you have to apply to the institute or university for permission to use the data. This can be done by a free format application or by using the model application form (http://www.om.fi/23957.htm).

The costs of receiving data and making an application for permission

According to the Decree on Criteria for Charges Payable to the State (211/1992) the statistical authorities collect charges on the total cost of producing databases, including both the direct costs and the proportion of administrative and capital costs and other overall costs in producing the performance in question. Negotiations need to be conducted with each of the statistical authorities concerning the costs of the data collection. For example, Statistics Finland collects charges based on the number of databases used in data compilation, the amount of statistician work involved and on storage of the original data with personal identification codes. Similar principles of cost collection concern other statistical authorities involved in the data compilation as well. Other arrangements can sometimes be made when the authorities and the research project have overlapping research interests. If a register authority, which commonly may also be a research institute, has a strong interest in the research questions, they may be willing to become partners in the project and thus cover their own costs in the data collection and compilation. It is worth negotiating these questions during the permission application procedure since many authorities also make the decision on the costs at the same time as they decide about the data delivery.

4.4 The formal file description to be sent to the Data Protection Ombudsman

A description of the data file used in the research project should be available to anyone. The file description is prepared and stored by the register holder before the launch of the project. When sensitive personal information is used in scientific research a notification needs to be made to the Data Protection Ombudsman 30 days before accessing the data. The notification is made by sending a data description file to the Ombudsman's office. There is a standardised form for the scientific research data files available at the website of the Ombudsman's office (www.tietosuoja.fi). The file description defines the register keeper, i.e., the organisation or research team that uses the register database and is responsible for it, the responsible leader of the research project, the members of the team that use the database, the purpose of the study, the information content of the register, the principles of data protection that are in use, and the plan for archiving or destroying the data after the project is completed.

4.5 Ethical review and application for permissions in the REDD project

In the following, it is in shortly presented how the confidentiality was secured and the permissions for using register data were applied from the register authorities in the REDD project.

Ethical review at STAKES

The plan of the REDD project was sent to the STAKES Ethical Review Board on January 28, 2004 and the notice of approval was received on February 4, 2004.

Permissions for using register data from different authorities

The time line for the process of the filing and approval dates of the applications as listed in the REDD project is shown in Figure 1. In the following text we go through the process in detail.

a) STAKES and the city of Helsinki

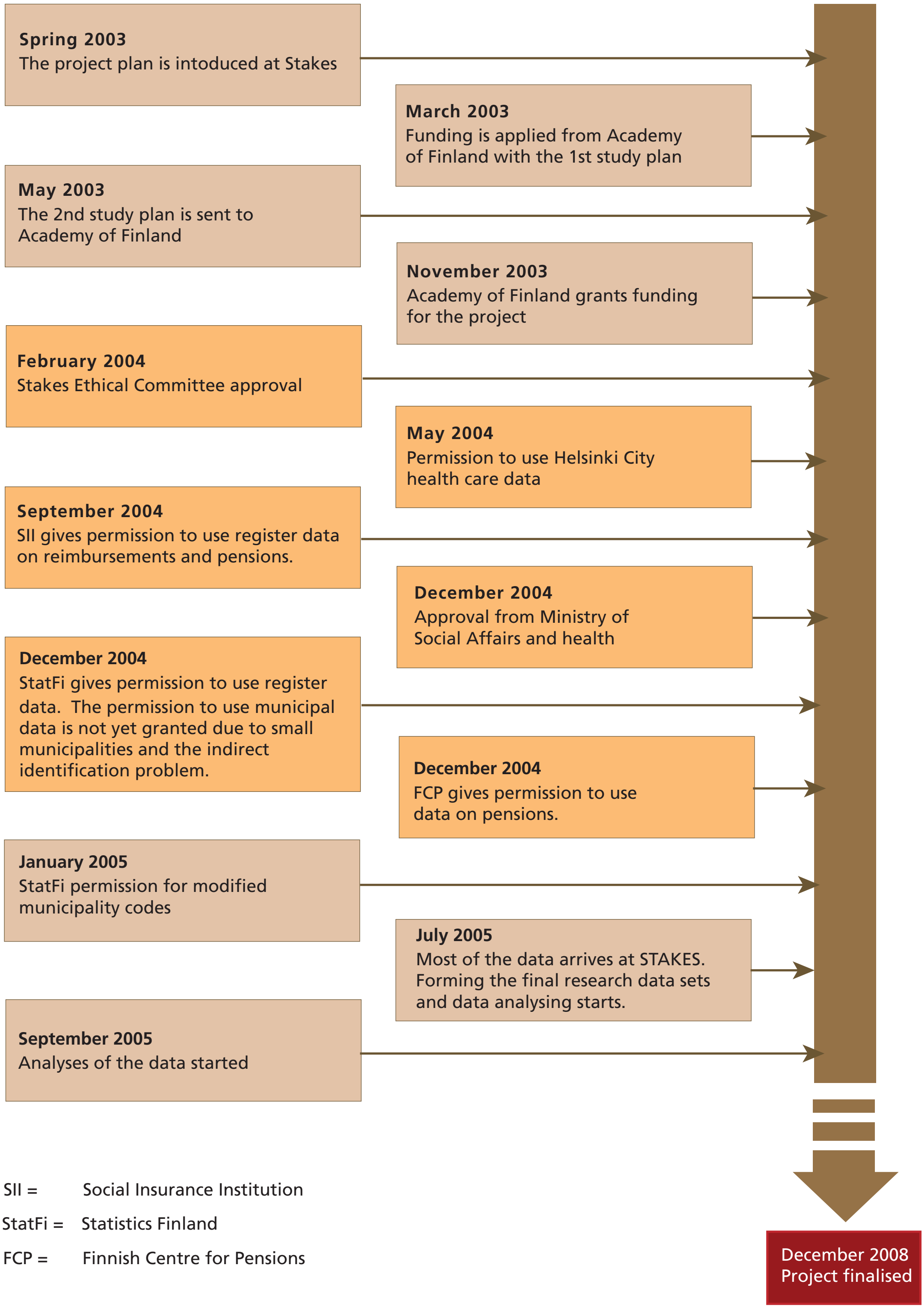
STAKES has a double role in register-based health research. It is simultaneously a research institute and a statistical authority. These roles are kept separate and STAKES-based research teams apply for permissions to use STAKES register data in the same way as projects conducted outside STAKES. In the REDD project, individual level data on medical treatment is collected from more than one administrative register (HILMO at STAKES and the Helsinki health care register) which are held by authorities who work under the Ministry of Social Affairs and Health. Thus, the project needed to be approved by the Ministry of Social Affairs and Health instead of the respective register authorities separately. Additionally, STAKES is the research site for the two projects the data of which are used in the REDD project. The application was submitted to the Ministry on May 11, 2004 and notice of approval was received on December 21, 2004.

In addition to the approval from the Ministry of Social Affairs and Health, permission to use the City of Helsinki data on utilisation of ambulatory health services and mental health services was applied for from the Helsinki Health Centre on April 27, 2004 and it was approved on May 5, 2004. Since the City of Helsinki has a strong interest in the research questions studied in the REDD project, it agreed to become a partner in the REDD project and experts from the Helsinki Health Centre agreed to participate in the project as external advisers in conjunction with their own work duties.

b) Statistics Finland

The REDD project exploits three registers/statistics maintained by Statistics Finland; the Causes of death statistics, and registers on Populations Censuses and Employment statistics. Additionally, since data from other sources was combined with the Statistics Finland data, as mentioned earlier, a detailed description of the other data sets as well as a logistics plan for the data compilation processes was included in the application.

In the REDD project, the study data include both individual and area level data. The fact that information about place of residence and several individual sociodemographic variables is included in the data brings up the question of the indirect identification. This was one of the issues considered when Statistics Finland deliberated permission on the compilation of the REDD data.



Spring 2003
The project plan is intoduced at Stakes

March 2003
Funding is applied from Academy of Finland with the 1st study plan

May 2003
The 2nd study plan is sent to Academy of Finland

November 2003
Academy of Finland grants funding for the project

February 2004
Stakes Ethical Committee approval

May 2004
Permission to use Helsinki City health care data

September 2004
SII gives permission to use register data on reimbursements and pensions.

December 2004
Approval from Ministry of Social Affairs and health

December 2004
StatFi gives permission to use register data. The permission to use municipal data is not yet granted due to small municipalities and the indirect identification problem.

December 2004
FCP gives permission to use data on pensions.

January 2005
StatFi permission for modified municipality codes

July 2005
Most of the data arrives at STAKES. Forming the final research data sets and data analysing starts.

September 2005
Analyses of the data started

December 2008
Project finalised

SII = Social Insurance Institution
StatFi = Statistics Finland
FCP = Finnish Centre for Pensions

FIGURE 1. The realisation of the REDD project.

For this part of the decision-making, Statistics Finland needed a detailed plan of the variables and their categories to be used in the study. For the REDD project, these plans were included in the annex of the application. In the course of the process, the variables to be released to the REDD researchers and the principles of how to avoid indirect identification were negotiated with personnel from Statistics Finland at several meetings. Finally, several limitations were applied to the REDD data as a condition for linkages. For instance, datasets for each REDD sub-study have been made anonymous separately. This means that every person that appears in more than one datasets has different sub-study specific identification codes in each of them. In small areas, socioeconomic information was considered to compromise anonymity, so municipalities had to be combined into units of more than 5000 residents. In the area codes released to the REDD researchers, small municipalities were aggregated into area units of more than 5000 inhabitants by combining municipalities according to their health care collaboration.

Application for permission to use data from the registers of Statistics Finland was submitted to Statistics Finland on June 7, 2004. The use of the first two data sources was granted on December 1, 2004. The use of Employment statistics had to be modified in terms of small areas (>5000 inhabitants). With these conditions the use of Employment statistics and Census data was approved on January 20, 2005 and for population at risk tables on December 2005.

c) Social Insurance Institution (SII)

Since the Social Insurance Institution has a strong interest in research questions investigated in the REDD project, it agreed to become a partner in the project. STAKES and the Social Insurance Institution signed an agreement of co-operation on September 1, 2004. The agreement of co-operation included information about the purpose of the project, the register sources used in the project, the research team and its leader, the plan for the realisation of the project, the responsibilities of both parties, the supervision of the project, the timetable of the project, reporting, the principles of data protection, and the plan for archiving or destroying the data after the project is completed. The experts of the Social Insurance Institution agreed to participate in the REDD project as external advisers in conjunction with their own work duties. As annexes, the agreement included the project plan and a logistic plan for the collection of the data. Permission to use the register data of the SII was included in the agreement, so that the REDD research team would have access to three data sources of the Social Insurance Institution: (1) Register of rights of special reimbursements for medicine costs, (2) Register on reimbursed prescription costs, and (3) the Register on pension decisions.

d) Finnish Centre for Pensions (FCP)

Data on early retirement and disability pensions were applied from the Finnish Centre for Pensions. Before granting the permission, the Finnish Centre for Pensions consults all the relevant pension institutions, which may prolong the application process. The Centre does not utilise a standardised application form. Instead, a free format application was used. The application was sent on September 10, 2004 and the permission to use the data was received on December 22, 2004.

e) Other research projects

In addition to data coming directly from register authorities, the REDD project exploits two research databases that are formed by other STAKES research projects, namely the projects of the Hospital Benchmarking System, co-ordinated at the Centre for Health Economics at STAKES, and Diabetes in Finland carried out in a collaboration with STAKES, the Diabetes Association and the Social Insurance Institution. In both cases, permission to use the project data was applied for from the management group of the project using a free format application since neither of these projects utilises a standardised application form. Permission to use the Benchmarking data was received on May 26, 2004 and for the Diabetes in Finland data on November 26, 2004.

The formal file description

The REDD data file description was compiled and sent to the Ombudsman's office on January 27, 2005. It is stored in the STAKES registry office.

5 Defining data linkages, variables and study populations in the REDD project

After setting the research objectives, specific research questions for profiling appropriately the phenomenon of interest need to be defined. In register-based health services research, defining research questions and variables is often a continuous process. Since the structure and content of existing register data are given, research questions have to be formulated while taking this fact into account. It is also possible that some research questions that are theoretically grounded cannot be directly investigated using existing data only. On some occasions in the study process, the data may also be revealed to differ from the researchers' preconceived notion on the content and quality of the data, which may prompt modifications for study settings and variables.

In the planning stage of the REDD project the research group defined the initial research questions based on previous research and the literature available from the research area. The specific conditions that the REDD project examines were originally chosen because they are common, serious health problems that affect the length and quality of life. These conditions also respond to medical care, so health service interventions can improve the natural history of the disease. The list of conditions was reviewed to see whether there was evidence of socioeconomic or area differences in treatment rates, theoretical models, or studies from other countries or studies of conditions with similar features.

When research questions are defined they need to be examined in relation to the available register data. Can all the necessary information be obtained reliably from the registers? The question is more complicated than it might seem by looking at the file descriptions. Administrative practices or the less apparent characteristics of the variables may cause the data to be different to what the researcher expects. If the registers do not provide enough accurate and adequate information, the research questions need to be modified. The process continues until it can be stated that the data is qualified to produce valid results to the questions presented. Even at the late stages of analysis, issues can come up that require a modifying of the variables, target populations or research questions. Many difficulties encountered during the process can be avoided by having a close contact with the register authority and personnel responsible for maintaining the register. Valuable information on the strengths and shortcomings of the register data can be obtained from these sources.

In this chapter, the examples of issues related to the planning of the REDD study data are given. First, the importance of a sensible definition of the study setting in the planning of register linkages for study data is underlined. Then examples on the definition of cases and individual and area level variables are given, and finally a short description of the different types of register linkages carried out for compiling the REDD study datasets are given.

5.1 Defining the study design and register linkages in the REDD project

The data obtained from health care registers contain information on the use of services or on administrative decisions, but in the study these register data refer to events taking place around some time point. In the REDD study the primary interest is to analyse the use of health services in population groups defined according to their socioeconomic standing or area of residence. This means that analyses also require data on the population at risk.

For planning the study designs and register linkages, it is important to understand to which population and to which time period (person-time) the data obtained from administrative registers are referring. In epidemiological research, this population definition in terms of person-time that is meant to be captured in a study is sometimes also called a study base or base population (Miettinen 1985). Precise information on the study base is important in planning study designs, for instance, for the accurate definition of study cohorts and their follow-up.

In the REDD project, the population of interest varies from sub-study to sub-study but in general terms it was defined to be the Finnish residents in the period 1991–2003. Thus the study population is open, which means that births and people immigrating to Finland are added to the study population. On other hand, due to a long study period, deaths have to be taken into account in forming the person-time data for the study population. From administrative registers, it is not feasible to get data that match exactly with the given definition of the base population or at the least the linkages for obtaining the data would be very complicated. The limitations are mostly small but as said above they are important in planning study designs and register linkages. For instance, various changes in individual sociodemographic variables, such as marriages, graduations, migration to and within the country, and the time point when the change occurred are in principle recorded in administrative registers but obtaining these data from primary registers would be costly and require a considerable amount of work. To simplify the linkages, it was decided that in the REDD project the base population is operationalised as a set of annual closed cohorts. Accordingly, the population at risk is a cohort defined in the beginning of each year separately and sociodemographic variables in the REDD data, such as education, marital status and place of residence, are referring to the situation at the beginning of that year. The cohorts' health care use, mortality and other relevant events are then followed in the registers for the subsequent year. This definition of the study population allowed the direct use of data in the Populations Censuses and registers for employment statistics that use December 31st as a reference date. Population at risk tables were defined accordingly as yearly cohorts using December 31st as a reference date. Tabulations were made using the sociodemographic and area variables of interest in the REDD study populations. The same population at risk tables are planned to be used in all sub-studies.

In some sub-studies the definition of the study population in terms of person-time brought about certain shortcomings. For instance, in the sub-study on amenable causes of death, socioeconomic differences in perinatal mortality cannot be analysed. In the Populations Census and employment statistics data, children receive socioeconomic data describing their families. However, these registers contain only the children who are alive on the reference date. Due to the study population definition and linkage scheme most children who died from perinatal causes cannot be linked to their families. Moreover, the data do not cover the changes in sociodemographic factors occurring between the reference dates. Consequently, a person who moves to another municipality or graduates during the year contributes only to the municipality of residence or educational group in the beginning of that year.

5.2 Defining cases in linked register data

In this context, a case is defined as a person identified in register data either afflicted by a specific condition or having undergone an operation. In the planning of the study data and register linkages for the REDD project, several registers were used to identify cases for different sub-studies. These registers were HILMO, the Causes of death statistics, and registers on reimbursed prescription costs and rights for special reimbursement. The detailed case definitions including diagnosis and procedure codes of the sub-study datasets are given in appendix 2. In the following, examples on issues to be considered in case definitions are addressed.

Defining cases and variables in the HILMO database can be laborious and time-consuming depending on the timeframe and the extent of the condition(s) that the project involves. Bundling up relevant diagnoses and translating codes between different ICD versions and between different procedure code classifications used often requires medical competence. Moreover, changes in the content and coding of the HILMO register need to be taken into account.

In the planning of the REDD study data, the HILMO register was used as a sole source for case definitions in three sub-study datasets: 1) elective surgery, 2) ambulatory care sensitive conditions, and 3) psychiatric hospitalisations. In the sub-study on elective surgical procedures, the cases were identified in the HILMO database by searching all elective procedures selected for the study recorded in the register between 1991–2003. The elective procedures included in the study were coronary revascularisations, hip and knee replacements, lumbar disc operations, cataract operations, hysterectomy and prostatectomy. In the sub-study on ambulatory care sensitive conditions, the relevant conditions were correspondingly identified in the HILMO register using a main cause of hospitalisation (i.e. the first discharge diagnosis) as an indicator. The definition of the conditions was based on earlier research and included the following diagnoses: asthma, COPD, hypertension, angina pectoris, and heart failure. These are considered to be conditions that should not require hospital care but treated successfully at ambulatory care (Bindman et al. 1995, Basu et al. 2002). In the sub-study on the psychiatric hospital care the cases were defined as all hospital admissions with psychiatric ICD codes.

As in the sub-studies based on the HILMO data, the registers were used to define cases in a similar straightforward manner for the sub-studies on amenable mortality and use of antidepressants. In the former, ICD codes for deaths from diseases that should not usually be fatal were defined using the available literature and making adjustments for the Finnish situation. These diagnostic codes were then used to identify cases in the Causes of death statistics. The definition of amenable mortality in the REDD project were based on a recent review by Nolte and McKee (2004) providing a comprehensive review of the research up to date on amenable mortality and the classifications used in earlier studies. The review also includes an updated list of amenable causes of death taking into account recent advances in medical treatments and a comparative analysis of amenable mortality in European countries using the WHO statistical records. The classification of amenable causes of deaths by Nolte and McKee was chosen for use after some modifications in the REDD project. The list was checked and causes not included in routine statistics in the WHO statistical records though considered amenable by Nolte and McKee were added to the list. These include chronic rheumatic heart disease, malignant neoplasm of testis, Hodgkin's disease, septicaemia and malignant neoplasm of colon. Additionally, asthma and COPD were included in the list. Deaths due to misadventures to patients during surgical and medical care that appear in Nolte's and McKee's list were not included in the REDD project data set, since these causes cannot be determined from the Finnish Causes of death statistics. The Causes of death statistics were also used to define a subpopulation of the coronary heart disease sub-study by extracting all CHD-related deaths during 1994–2003.

For the study on use of antidepressants, the register on reimbursement of prescription costs was used in defining users of these medications. All filled prescriptions for the medicines with ATC code N06A were searched for in the register for the years 1994–2003.

The linked register data from different sources provides an opportunity for a complex definition of cases based on several registers. In the REDD project this opportunity was used in the sub-study on coronary heart disease so as to provide comprehensive data on patient with diagnosed disease. In register-based studies where cases are defined from different registers with dissimilar codes, it is of primary importance that the equivalency between search criteria is assured. In the REDD sub-study on coronary heart disease the cases were defined using HILMO, and registers coming from Statistics Finland and the Social Insurance Institution. One of the objectives of the definition was to identify all incident coronary heart disease cases independent of presenting symptoms among the Finnish population aged 40 years or older in 1994–2003. Cases were extracted from the Causes of death statistics and the HILMO database by searching ICD9 and ICD10 codes for coronary heart disease. Additionally codes for invasive coronary operations (angiography, angioplasty (PTCA), coronary artery bypass operation (CABG)) were searched in HILMO. From the register on the right for special reimbursement the codes assigned by the Social Insurance Institution for coronary heart disease (206) were searched for. Additionally, the register or reimbursement of prescription costs was searched for reimbursed medicine purchases with ATC code for nitrates (C01D). According to earlier research, nitrate prescriptions are a sensitive and specific marker of a doctor's diagnosis of angina and nitrates are only used for treating coronary heart disease (Cannon et al 1988).

5.3 Defining individual level variables

In defining individual level variables in register-based research, the researchers are dependent on variables in primary source registers. In planning study data, it is important to be fully informed on the content of source registers and on possible changes in the content and variable definitions during the study period. In the following, examples on considerations in variable definitions are described for the REDD project.

Statistics Finland

In the REDD project, data from Populations Censuses and Employment statistics of Statistics Finland are used to obtain sociodemographic background information on study populations. In the REDD project, it was planned that this information would be identical for each sub-study (appendix 3). The content of Populations Censuses compiled for every fifth year and the annual Employment statistics are mostly identical but there are some exceptions. Detailed information on the content of the registers and classifications used is available at the website of Statistics Finland (www.stat.fi).

Sociodemographic data for the REDD project was obtained for the period 1990 to 2002 (study period 1991–2003). Information about family, housing, place of residence, employment, education and income used in the final study data sets is taken from the year preceding data entry, since the Populations Census and employment statistics data have the reference date of December 31st in each year.

Because of the potential for identification, Statistics Finland did not release exact date of birth for the research subjects, but instead, age rounded into months at the end of the year. Age is given for each year separately. Missing information in the variable indicates that the person was not resident in Finland at the end of the year in question. This information is often necessary when final study populations are defined.

Two versions of classifying educational level were included in the data: the Finnish classification issued by Statistics Finland and the international ISCED classification issued by UNESCO. The Statistics Finland classification corresponds quite accurately to the categories used in the ISCED system, which is recommended for international publications. The ISCED classification is only available for the Populations Census years of 1990, 1995 and 2000. Both classifications are formed from the register of educational achievement, and, thus, those with basic education only are represented with missing values.

Data on socioeconomic status based on occupation are only available for the Census years. The Statistics Finland registers use two official classifications of socioeconomic status issued in 1980 and 1989. The third available classification, derived socioeconomic status, is based on the classification originally developed in a long-standing Finnish research project on socioeconomic mortality differences (Valkonen et al 1993). The definition of derived socioeconomic status is available at the web site of the Population Research Unit of the Department of Sociology of the University of Helsinki (www.valt.helsinki.fi/sosio/pru/). The classification of derived socioeconomic status is often used in the analyses, since it is backdated to the 1970 Census for those outside the labour force. However, it is only available for the 1990 and 1995 Populations Censuses because it is based on the 1980 classification of socioeconomic status, which was not used in the Populations Census in 2000. For the 2000 Census, socioeconomic status is only available for those belonging to the work force at the end of 2000 using the 1989 classification. For planning the REDD data, the 1980 and 1989 classifications of socioeconomic status were compared in a sample of the 1995 Populations Census data. The comparison showed that the classifications assigned the sample into socioeconomic groups differently to the extent that the 2000 socioeconomic status could not be used in updating occupational social class. Although socioeconomic status does not usually change among the adult population in the short term, the comparisons of socioeconomic differences in terms of occupation-based socioeconomic status may be somewhat biased for the younger age groups across the years.

Although the REDD dataset does not include any direct identifiers, exceptional characteristics such as very large income can cause a potential for indirect identification. To release these kinds of data is against the policy of Statistics Finland. Potential indirect identification can be avoided by variable definitions. Variable coding can be modified and categories compiled so that the data will not compromise anonymity. For instance, income variables in the REDD project were rounded so that maximum personal net income was EUR 50 000 and maximum family net income EUR 100 000 per year. For the same reason information on language was defined as Finnish, Swedish and other. Furthermore those aged over 100 years at the time of data entry were defined as 99.99–years-old.

Since one of the main aims of the REDD project is to analyse the effect of area characteristics and regional variation in health services use, special effort was made in order to form meaningful totalities from the municipalities or health centre areas with less than 5000 inhabitants. Statistics Finland thought that using area units with less than 5000 inhabitants would risk the anonymity of the data. These municipalities were combined with municipalities with which they co-operate in the organisation of health and/or social services according to their www-pages or information received from these municipalities, with new codes being created to them. If more than two municipalities with less than 5000 inhabitants organised their services together, any number of combinations of at least 5000 or more inhabitants was created (Appendix 4).

Furthermore, some municipalities have merged with larger municipalities during the study period. Since municipality codes needed to be defined in the same way for each year, the 2004 distribution of municipalities was used. Small area codes were obtained for the ten largest cities (Appendix 5).

Besides the sociodemographic data, the basic data from Statistics Finland for all REDD sub-study datasets included data from the Causes of death statistics. These data were obtained for the whole study period. Information was extracted from the longitudinal causes of death database linking all causes of death statistics from the years 1991–2003. Both main and immediate causes of death were requested to determine the cause of death in the REDD-studies. Data on death included information also on contributing causes and intermediate phase causes of death. Information on where the patient died, whether death was alcohol related and how the cause of death was determined were included in the data.

STAKES

Hospital discharge records linked from the HILMO register to the REDD data vary from sub-study to sub-study according to the sub-study objectives and definition of study cases. However, the linked records contain all variables in the basic register. Supplementary entries for heart patients were only included in the datasets on coronary heart disease and diabetes. Correspondingly, supplementary entries for psychiatric patients were included in the dataset on psychiatric hospital care.

In the REDD study period of 1991–2003 the content, definitions and classifications of the HILMO register have changed several times. These changes have briefly been described in chapter 3.1. For the linkages, the comparability of the register content in each study year was assessed and a new list of variables for the longitudinal datasets was redefined. This list of variables in Appendix 6 separates the variables with different definitions. Variables for the analysis are formed from the relevant events and diagnoses when the final study datasets are constructed.

Social Insurance Institution

Special reimbursement rights are entered in the SII data system with a special SII code, not with the ICD10 code. Therefore before the data request is made to the SII, the ICD10 codes have to be converted into SII codes. In the prescription register, drug types are defined using the ATC-classification. The 1994 data do not cover all pharmacies.

Information of prevalent disability pension and early retirement was taken from the year before the beginning of the follow-up and each first incident pension during the follow-up was then gathered from the yearly pension registers (Appendix 7).

Finnish Centre for Pensions

The pension data from the Finnish Centre for pensions includes mainly the same variables as the Social Insurance institution pension data (Appendix 8).

Helsinki City

Data from Helsinki City include ambulatory visits for 1998–2003. Due to modification of the structure of the database in 2001 some variables only exist for the 1998–2001 period and others only for the 2002–2003 period (Appendix 9).

5.4 Defining spatial variables

Area-level variables were identical for each sub-study. Clinical data was gathered from the SII and the Finnish Centre for Pensions' registers, and sociodemographic data from the SOTKA database. SOTKA contains municipal-level information on population, economy, housing, morbidity as well as labour, costs and usage of the health care services. It is maintained by STAKES, which collects additional information from Statistics Finland, the Social Insurance Institution and the Finnish Municipal Association among others. The database has over 3000 facts and 200 proportional indicators on every municipality in Finland. SOTKA is updated regularly and the information is mostly available the following year. Because in the REDD project the basic area division is based on municipalities and municipal associations, the project is able to use this kind of municipal-level information to analyse area-level variation. The REDD project uses area variables, such as unemployment rate, single parent households, rate of households that live under the poverty line, crime rates, and educational level in the area (Appendix 10).

The REDD project uses area-level information on health and health care as well as on sociodemographic factors. For area health indicators, the Social Insurance Institution provided the project with the annual municipal numbers of special reimbursement rights according to diagnosis groups, reimbursements for medicine use, reimbursements for public health care and reimbursements for private sector doctor fees. The Social Insurance Institution also delivered tabulated data on early retirement and disability pensions by specific diagnosis groups. Data on the disability pensions were gathered for mental disorders, cardiovascular and musculoskeletal diseases, poisonings and injuries, and other diseases together. Municipality-level information was only available for prevalent pensions for the whole study period, with data on incident pensions only from 2000 onward (Appendix 10).

5.5 Data linkages

In advanced register-based research, study data often compile information from several primary register sources. In the REDD project, the purpose was to comprehensively cover information available on the interest areas of the study in the administrative registers. In the REDD project, the data compilation process varies between sub-studies. For the eight sub-studies of the project, five different principal methods of compiling study data sets were used.

(1) Straightforward compiling was used when all the information used in the sub-study could be obtained from a single administrative register authority and the target cases were extracted from a single register. The sub-study examining avoidable deaths exemplifies this, the administrator being Statistics Finland. The data include all persons whose cause of death was indicated to belong to the avoidable causes in the Causes of death statistics. Data from the Employment statistics, Populations Censuses and area level indicators were then linked to these individuals. Since avoidable deaths are rare, frequencies are likely to remain very small in some regions and to avoid the possibility of indirect identification the data were delivered in tabulated form.

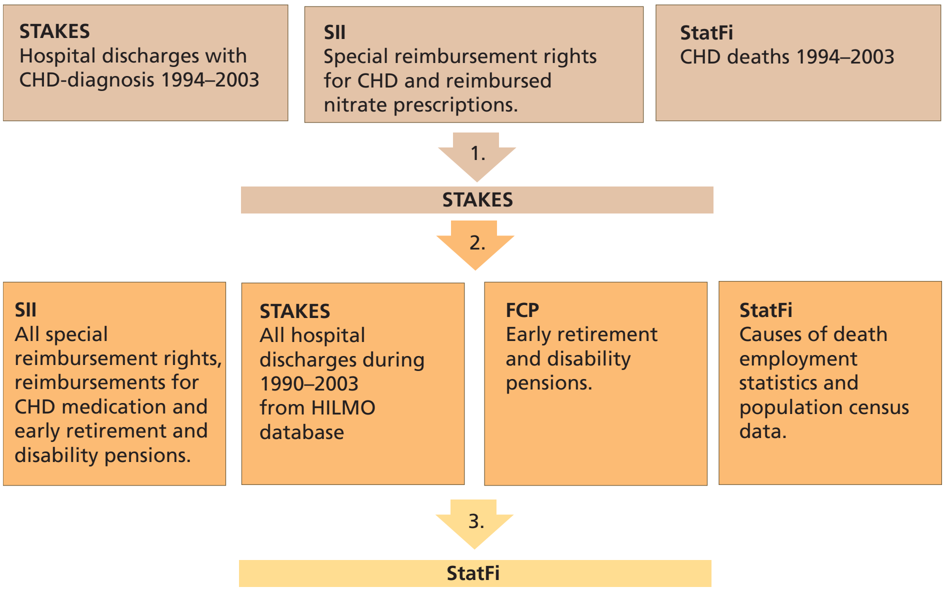
(2) The second method to define the target cases for the REDD sub-studies was to identify individuals with a medical condition, procedure, or medical treatment of interest. In three of the REDD sub-studies the target cases were obtained solely from HILMO. The elective surgery sub-study covers individuals that had undergone specific operations. The data for the ambulatory care sensitive conditions sub-study and psychiatric hospital care sub-study cover people that had hospital admissions with specified diagnoses. Personal identification codes and the data for hospital care were then sent to Statistics Finland so as to link data from the Censuses, Causes of death statistics and Employment statistics to each individual. Personal identification numbers were then replaced with artificial ID-codes before returning data to the research group.

The sub-study examining the use of antidepressants followed much of the same order in data compiling, but the target cases were picked out from the Social Insurance Institutions register for reimbursed prescriptions. Personal identification codes and relevant medical information for individuals who had sought reimbursement for antidepressant use were sent to Statistics Finland for linking data from the Censuses, Causes of death statistics and Employment statistics to each individual. Personal identification codes were then replaced with artificial ID-codes before returning data to the research group.

(3) The third method was applied in the City of Helsinki sub-study where data on the population at risk was used as a base to compile data. The City of Helsinki maintains a register of all individuals that have been residents of the City. Personal identification codes for persons who had been residents at least on one December 31st in 1997–2002 were selected and used to collect hospital care data from HILMO and the Benchmarking database, the Social Insurance Institution data for special reimbursement rights, reimbursement of use of private sector investigations and care, the Social Insurance Institution and the Finnish Centre for Pensions data for disability pensions and early retirement and the registers of the City of Helsinki on use of municipal health services and mental health services. These data were sent to Statistics Finland so as to link with data from the Censuses, the Causes of death statistics and Employment statistics for each individual. Personal identification codes were then replaced with artificial identification codes before returning data on health care users to the research group. Data were collected only for residents on December 31st in each year since the City of Helsinki register on municipal residents covers limited number of changes of address.

(4) The fourth method of compiling data for the REDD project was exploiting already existing research data sets. The diabetes sub-study made use of a register-based dataset collected from HILMO, the Causes of death statistics, special reimbursement register and the prescription register in the Social Insurance Institution, the Finnish Register for Visual Impairment, and the Birth Register. The research group applied for permissions to use the already existing dataset from the relevant authorities and from the Diabetes in Finland project group. The data set did not originally include information about the socioeconomic background of the diabetic population and the supplementary information on specialist treatment for patients with coronary heart disease from HILMO. Therefore, the data was complemented with this information from HILMO in STAKES and from the Censuses and Employment statistics in Statistics Finland for each individual. Personal identification codes were then replaced with artificial identification codes before the data were returned to the research group.

(5) For the fifth method of constructing the datasets, three stages needed to be undertaken; in REDD this was done in the sub-study on coronary heart disease. The first step included picking out the target cases, i.e., searching for all individuals that have coronary heart disease. Several registers were used in this study to identify cases: HILMO, the register on entitlements for special reimbursement and the prescription register in the Social Insurance Institution as well as the Causes of death statistics in Statistics Finland. Personal identification codes for individuals who had been hospitalised with, achieved special reimbursement right for, used nitrates or died of coronary heart disease were extracted from each of the registers. Since most target individuals appeared in several of the registers, the datasets of identification codes were merged by the statistical authorities in STAKES. At the second stage, all HILMO information, information concerning special reimbursement rights, medicine used for cardiovascular diseases (ATC codes C01-C10) in the Social Insurance Institution, and information on disability pension and early retirement from the Social Insurance Institution and the Finnish Centre for Pensions were linked to these individuals by the relevant authorities. Third, all data were sent to Statistics Finland for linking data from the Censuses, Causes of death statistics and Employment statistics for each individual. Personal identification codes were then replaced with artificial identification codes before returning data to the research group. The compilation of the REDD coronary heart disease study data is schematically presented in the Figure 2.



The purpose in data collection to identify a total CHD population in Finland in 1994–2003 to enable research on health care use and outcomes of CHD

1. Personal IDs collected from all relevant register authorities. IDs from different register sources merged at STAKES
2. IDs sent to register authorities for compliation of relevant data for the research questions
3. Register authorities send the relevant data to StatFi, where IDs replaced by artificial research numbers

FIGURE 2. Data compilation process for the Coronary heart disease sub-study

It was agreed that the register authorities in STAKES would carry out the merger of the personal identification codes to guarantee that all register authorities would receive the same identification codes of cases for the next step of data compilation. The data collection was done in three steps to guarantee that all relevant data would be collected for all the cases from each of the registers used.

(6) Additionally, the REDD project exploits area level information about socioeconomic characteristics of the municipalities, structural characteristics of the health care system, time and distance to care, and health services use and morbidity. These data are obtained from SOTKA and the Social Insurance Institution, and are attached to the data sets through the municipality code. The 2004 division of municipalities is used for all the study years. In the case of merged municipalities (< 5000 inhabitants, administrative mergers) a population size adjusted mean of the data is used.

In the REDD project, other statistical authorities sent their identifiable source data to Statistics Finland, which linked the datasets using personal identification numbers and formed study data according to the detailed plans from the REDD researchers. After linkages, personal identification numbers were replaced with artificial identification codes before handing the data forward to the research team.

6 Forming and analysing final research datasets

Since register data come in different forms in the data compilation process, e.g. some of the data are event based and some are individual based, final study data sets need special definitions in each sub-study. Further, at the data compilation phase, broad definitions of conditions or procedures are often used to guarantee the inclusion of all relevant cases, so the case definitions need to be reconsidered when forming the final sub-study datasets. Additionally, in some cases the target population needs to be redefined to take into account the imprecision of register data in the definition of cases.

This chapter gives two examples of this process in the REDD project: A sub-study of variations in the use of common surgical procedures and a sub-study of socioeconomic and spatial differences in mortality among diabetic people. We describe briefly how study questions, settings and target populations were defined in these studies, the testing and cleaning done to the datasets, definition of variables in these sub-studies and finally, preliminary plans for analyses.

6.1 Studying variations in the use of common surgical procedures

Study questions and definition of study cases and design

The sub-study on the use of common surgical procedures will explore trends in the socioeconomic and spatial patterning of hospital care in Finland from the early 1990s to the early 2000s. It will analyse access to elective interventions, such as coronary revascularisation, hip and knee replacement, cataract extraction, prostatectomy, and hysterectomy. The focus is on the following issues:

- How is the overall socioeconomic and geographical patterning of hospital use modified by spatial and socioeconomic differentiation across the country, and in the capital region? Specifically, what impact do spatial differentiation and deprivation have on socioeconomic differences in hospital care? Have aggravated levels of individual deprivation (long-term or repeated unemployment and income inequalities) resulted in an underclass who have poor access to health services?
- To what extent do organisational factors in health services, such as the specialisation of local services, and public provision of specialist ambulatory care contribute to the changing patterns of socioeconomic inequalities in access to and utilisation of hospital care?
- What is the impact of private service provision and hospital referrals from private practitioners on socioeconomic and social inequities in access to hospital care and on changing patterns of hospital use?
- What impact does co-morbidity have on socioeconomic and spatial differences in health service use?

The dataset consists of all hospital discharges for the elective procedures of interest between 1991 and 2003. Each discharge record for the study years was considered as a case. This data were linked to sociodemographic data from Statistics Finland for the years 1991 to 2003. Each record was linked with information from the year before data entry as well as the derived socioeconomic position from the Census of 1990 for cases entering the data between 1991 and 1995 and from the Census of 1995 for cases entering the data between 1996 and 2003.

Testing and cleaning the data

The definition of procedure codes is presented in Appendix 11. The data for hip and knee replacements was compared to numbers from the National Agency for Medicines' register of orthopaedic endoprostheses (Nevalainen 2003) and data on angioplasties (PTCA) and bypass operations (CABG) to the Finnish Heart Association's register information (Finnish Heart Association 2004) for each of the study years. The comparison showed a number of CABGs not accounted for in the Finnish Heart Association and a closer examination of the HILMO data revealed that in some cases when the hospital admission was followed by a hospital or department transfer the procedure codes were registered in multiple discharge records. However, data on hospital admissions without procedures were not collected in the dataset. Hospital admissions, which started on the day (+/-1 days) of the previous hospital discharge were defined as belonging to the same admission (i.e. ward or hospital transfers). Additionally, CABGs reported from hospitals that did not have these facilities in the hospital admission year according to the Finnish Heart Association register and that were not connected to admissions to hospitals that did have these facilities were removed from the data as misclassifications. Data on knee and hip replacements were similar in terms of first operations, but data on reoperations differed from those reported in the Endoprosthesis register of the National Agency for Medicines, which only reports reoperations performed at least four months after the initial operation (reoperations within four months are classified as complications). Since the aim was to examine elective surgery, hip replacement operations due to fractures (ICD10 code S72) were removed from the data. Additionally, it was assumed that first endoprosthesis operations could only be performed on one person twice. Data on hysterectomies and prostatectomies were checked by sex. Data on cataracts done in the private sector was compared to Social Insurance Institution data on reimbursed cataract operations in 2002 and 2003. The comparison showed that almost two-thirds of reimbursed operations were missing in the HILMO data suggesting that they have been done policlinically. Further analysis of cataracts was therefore postponed and it was decided to compliment the dataset with data on these operations.

Patients under age 25 at the beginning of the entry year were removed from the data. The missing values of age in the sociodemographic dataset indicate that the persons in question were not permanent residents in Finland and were therefore removed, since the database does not include data on operations done in other countries.

Definition of variables

Year of data entry was defined by hospital admission instead of discharge. Age was defined as age on December 31st in the year before data entry to match the age groups in the population at risk tables. Finally, age was classified into five year age bands. Data on socioeconomic status: family net income per OECD consumption unit and education were taken from the year before data entry, derived socioeconomic position from the Census (see chapter 3.2, and appendix 3) before data entry, and for the years 2001–2003 from the 1995 Census, since because of changes in classification, a compatible variable of derived socioeconomic status is not included in the 2000 Census. The area identifications, namely hospital district, health centre area, municipality, and municipal sub-area codes, were referring to December 31st of the year before data entry. Marital status, native language, main activity and number of months unemployed were derived from the year before data entry. A separate variable for the year of data entry was defined.

Population at risk tables

For statistical analyses, the population at risk was defined as the Finnish population in the age group 25+ years. Tabulated data on the population at risk were derived from the Employment statistics for study years by each of the demographic and socioeconomic variables and area variables used in the study.

Preliminary plan for analysis

The preliminary plans to analyse socioeconomic and regional variations in the use of common surgical procedures address three overall research questions:
1) How have socioeconomic differences in the procedures developed during the study period in the whole country?

- Age-standardised procedure rates/100 000 each procedure by income quintile for each year to be reported in the Inequity report 2006.
- Concentration indices and their 95% confidence intervals for all study years.
- Explorative analysis of the lowest income quintile.

2) What kind of regional differences are there in elective surgery?

- Are there hospital district differences, university hospital district differences or differences between types of municipalities in procedure rates?

3) Can a development of an underclass be detected in the data? Exploratory analysis of the contribution of socioeconomic variables, unemployment, living alone and region in differences in use of procedures.
4) What is the contribution of area level factors to socioeconomic differences in the use of common surgical procedures? (Multilevel modelling)

- Can the socioeconomic differences be explained by the way health care is organised at the local level (for instance in terms of public provision of specialist ambulatory care, provision of private care, referral practices (HILMO), or specialisation of local hospital services)?
- Are there spatial differences in the distribution of socioeconomic differences and how can these potential differences be explained? What is the relative contribution of area level deprivation, the supply of services, and the organisation of services?

6.2 Socioeconomic and spatial differences in mortality among diabetic people

Study questions and definition of study cases and design

The study on mortality among diabetic people will assess socioeconomic and regional differences in mortality among diabetic people in Finland 1991–2003. The main focus is on the following issues:

- To evaluate how socioeconomic mortality differences among diabetic people in Finland have changed compared to an earlier study (Forssas et al. 2003) that analysed the years 1981 to 1985 and 1991 to 1996, and showed emerging socioeconomic mortality differences among diabetic people.
- To examine whether socioeconomic mortality differences among diabetic people have increased in some regions, and whether the potential regional inequities in mortality reflect poor quality of care in the whole diabetic population or in some socioeconomic groups.
- To examine the possible causes of socioeconomic and regional mortality differences among diabetic people.

The study population (N=308 447) has been modified from the diabetes register for the years 1988 to 2002, which was gathered in a collaboration project 'Diabetes in Finland' by STAKES, the Diabetes Association and the Social Insurance Institution (see chapter 3.1). The project used three main indicators based on health registers to establish the diabetic population:

- Any diagnosis of diabetes classified according to ICD-9 or ICD-10 in HILMO between 1988 and 2002
- Entitlement for reimbursement for antidiabetic medication in the Social Insurance Institution register on entitlements for special reimbursement between 1964 and 2002
- Entries of diabetes drugs reimbursed in the Social Insurance Institution register on reimbursed prescriptions between 1994 and 2002.

In addition, the diabetes register data were complemented with information from various registers including HILMO, the register on entitlements for special reimbursements, the register on reimbursed prescriptions, the Causes of death statistics, Birth Register and Visual Impairment Register.

In the REDD project, the study population consisted of people between 25 and 84 years suffering from diabetes at the end of 1990 and those who contracted diabetes by the end of 2002. Deaths were followed from the beginning of 1991 to the end of 2003. In order to investigate socio-economic mortality differences, sociodemographic data from the annual Employment statistics and longitudinal Census database were linked to the data.

Testing and cleaning the data

To further specify and clean the study data some cases were excluded from further analyses. Those who died at the year of the onset of diabetes or before 1991, those aged under 25 or over 85 during the entire follow-up and those with the onset of diabetes after 2002 were ruled out. Moreover, those with gestational diabetes and some cases with insufficient information were excluded. After the cleaning, the study population comprised 270 034 persons.

Population at risk tables

We formed the yearly tabulated data of the population at risk consisting of people considered diabetic at the end of the preceding year. The number of persons at risk at the beginning of the year as well as person days in age groups and other subgroups were calculated for the analysis. Likewise, the number of all deaths and deaths from specific causes were calculated for the same subgroups. Since age was defined by month and year, birthday was estimated to be the 15th of the month. Diabetic people under 25 years entered the population at their estimated 25th birthday. Exit from the population occurred at the equally estimated 85th birthday or death occurring before. Transition to the consequent age group occurred on the birthdays. The total amount of person years in the study data was 1 598 863 and the number of deaths 82 486.

Definition of variables

In defining diabetes type, the REDD dataset uses the classification utilised in the Diabetes in Finland project (Niemi and Winell 2005). Accordingly three diabetes groups are used: 1) insulin-dependent diabetes (type 1), 2) non-insulin-dependent diabetes (type 2), and 3) undefined type of diabetes. Due to the rules of definitions, it was only possible to define the type of diabetes from the beginning of 1995, as the register on reimbursed prescriptions was launched in 1994.

The duration of diabetes was defined as time from the onset of diabetes in years to the beginning of each year.

Data on socioeconomic status were derived from the Employment statistics and Populations Censuses in a corresponding way to the sub-study on common surgical procedures. Family net income per OECD consumption unit and education were taken from the preceding year; the derived socioeconomic position was taken from the Census before the index year, and for the years 2001 to 2003 from the 1995 Census due to the incompatible classification of socioeconomic status in the 2000 Census mentioned earlier. Area level variables (hospital district, health centre area, municipality, and municipal sub-area codes) were taken from the last day of the preceding year. Marital status, native language, main activity and number of months unemployed were similarly derived from the preceding year.

Deaths in the study population between 1991 and 2003 were obtained from the Causes of death statistics. Information on cause of death was limited to the underlying cause. Between 1991 and 1995, causes of deaths in Finland were classified according to the International Classification of Diseases, 9th revision (ICD-9), and after that according to the 10th revision (ICD-10). These diagnoses were converted to the codes of the 8th revision (ICD-8) in order to give comparable information on the mortality changes between this study and the previous study analysing mortality between 1981 and 1985.

Preliminary plan for analysis

According to the preliminary plans, the sub-study on socioeconomic and regional mortality variations among diabetic people will address the following main research questions:

1) How have socioeconomic differences in overall mortality and mortality from specific causes changed among diabetic people between 1991 and 2003?

- Has the trend of increasing socioeconomic mortality differences observed in earlier research continued?

2) What are the regional differences in mortality among diabetic people?

- Are there regional variations in diabetic people's overall and cause-specific mortality by hospital districts, health centre areas, municipalities and municipal sub-areas?
- Are there regional variations in the extent of socioeconomic differences of diabetic people's mortality?

3) Do contextual area-related and health care factors explain regional and/or socioeconomic variations in mortality among diabetic people?

- Are high mortality or larger socioeconomic differences in mortality related to regional factors, thus potentially indicating sub-standard quality in diabetes care?

7 Summary and conclusions

Finland offers good possibilities to use administrative health and health care registers in research. The register system is comprehensive and the quality of register data is generally adequate for most research purposes. Legislation regulating data confidentiality and databases gathered by official agencies and register authorities is in principle supporting research use of administrative registers. Moreover, in addition to sensible data protection legislation, the unified personal identification code applied to virtually all administrative registers containing individual information allows the effective use of register information through linking information in different registers. Finland is also one of the few countries that include comprehensive data from the public and private services in their health registers. It is not surprising that the favourable research environment for register research has resulted in internationally acknowledged quality health and health services research.

This report aimed to give a general overview of the process of conducting register-based health services research in Finland. It was written on the premise that reporting the process of planning and executing data compilation in one research project may be useful for other researchers planning register-based health services research. For the most part, register-based research projects do not, in general, differ much from other ways of collecting research data. However, they have some specific features which are related, for instance, to the fact that these studies are based on secondary data not originally designed for research. The process for applying for permission to use administrative registers for research purposes and the conditions under which the register authorities can release their data for researchers are determined by specific laws. However, issues related to study design and data structure as well as variable and case definitions need to be solved study by study.

The current report presents the process for conducting register-based health services research using one research project ‘Regional disparities, social segregation and socioeconomic patterning: Where do inequities in access to health care arise?’ (REDD) as a study case. As the report shows, the project nevertheless provides many lessons that can be generalised. The overall aim of the REDD project is to evaluate the association between area variation, urbanisation, socioeconomic patterning, the development of social segregation in Finland and their contribution to the changing regional and socioeconomic gradients in health and access to health care. The REDD project is a good example on how register data enable us to examine issues that could not be studied otherwise or that would be very expensive to study.

While the report is meant to be a sharing of experience in the research use of administrative registers, the report mainly addresses issues related to specific features of administrative registers used in the REDD project, the legal and administrative regulations of data availability for research purposes, and the methodological questions specific for register-based health services research. To improve comprehensibility, the report includes the research background of the REDD project and gives two examples on REDD sub-studies with descriptions on the construction of sub-study datasets, the formulation of specific research questions and some starting points for preliminary analyses. In particular, the report has tried to address various challenges and shortcomings met during the implementation of the REDD project. For planning the time budget of corresponding projects, the report also gives the timeline of the REDD project, which indicates rough estimates for time required for the different phases of the project.

The use of administrative register data for research has several advantages. Administrative registers allow the formation of large datasets at relatively low costs and thus enables the examination of diverse phenomena related to health and health care. Examples on the use of administrative registers in health and health services research are given in many publications reviewing the topic, such as Gissler (1999), Gissler and Haukka (2004), Keskimäki et al. (1997), Lewsey et al. (2000), Rosén (2002), and Valkonen et al. (1997).

The REDD project and its research questions show effectively the potential of health care registers for research use. Without comprehensive data it would not be feasible to explore the relationship of individual level sociodemographic factors and factors related to municipal or sub-area level contextual and health care features. Comprehensive register data allow the analysis of various population subgroups and the study of relatively rare health conditions and hospitalisations due to specific diagnoses and procedures. Continuously accumulating administrative registers also give an opportunity to investigate time-trends and to construct study designs based on long-term follow-up of the study population. Individuals can be linked using secure methods, a series of unique identifiers and strict access rules that mean that individuals cannot be identified, either directly or indirectly. In this way it is possible to ensure continued use of the individual but non-attributable data that is necessary for a lot of public health research so as to answer questions that could not be answered using methods that are based on explicit consent given the inevitable attrition rates.

While it is not the focus of this report, such large data sets are often used to identify rare adverse outcomes from common interventions, and to identify sub-populations that are more likely to be affected. In equity research, these subpopulations are often vulnerable groups among whom limited use of services is a concern. Few other research methods are able to examine the use of routine services and outcomes of interventions among this group. People at high risk of social exclusion are rarely recruited into randomised controlled trials and the outcomes achieved by special projects may not be transferable to routine practice. In terms of statistical analysis, administrative registers often show advantages as well. In register data, the absence of response bias introduced by systematic differences in response rates, consistency of data definitions and the size of the data sets ensure robustness and precision in statistical modelling.

Despite many good properties in terms of applied research, administrative register data also offer several methodological challenges. Transforming data from original registers to the form of a study dataset suitable for the actual research questions of interest requires good knowledge of the contents and features of the registers used. Careful planning and awareness about the deficiencies and limitations of the data are required for a successful study and valid results. A major limitation in administrative registers is evident: entries in register records are not usually theoretically grounded but instead based on administrative practices and needs. For this reason, study data obtained from administrative registers are often poor regarding research questions and theoretical content. To some extent this shortcoming can be circumvented by supplementing the study data by using several primary register data sources. In REDD sub-studies this has been attempted by supplementing hospital discharge data with data from the Social Insurance Institution register on eligibility to special reimbursement, thus indicating comorbid conditions. In general, however, available data and variables in registers limit the potential research questions to be investigated.

To understand the process of the compilation of the primary register to be used in the study and the way the entries are made into the register is important for assessing the validity and reliability of the data. The quality of register data depends on various factors related to the compilation process of the register. For instance, if the data is based on administrative decisions, such as health insurance benefits, the data is probably rather reliable but the coverage of the register may be deficient, which means that while most entries are recorded correctly, the number of false negatives may be high. Moreover, most authorities pay attention to the completion of data on so-called compulsory variables, particularly where these are associated with financial incentives. Finland is no different from other countries in this regard. The reliability of entries on items based on discretionary decisions may also be inadequate. In HILMO the diagnosis of myocardial infarction (MI) which is based on clear diagnostic criteria is considered to be recorded accurately, but the diagnosis of unstable angina—which is more discretionary compared to MI in clinical settings—is not as consistently recorded in the discharge register (Häkkinen et al. 2002).

The design of study data compilation from original registers is of primary importance in register-based research because the compilation process defines the structure of study data as well as the study designs that can be applied in the study. Therefore the planning of a register-based study should comprise a careful consideration of possible study designs to be applied on the planned study datasets. Particularly in complex datasets that gather cases from several registers, these requirements of the study designs are to be taken into account in the practical planning of the 'logistics' of the data compilation.

Another important issue in planning the compilation of register-based study data is the definition of the study population and cases that use information given in registers. In contrast to survey research, register-based health services research often deals with total patient populations and therefore carefully planned linkage schemes are needed to ensure that all relevant individuals are captured in linkages. An inadequate definition of inclusion criteria can exclude some of the relevant population. If the study applies a complicated definition of cases, such as the REDD sub-study on coronary heart disease, which defines cases in four registers, it should be taken into account that the definitions of cases are not necessarily compatible in different registers. In administrative registers the definitions of medical conditions are not always fully consistent with clinical criteria and some parts of the relevant population may be missed by using registers. Considerations need to be made on whether registers can be combined to capture as large a portion of the population as possible.

Although all cases can be extracted from the registers and cases from different sources can be linked, it does not necessarily follow that all cases in the target population are included. Two examples are persons with diabetes and those with coronary heart disease. A proportion of diabetic and coronary heart disease patients do not appear in any of the health care registries because they have not used hospital treatment, do not fulfil the criteria for special reimbursement right for these diseases or have not applied for reimbursements for their medication. It is important to be clear on what part of the clinical cases is not included in the registers used and whether any information is available about the size and sociodemographic or other relevant distribution of the part of the relevant population that is missed.

In studies investigating time trends, shortcomings in the coverage of registers may further be complicated by a changing proportion of cases not captured. For myocardial infarction, the introduction of troponins with good sensitivity and specificity for diagnostic testing in the late 1990s potentially changed the numbers and severity of diagnosed infarctions (Pajunen et al. 2004). The increased screening of diabetes may have resulted in a corresponding expansion of the number of incident diabetic patients and also changed the severity of cases. This may partly explain the rapid increase in the prevalence of diabetes in the 1990s (Niemi and Winell 2005). Additionally, while cross country comparative analysis may be a major strength of methods that use register-based and administrative data, detailed work is required to ensure cross-cultural compatibility.

When defining the codes used for identifying the relevant population and variables in the data set, it is of primary importance to be familiar with the definitions used in the registers and their changes in time. A considerable amount of the data recorded in the administrative registers is inexact and the researchers using these registers need to be aware of the administrative practices used in the registers and changes in them. A good contact with the experts who maintain and supervise the registers used for research is important already in the planning phase, since relevant information about the content and coverage of the registers will otherwise be missed. However, the register experts are experts in the register data, but not necessarily on the specific research questions and settings which are of interest in the study. Some characteristics of register data can only be detected when data is used for research purposes. Therefore contacts with researchers with prior experience of the data may be helpful at the planning stage.

While it is important to be familiar with the registers and the definitions used in them, good knowledge about the requirements of data protection legislation for register-based research and about the application process is also essential when planning a register-based study. Additionally, the restrictions in data release will need to be taken into account already in the planning phase, since they should be taken into account both in the planning of the data linkages and in the application process. Due to restrictions on releasing data with personal identifiers, Statistics Finland acted as the authority that carried out the linkages of the identified sub-study data sets in REDD and removed the personal identification numbers before handing the data over to the research team. Another example is avoiding indirect identification of individuals for municipalities with less than 5000 inhabitants, for which a new health centre area code needed to be developed in the REDD project to ensure meaningful small area units for research purposes of the project.

Despite the position of administrative registers in Finnish health research, the research potentials of registers are not fully utilised. In health care, the use of administrative registers in research and development is still unsystematic. There are few textbooks or scholarly articles on the underlying theory, the most appropriate methods of analysis, and a limited number of research groups. Those that do exist, however, tend to be international in their outlook, recognising the common origins of many health problems, the nature of the effective solutions, and the importance of influencing countries that are establishing or rebuilding their health systems so that they can build in the ability to undertake high-quality low-cost analysis of access to care, treatment rates and outcomes through using population registers and administrative databases. Biogenetic research has only started to realise the potential of register information. It may also be possible to nest other study types within the databases or registers. Methods for register-based research need to be refined. And finally, register authorities should acknowledge that scientific research is an essential aspect to be considered in developing and maintaining administrative registers. Legislation and institutional regulations on the use of register data as well as the content of the registers should be developed while also taking research interests into account. Administrative registers in Finland are an asset and provide a competitive edge for many fields of research. In an international context, the research community in Finland is responsible for demonstrating an example of how these registers can be used effectively for high quality research while also respecting confidentiality.

References

Act on the National Research and Development Centre for Welfare and Health (1073/1992)

Act on the Openness of Government Activities (621/1999) http://www.finlex.fi/en/laki/kaannokset/1999/en19990621.pdf (Amendments up to 1060/2002 included)

Act on the Nation-wide Registers on Health care (A 556/1989)

Act on the Statistical Action of Stakes (409/2001)

Basu J, Friedman B, Burstin H. Primary care, HMO enrollment and hospitalization for ambulatory care sensitive conditions: a new approach: Medical Care 2002; 40: 1260-1269.

Bindman AB, Grumback K, Osmond D, Komaromy M, Vranian K, Lurie J et al. Preventable hospitalizations and access to health care. Journal of American Medical Association 1995; 274: 305-311.

Cannon PJ, Connell PA, Stockley IH, Garner ST, Hampton JR. Prevalence of angina as assessed by a survey of prescriptions for nitrates. Lancet 1988; 1: 979-981.

Csonka P, Mertsola J, Klaukka T, Kaila M, Ståhlberg M-R, Ashorn P. Corticosteroid therapy and need for hospital care in wheezing preschool children. European Journal of Clinical Pharmacology 2000; 56: 591-596.

Decree on Criteria for Charges Payable to the State (211/1992)

Finnish Heart Association. Suomen sydäntutkimus- ja toimenpiderekisteri 2003. (The Finnish: Register for heart disease examinations and procedures 2003) (http://www.sydanliitto.fi/kaikki_sydamesta/tilastot_ja_rekisterit/fi_FI/rekisterit/)

Forssas E, Keskimäki I, Koskinen S, Reunanen A. Widening socioeconomic mortality disparity among diabetic people in Finland. European Journal of Public Health 2003; 13: 38-43.

Gardner K, Chapple A. Barriers to referral in patients with angina: qualitative study. BMJ. 1999 Aug 14; 319: 418-421.

Gissler M. Routinely collected registers in Finnish health research. In: Alho J, ed. Statistics, Registries, and Science – Experiences from Finland. Otava, Keuruu: Statistics Finland, 1999: 241-254.

Gissler M, Haukka J. Finnish health and social welfare registers in epidemiological research. Norsk Epidemiologi 2004; 14: 113-120 113.

Haahtela T, Klaukka T. Societal and health care benefits of early use of inhaled steroids. Thorax 1998; 53: 1005-1006.

Haahtela T, Klaukka T, Koskela K, Erhola M, Laitinenen LA. Asthma programme in Finland: a community problem needs community solutions. Thorax 2001; 56: 806-814.

Hetemaa T, Keskimäki I, Manderbacka K, Leyland AH, Koskinen S. How did the recent increase in the supply of coronary operations in Finland affect socio-economic and gender equity in their use? Journal of Epidemiology and Community Health 2003; 57: 178-185.

Hetemaa T, Keskimäki I, Salomaa V, Mähönen M, Manderbacka K, Koskinen S. Socioeconomic inequities in invasive cardiac procedures after first myocardial infarction in Finland in 1995. Journal of Clinical Epidemiology 2004; 57: 301-308.

Häkkinen U, Idänpään-Heikkilä U, Keskimäki I, Rauhala A, Klaukka T, Teitto E. Akuutin sydäninfarktin hoitokäytäntöjen, kustannusten ja vaikuttavuuden vertailu. (In Finnish: Treatment practices, cost and cost-effectiveness in the care of acute myocardial infarction). STAKES, Aiheita 9/2002.

Isometsä ET, Katila H, Aro T. Disability pension for major depression in Finland. American Journal of Psychiatry 2000; 157: 1869-1872.

Jepson GMH. How do primary care system compare across Western Europe. The Pharmaceutical Journal 2001; 267: 269-273.

Kainulainen S, Rintala T, Heikkilä M. Hyvinvoinnin alueellinen erilaistuminen 1990-luvun Suomessa. Kahtiajakautuva Suomi? -tutkimusprojektin julkaisu (In Finnish with an English abstract: Regional differentiation of well-being in Finland in the 1990s. Report on the Splitting Finland? Project). Saarijärvi: Stakes. Tutkimuksia 114, 2001.

Kalimo E, Klaukka T, Lehtonen R, Nyman K. Suomalaisten terveysturva ja sen kehitystarpeet. Terveysturvan väestötutkimuksen 1987 päätulokset. (In Finnish with an English summary: Health security in Finland and needs for development. Main results of a nationwide health security survey conducted in 1987). Kansaneläkelaitoksen julkaisuja M:81, Helsinki 1992.

Kelan tilastoryhmä. Kelan sairausvakuutustilasto 2004. (In Finnish: Social insurance statistics 2004). Official Statistics of Finland, Social security. Social Insurance Institution, Helsinki 2005.

Kennedy S, Payne N, Saul C, Luff D, McKee K, Shippam A. Reducing barriers to health care. A study to measure and understand inequalities in health resulting from an inequitable balance between health service use and need for health services. ScHARR Report Series No. 9. Sheffield: ScHARR, University of Sheffield, 2002.

Keskimäki I, Aro S. Accuracy of data on diagnosis, procedures and accidents in the Finnish Hospital Discharge Register. International Journal of Health Sciences 1991; 2: 15-21.

Keskimäki I, Salinto M, Aro S. Private medicine and socioeconomic differences in the rates of common surgical procedures in Finland. Health Policy 1996; 36: 245-259.

Keskimäki I, Salinto M, Koskinen S. Viranomaisrekisterit terveyden ja terveyspalvelujen käytön tutkimusaineistona (In Finnish: Administrative registers as research data for health services use). Suomen Lääkärilehti 1997; 52: 585-591.

Keskimäki I, Koskinen S, Salinto M, Aro S. Socioeconomic and gender inequities in access to coronary artery bypass grafting in Finland. European Journal of Public Health 1997; 7: 392-397.

Keskimäki I, Karvonen S, Sund R, Leyland AH. Monitasomallien käyttö terveystutkimuksessa. (In Finnish with an English abstract: Multilevel models in health research). Sosiaalilääketieteellinen Aikakauslehti 2001; 38: 327-335.

Keskimäki I. How did Finland's economic recession in the early 1990s affect socio-economic equity in the use of hospital care? Social Science and Medicine 2003; 56: 1517-1530.

Kieseppä T, Partonen T, Kaprio J, Lönnqvist J. Accuracy of register-and record-based bipolar I disorder diagnoses in Finland: a study of twins. Acta

Neuropsychiatrica 2000; 12: 106-109.
Klaukka T. Tietoja erityiskorvaukseen oikeuttavien kansantautien potilasmääristä vuonna 2000. (In Finnish: Information about numbers of patients having diseases entitling to special reimbursement of medicine costs in year 2000). Suomen Lääkärilehti 2001; 56: 902-904.
Koivusalo M. Decentralisation and equity of healthcare provision in Finland. BMJ 1999; 318: 1198- 1200.
Kokko S. General Practice in the Nordic Countries. 13th Nordic Conference in General Practice. Helsinki, 31.8. - 3.9.2003.
Kortteinen M, Vaattovaara M. Onko osa Helsingistä alikehityksen kierteessä? (In Finnish: Is a Part of Helsinki in a Vicious Circle of Underdevelopment?) Yhteiskuntapolitiikka 2000; 65: 115-124.
Lahti RA. From findings to statistics: An assessment of Finnish medical cause-of-death information in relation to underlying-cause coding. Department of Forensic Medicine, University of Helsinki, Helsinki 2005.
Laki sairausvakuutuslain muuttamisesta ja väliaikaisesta muuttamisesta n:o 885/2005.
Laki sosiaali- ja terveysalan tutkimus- ja kehittämiskeskuksen tilastotoimesta 23.5.2001 /409. http://www.finlex.fi/fi/laki/ajantasa /2001/20010409
Lewsey JD, Leyland AH, Murray GD, Boddy FA. Using routine data to complement and enhance the results of randomised controlled trials. Health Technology Assessment 2000; 4: 22.
Leyland A, Goldstein, H. Multilevel Modelling of Health Statistics, Chichester: John Wiley & Sons, 2001.
Leyland AH, Groenewegen PP. Multilevel modelling and public health policy. Scandinavian Journal of Public Health 2003; 31: 267-274.
Lindeman S, Hämäläinen J, Isometsä E, Kaprio J, Poikolainen, Heikkinen M, Aro H. The 12 month prevalence and major risk factors for major depressive episode in Finland: representative sample of 5993 adults. Acta Psychiatrica Scandinavica 2000; 102: 178-184.
Linna M, Häkkinen U. Sairaaloiden tuottavuus 1991–1994. (In Finnish: Productivity of hospitals 1991–1994). In Häkkinen U, Asikainen K, Linna M, Eds. Terveyspalvelujen tarve ja kustannukset alueittain sekä sairaaloiden tuottavuus 1990-luvulla. STAKES, National Research and Development Centre for Welfare and Health, Helsinki 1996, pp. 238-53.
Malinen P, Keränen R, Keränen H. Rural area typology in Finland - a tool for rural policy. Research Reports 123. University of Oulu, Research Institute of Finland, 1994.
Miettinen OS. Theoretical epidemiology. Principles of occurrence research in medicine. New York: John Wiley & Sons 1985.
Mähönen M, Salomaa V, Brommels M, et al. The validity of hospital discharge register data on coronary heart disease in Finland. European Journal of Epidemiology 1997; 13: 403-415.
Mäkikyrö T, Isohanni M, Moring J, Hakko H, Hovatta I, Lönnqvist J. Accuracy of register-based schizophrenia -diagnoses in a genetic study. European Psychiatry 1998; 13: 57-62.
Naderi A, Mace J. Education and earnings: a multilevel analysis. A case study of the manufacturing sector in Iran. Economics of Education Review 2003; 22: 143-156.
National Agency for Medicines and Social Insurance Institution. Finnish statistics on medicines 2004. Helsinki, 2005.
Nevalainen J. The 2002-2003 implant yearbook on orthopaedic endoprostheses. Finnish Arthroplasty Register. http://www.nam.fi/uploads/julkaisut/Orthopaedic_Endoprostheses_2003_v.pdf
Niemi M, Winell K. Diabetes Suomessa. Esiintyvyys ja hoidon laadun vaihtelu. (In Finnish: Diabetes in Finland. Prevalence and variations in the quality of care). Stakes, Reports 8/2005 http://www.stakes.fi/verkkojulk/pdf/rap05/Ra8-2005.pdf
Nolte E, McKee M. Does health care save lives? Avoidable mortality revisited. 2004. www.nuffieldtrust.org.uk/ecomm/files/21404avoidablemortality.pdf
Office for National Statistics. General practitioners, dentists and opticians by NHS Regional Office area, 30 September 2001: Regional Trends 37. http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=5945
Oster A, Bindman AB. Emergency department visits for ambulatory care sensitive conditions: insights into preventable hospitalizations. Medical Care 2003; 41: 198-207.
Pajunen P, Koukkunen H, Ketonen M, et al. The validity of the Finnish Hospital Discharge Register and Causes of Death Register data on coronary heart disease. European Journal of Cardiovascular Prevention and Rehabilitation2005; 12: 132-137.
Pajunen P, Pääkkönen R, Juolevi A, Hämäläinen H, Keskimäki I, Laatikainen T, Moltchanov V, Niemi M, Rintanen H, Salomaa V. Trends in fatal and non-fatal coronary heart disease events in Finland during 1991-2001. Scandinavian Cardiovascular Journal 2004;38(6):340 - 344.
Personal Data Act 1999. Henkilötietolaki (523/1999).
Personal Data Act (523/1999)
Poikolainen K and Eskola J. The effect of health services on mortality: decline in death rates from amenable and non-amenable causes in Finland, 1969-81. Lancet 1986: i; 199-202.
Poikolainen K and Eskola J. Regional and social class variation in the relative risk of death from amenable causes in the city of Helsinki, 1980-1986. International Journal of Epidemiology 1995: 24; 114-118.
Purola T, Kalimo E, Sievers K, Nyman K. Sairastavuus ja lääkintäpalvelusten käyttö Suomessa ennen sairausvakuutusta (In Finnish: Morbidity and the utilisation of medical services in Finland before the sickness insurance). Kansaneläkelaitoksen julkaisuja A:1, Helsinki 1967.
Purola T, Nyman K, Kalimo E, Sievers K. Sairausvakuutus, sairastavuus ja lääkintäpalvelusten käyttö (In Finnish: Sickness insurance, morbidity and the utilisation of health services). Kansaneläkelaitoksen julkaisuja A:7, Helsinki 1971.
Puumalainen I, Helin-Salmivaara A. Masennuksen lääkehoito, kenelle, mitä ja millä hinnalla? (In Finnish: Drug treatment for depression, who receives treatment, with what are they treated and at what cost?). Suomen Lääkärilehti 2003; 58: 524-528.
Rapola JM, Virtamo J, Korhonen P, et al. Validity of diagnoses of major coronary events in national registers of hospital diagnoses and deaths. European

Journal of Epidemiology 1997; 13: 133-138.

Rice N, Leyland A. Multilevel models: applications to health data. Journal of Health Services Research and Policy 1996; 1: 154-164.

Riihinen O. Teollistuvan yhteiskunnan alueellinen erilaistuneisuus (In Finnish: Regional differentiation in an industrialising society). Kuopio: WSOY 1965.

Rosén M. National Health Data Registers: a Nordic heritage to public health. Scand J Public Health 2002; 30: 81-85.

Salmela R, Koistinen V. Yleissairaaloiden poistoilmoitusrekisterin kattavuus ja luotettavuus (In Finnish: Is the discharge register of general hospitals complete and reliable?). Sairaala 1987; 49: 480-482.

Salminen J, Saarijärvi S, Raitasalo R. Depression and disability pension in Finland. Acta Psychiatrica Sacndinavica 1997; 95: 242-243.

Salomaa V, Miettinen H, Niemelä M, Ketonen M, Mähönen M, Immonen-Räihä P, et al. Relation of socioeconomic position to the case fatality, prognosis and treatment of myocardial infarction events; the FINMONICA MI Register Study. Journal of Epidemiology and Community Health 2001; 55: 475-482.

Siirilä, S. Alueellisen erilaistuneisuuden vanhat ja uudet piirteet. (In Finnish: The old and new characteristics of regional differentation). In O. Riihinen (Ed.), Suomi 2017. Jyväskylä: Gummerus 1990.

Social Insurance Institution. Terveyspalvelujen kustannukset ja rahoitus Suomessa 1960-1994 (In Finnish: Costs and financing of health services in Finland 1960-1994) Publications of the Social Insurance Institution T9:52. Social Insurance Institution, Helsinki 1996.

Social Insurance Institution. Statistical Yearbook of the Social Insurance Institution. Social Insurance Institution Publications T1:36, Helsinki 2001.

Sorvaniemi M, Helenius H, Alho A, Kesti S, Salokangas RKR. Improved diagnostic assessment of major depression in psychiatric outpatient care in Finland. Psychiatric Services 1998; 49: 384-386.

STAKES 2006. http://www.stakes.info/cognos73/cgi-bin/ppdscgi.exe?DC=Q&E=/Hilmo/teemanetti

Statistics Act 2004. Tilastolaki (280/2004). (http://www.stat.fi/meta/lait/tilastolaki.html)

Statistics Finland. Use of Register and Administrative Data Sources for Statistical Purposes. Best Practices of Statistics Finland. Handbooks 45. Statistics Finland, Helsinki 2004. http://www.stat.fi/tup/julkaisut/kasikirjoja_45_en.pdf

Statistics Finland 2006. The quality report of the causes of death register http://www.stat.fi/til/ksyyt/ksyyt_2004-12-20_laa_001.html

Taipale V. Mielekäs elämä (In Finnish: Meaningful life) Working Group Reports 1996: 20. Ministry of Social Affairs and Health, Helsinki 1996.

Tod AM, Lacey EA, McNeill F. 'I'm still waiting...': barriers to accessing cardiac rehabilitation services. Journal of Advanced Nursing 2002; 40: 421.

Tod AM, Read C, Lacey A, Abbott J. Barriers to uptake of services for coronary heart disease: qualitative study. BMJ. 2001; 323:214 (Accessed from www.bmj.com on 15/9/03).

Vaattovaara, M, Kortteinen, M. Beyond Polarisation versus Professionalisation? A Case Study of the Development of the Helsinki region, Finland. Urban Studies 2003; 40:2127-2145

Valkonen T, Koskinen S, Martelin T, eds. Rekisteritiedot yhteiskunta- ja terveystutkimuksen aineistona. (In Finnish: Register data in social and health research). Helsinki: Gaudeamus 1997.

Valkonen T, Martelin T, Rimpelä A, Notkola V, Savela S. Socio- economic mortality differences in Finland 1981-90, Statistics Finland, Population 1993, 1, Helsinki 1993

Valtioneuvoston asetus N:o 1108/2005 lääketieteellisin perustein vaikeiksi arvioitavista sairauksista, joiden lääkehoidon kustannuksista sairausvakuutuslain 5 luvun 6§:n 2 momentin perusteella korvataan 72 tai 100 prosenttia.

Valtioneuvoston päätös N:o 614/1986 vaikeista ja pitkäaikaisista sairauksista ja niiden hoitoon käytettävistä sairausvakuutuslain mukaan kokonaan tai 90-prosenttisesti korvattavista lääkkeistä annetun valtioneuvoston päätöksen 1 ja 2 §:n muuttamisesta.

Appendices

Appendix 1. The research group

Person in charge of the project

Ilmo Keskimäki [1], MD, DmedSc (PhD), Docent
Research Director, Head of the Outcomes and Equity Research Group

Other members of the research group

Sakari Karvonen [2], MSc, PhD (sociology), Docent, Head of the Welfare Research Group
Matti Kortteinen [2], MSc, PhD (sociology), Docent, Senior Research Fellow
Alastair Leyland [3], MSc, PhD (statistician), Head of Programme
Erja Forssas [1], MSc (health care), Research Fellow
Martti Arffman [1], MSc, (statistician, systems analysis)
Mira Kajantie [1], MSc student (statistician, -31.12.2005)
Sonja Lumme [1], MSc, Doctoral Student (statistician)
Kristiina Manderbacka, MSc, PhD (sociology), Senior Researcher
Alison McCallum, [1,4,5] MBChB (MD), MSc (public health medicine) FFPHM, Research Fellow

Affiliations

[1] National Research and Development Centre for Welfare and Health (STAKES),
Outcomes and Equity Research
[2] National Research and Development Centre for Welfare and Health (STAKES),
Welfare Research
[3] MRC Social and Public Health Sciences Unit, University of Glasgow,
Measuring Health and Variations in Health Programme
[4] NHS Lothian Health Board, Director of Public Health, Edinburgh
[5] Department of Primary Care and Population Sciences, Royal Free and University College
London Medical School

Appendix 2. Codes and variables used for defining populations for the REDD-project.

Note: Codes are for the primary data collection from the original data sources and in some cases definitions are more wide-ranging than the ones used to define final study populations. Definitions may also vary between sub-studies because of the differences in contents and objectives of the studies.

1 Elective surgery

Source:	HILMO -database (incl. demanding heart patient's extra sheet (DHES))
Variables	All procedure variables - NOMESCO: PTMPK1-3, MTMP1K1, MTMP2K1 - Hospital League: TMP1-3 - DHES: TMP94, TMPTYP1-3
Years:	1991–2003
Codes:	

OPERATION	NOMESCO	HOSPITAL LEAGUE	DHES
CABG	FNA-G		2, 11, 25
PTCA	FN1AT, FN1BT, FN1YT		1, 82-84
angio	FN1AC, FN1BC, FN1CC,		61, 81
catheterisation	TFC00,TFC10, TFC20, TFC99	5311-5319	
other cardiac procedure	FNH-W, FN1ST, FN1XT, TFN, XFN, FN1BE, FN1BG, TFB10, TFB15, TFC25, TFC30, TFC40, TFC10, TFE, TFF, TFJ, TFP, TFW		
hip endoprosthesis	NFB, NFC, NFG, NFJ	9291-9294, 9321-9329, (9132 if ICD9=820 or ICD10=S72,0 or S72.1)	
knee endoprosthesis	NGB, NGC, NGG	9313, 9314, 9331-9339, 9281	
spinal disc operations	NAB, NAC, NAG, ABC	9171-79, 9181-89, 9191-9199, 9201-03, 9209, 9211-12, 9219, 9221-22, 9229	
cataract operation	CJA, CJC, CJD, CJE, CJF	3621-33, 3639	
hysterectomy	LCB, LCC, LCD, LEF13	8301-02, 8305-11, 8319, 8436, 7435	
prostatectomy	KEC, KED	7451-59	

2 Ambulatory care sensitive conditions

Source: HILMO -database
Variables: All main and side diagnoses
- ICD10: PDGO, PDGE, SDG1O, SDG1E, SDG2E, SDG2O
- ICD9: PDG, SDG1, SDG2, SDG3
Years: 1991–2003
Codes:

Condition	ICD10	ICD9
Asthma	J45, J46	493
COPD	J40, J41, J42, J43, J44	490, 491, 492 496
Hypertension	I10, I11, I12, I13	401-404
Angina pectoris	I20, I24 (excl. I24.1)	411, 413
Heart failure	I50, I51	428, 398.0A

3 Helsinki city data

Population: Base population consisted of every person living in Helsinki city district 31st December 1998–2003. Study data for the REDD -project included every person having one or more entries in health care registers and those who died as residents during those years.

Sources:

Authority	Register	Information
Helsinki City	Population register system (VTJ)	IDs for persons living in Helsinki 31.12. 1998-2003
STAKES	HILMO	All admissions 1991–2003
STAKES	Bencmarking database	Outpatient department visits 1998–2003
Social Insurance Institution	Register on entitlements for special reimbursement	All effective special reimbursement rights 1998–2003
Social Insurance Institution	Register on reimbursements for private health care	Reimbursed doctors appointments, dental care, physical therapy and examinations 1998–2003
Social Insurance Institution	Pension register	Disability pensions 1998–2003
Finnish Centre for Pensions	Pension register	Disability pensions 1998–2003
Helsinki City	Health care register	Use of public health care services 1998–2003 (excluding occupational health care)
Statistics Finland	Causes of death register	All deaths 1998–2003

4 Avoidable deaths

Source: Causes of death register
Variables: Basic causes of death
- Selected cause of death (VPKS)
- Statistical cause of death (TPKS)
Years: 1991–2003
Codes:

Cause of death	Age	ICD-8	ICD-9	ICD-10
Intestinal infections	1-14	000-009	001-009	A00-A09
Other infections (Diphtheria, Tetanus, Poliomyelitis)	1-74	032, 037, 040-043	032, 037, 45	A36, A35, A80
Whooping cough	1-14	033	033	A37
Measles	1-14	055	055	B05
Tuberculosis	1-74	010-019	010-018, 137	A15-A19, B90
Malignant neoplasm of colon and rectum(‡)	1-74	153-154	153-154	C18-C21
Malignant neoplasm of skin(‡)	1-74	173	173	C44
Malignant neoplasm of breast	1-74	174	174	C50
Malignant neoplasm of cervix uteri	1-74	180	180	C53
Malignant neoplasm of cervix uteri and body of uterus	1-44	182	179, 182	C54, C55
Hypertensive disease	1-74	400-404	401-405	I10-I13, I115
Cerebrovascular disease	1-74	430-438	430-438	I60-I69
Diseases of the thyroid	1-74	240-246	240-246	E00-E07
Diabetes mellitus	1-49	250	250	E10-E14
Epilepsy	1-74	345	345	G40-G41
All respiratory diseases (excl. pneumonia/ influenza)	1-14	460-466, 490-519	460-479, 488-519	J00-J09, J20-J99
Asthma (*)	15-49		493	J45, J46
COPD (*)	15-49		490-492, 496	J40-J44
Septicaemia	1-74	038	038	A40-A41
Malignant neoplasm of testis,(‡)	1-74	186	186	C62
Hodgkin's disease(‡)	1-74	201	201	C81
Leukaemia	1-44	204-207	204-208	C91-C95
Chronic rheumatic heart disease (‡)	1-74	393-396	393-398	I05-I09
Influenza	1-74	470-474	487	J10-J11
Pneumonia	1-74	480-486	480-486	J12-J18
Peptic ulcer	1-74	531-533	531-533	K25-K27
Appendicitis	1-74	540-543	540-543	K35-K38
Abdominal hernia	1-74	550-553	550-553	K40-K46
Cholelithiasis & cholecystitis	1-74	574-575	574-575.1	K80-K81
Nephritis and nephrosis	1-74	580-584	580-589	N00-N07, N17-N19, N25-N27
Benign hyperplasia of prostate	1-74	600	600	N40
Maternal death	All	630-678	630-676	O00-O99
Congenital cardiovascular anomalies	1-74	746-747	745-747	Q20-Q28
Ischaemic heart disease	1-74	410-414	410-414	I20-I25
Perinatal deaths, all causes excluding stillbirths	All	760-779	760-779	P00-P96, A33,A34

(‡) condition not included as 'amenable condition' in Denmark, Finland and Sweden in Nolte & McKee's study due to the classification of the WHO data based on ICD8 in these countries. Included in the REDD study

5 Psychiatric hospital care

Source:	HILMO -database (incl. yearly patient counts)
Variables:	All main and side diagnoses - ICD10: PDGO, PDGE, SDG1O, SDG1E, SDG2E, SDG2O - ICD9: PDG, SDG1, SDG2, SDG3
Years:	1994–2003
Codes:	

	ICD10	ICD9
All psychiatric conditions	F00-F99	290-319

6 Use of antidepressants

Source:	Register on prescription reimbursements
Variables:	Medicine code (ATC)
Years:	1994–2003
Code:	N06A

7 Incident coronary heart disease

Population:	Persons over 40 year's old and received CHD diagnosis.
Years:	1994–2003
Source 1:	HILMO -database
Variables:	All main and side diagnoses - ICD10: PDGO, PDGE, SDG1O, SDG1E, SDG2E, SDG2O - ICD9: PDG, SDG1, SDG2, SDG3 All procedure variables - NOMESCO: PTMPK1-3, MTMP1K1, MTMP2K1 - Hospital League: TMP1-3 - DHES: TMP94, TMPTYP1-3
Codes:	

CONDITION	ICD10	ICD9
CHD	I20, I21, I22, I23, I24, I25, I46.1, I46.9, R96, R98	410-414, 798, (excl. 7980A)

OPERATION	NOMESCO	HOSPITAL LEAGUE	DHES
Invasive heart operation	FNA-G, XFN, FN1AT, FN1BT, FN1YT, FN1ST, FN1XT, TFC00, TFC10, TFN10, FN1AC, FN1BC, FN1CC,TFN04, TFN40, TFN50	5311-15, 5341, 5342, 5349	11, 25, 61, 81, 82, 83, 84 and 1, 2 (1994)

Source 2: Social Insurance Institution
Codes:

REGISTER	VARIABLE	CODE
Special reimbursement r.	Disease code (SVA3)	206
Prescription register	Medicine code (ATC)	C01D

Source 3: Causes of death register
Variables: Basic causes of death
- Selected cause of death (VPKS)
- Statistical cause of death (TPKS)

Years: 1994–2003
Codes:

CONDITION	ICD10	ICD9
CHD	I20, I21, I22, I23, I24, I25, I46.1, I46.9, R96, R98	410-414, 798, (excl. 7980A)

Appendix 3. REDD data from Statistics Finland

1	TK_TUTNRO	Identification number (Statistics Finland)	
2	SP	Sex	1=men, 2=women
3-15	SIVS90-SIVS02	Marital status 1990–2002	1=single 2=married 3=separeted 4=widowed 5=divorced
16-18	KANSA00-KANSA02	Nationality 2000–2002	1=Finnish, 2=other
19-31	AKUN90-AKUN02	Municipality of residence (2004 division)	Only over 5000 resident municipalities
32-44	PTOIM190-PTOIMI102	Main activity (TVM) 1995	0=employed 1=unemployed 2=aged 0-14 years 3=student 5=conscript 6=retired (not due to unemployment) 7=retired due to unemployment 8=other, unknown
45-57	AMAS190-AMAS102	Occupational status 1990–2002	1=employee 2=entrepreneur 9=other, unknown, out of labour force
58-70	TOL**	Line of business 1990–2002	1988, 1995 and 2002 classifications
71-83	TYO90-TYÖ02	Months in labour force 1990–2002 (employed or unemployed)	00-12, 99=none
84-96	TYKE90-TYKE02	Months unemployed 1990–2002	00-12, 99=none
97-109	YOTUTK90-YOTUTK02	Secondary school graduate 1990–2002	0=no, 1=yes
110-121	SVATV90-SVATV02	Personal income subject to state tax 1990–2002	>50 000 € = 50 000 €
122-132	SVEROT91-SVEROT02	Personal taxes and payments 1990–2002	>50 000 € = 50 000 €
133-145	PEKO90-PEKO02	Number of family members (living in same household)	2-25, 00=not living with family
146-158	PEAS90-PEAS02	Family position 1990–2002	0=not a member 1=head of family 2=spouse 3=child 4=head of common-law marriage 5=spouse of common-law marriage 9=unknown
159-171	PETY90-PETY02	Family type 1990–2002	0=not a member of a family 1=married couple and no children 2=married couple and children 3=mother and children 4=father and children 5=cohabitation and children

Appendix 3...

			6=cohabitation and children from previous relationships 7=cohabitation and no children 9=unknown
172-184	LAPS1890-LAPS1802	Number of children under 18 years in the family 1990–2002	0–9
185-197	LAPS790-LAPS702	Number of children under 7 years in the family 1990–2002	0–9
198-210	AKOKO90-AKOKO02	Number of people living in the household 1990–2002	1–25
211-223	HAPE90-HAPE02	Housing tenure 1990–2002	Note! Classification changed 1998
224-236	HULU90-HULU02	Number of rooms (excl. kitchen) 1990–2002	
237	SOSE95	Socioeconomic status 1995 (Statistics Finland 1980)	
239-240	SOSE8995, SOSE8900	Socioeconomic status 1995 and 2000 (Statistics Finland 1989)	
240-252	KTUTKU90-KTUTKU02	Educational degree	1=intermediate grade 2=vocational 3=lower academic degree 4=higher academic degree 5=post graduate 6=unknown/basic education
	PETULO90-PETULO02	Family income subject to state tax 1990–2002	> 100 000 € = 100 000 €
	PEKYL90-PEKYL02	Number of consumption units in the family 1990–2002	
	PEVERO91-PEVERO02	Family taxes and payments 1990–2002	> 100 000 € = 100 000 €
	PENETTO91-PENETTO02	Family net income 1990–2002	> 100 000 € = 100 000 €
303-304	PEKAYTTO95, PEKAYTTO00	Disposable family income 1995 and 2000	> 50 000 € = 50 000 € (singles), 100 000 € = 100 000 € (families)
305-317	PERAUTO90-PERAUTO02	Possibility to use a car	0=no, 1=yes
318-319	JSES90, JSES95	Derived socioeconomic status 1990 and 1995	1=upper white-collar emploee 2=lower white-collar, supervising or independent work 3=lower white-collar, unindependent work 4=skilled worker 5=unskilled worker 6=worker, no data on level of skills 7=farmer 8=other entrepreneur 9=student 10=other 11=lower white-collar employee, data from the 1975 census
320-322	UISCED90, UISCED95, UISCED00	Degree of education (ISCED)	

Appendix 3...

Cause of death			
326	KUNTA	Place of recidence at the time of death	1=Finland 2=foreign 3=unknown
327	KUOLPV	Date of death	
328	PERUSTE	How cause of death was determined	
329	KPAIKKA	Place of death	Change in classification 1998
330	TPKS	Statistical cause of death	ICD10
331	TPKSAIKA	Statistical cause of death in longitudinal cause of death data	ICD10
332	HUUME	Alcohol related cause of death	1=main cause, 2=contributing cause
333	VPKS	Doctor defined cause of death	ICD10
334	VPKSR1	Parallel code for VPKS	ICD10
335	VKS	Immediate cause of death	ICD10
336	VVKS1	Cause for first intermediate stage	ICD10
337-339	M1-M3	Contributing causes of death 1–3	ICD10
340	VKSR1	Parallel code for VKS	ICD10
341	VVKS1R1	Parallel code for VVKS1	ICD10
Area			
	KUNTAMUOTO**	Type of municipality (city/other)	
	KUNTARYHMA**	Type of municipality federation	
	TAAJAMA_ASTE**	Proportion of the population living in densely populated areas in the municipality	
	TEKE_ALUE90-TEKE_ALUE02	Health centre area (STAKES)	< 5000 residents
	SAIRAANHOITOP_90_04-SAIRAANHOITOP_02_01	Hospital district 1990–2002	
383-395	IKAKK90-IKAKK02	Age in months 31.12 each year 1990–2002	If missing information, then not resident at that time
396	KIELI	Mother tongue	1=Finnish, 2=Swedish, 3=other
397-408	SNETTO91-SNETTO02	Personal net income 1991–2002	> 50 000 € = 50 000 €
	VATA90, VATA95, VATA00	Residential standard of equipment	1=well-equipped 2=poorly equipped 3=very poorly equipped 9=unknown
411-423	ALUE90-ALUE02	Small area code for large citys	

Appendix 4 Health centre codes in the REDD project

Municipality code	Name of municipality	Hospital district	Health centre area	New area code
220	Karjaa	01	607110	0160711000
044	Elimäki	08	901190	0890119000
045	Eno	12	901200	1290120000
049	Espoo	01	901240	0190124000
050	Eura	04	901250	0490125000
077	Hankasalmi	14	901400	1490140000
078	Hanko	01	901410	0190141000
084	Haukipudas	18	901460	1890146000
091	Helsinki	01	901520	0190152000
092	Vantaa	01	901530	0190153000
106	Hyvinkää	01	901620	0190162000
142	Iitti	07	901670	0790167000
143	Ikaalinen	06	901680	0690168000
145	Ilmajoki	15	901600	1590160000
146	Ilomantsi	12	901700	1290170000
149	Inkoo	01	901720	0190172000
164	Jalasjärvi	15	901780	1590178000
165	Janakkala	05	901790	0590179000
171	Joroinen	11	901840	1190184000
173	Joutseno	09	901800	0990180000
176	Juuka	12	901890	1290189000
179	Jyväskylä	14	901920	1490192000
186	Järvenpää	01	901980	0190198000
208	Kalajoki	18	902050	1890205000
213	Kangasniemi	10	902100	1090210000
217	Kannus	17	902130	1790213000
224	Karkkila	01	902190	0190219000
232	Kauhajoki	15	902260	1590226000
235	Kauniainen	01	902280	0190228000
245	Kerava	01	902370	0190237000
255	Kiiminki	18	902470	1890247000
263	Kiuruvesi	13	902540	1390254000
271	Kokemäki	04	902580	0490258000
276	Kontiolahti	12	902630	1290263000
285	Kotka	08	902700	0890270000
288	Kruunupyy	17	902730	1790273000
290	Kuhmo	19	902750	1990275000
309	Outokumpu	12	902890	1290289000
398	Lahti	07	902990	0790299000
408	Lapua	15	903080	1590308000
410	Laukaa	14	903100	1490310000

Appendix 4...

420	Leppävirta	13	903180	1390318000
426	Liperi	12	903240	1290324000
441	Luumäki	09	903380	0990338000
494	Muhos	18	903570	1890357000
499	Mustasaari	16	903610	1690361000
505	Mäntsälä	01	903660	0190366000
532	Nastola	07	903710	0790371000
535	Nivala	18	903740	1890374000
536	Nokia	06	903750	0690375000
540	Nummi-Pusula	01	903790	0190379000
543	Nurmijärvi	01	903820	0190382000
564	Oulu	18	903900	1890390000
604	Pirkkala	06	904190	0690419000
606	Pohja	01	904210	0190421000
607	Polvijärvi	12	904220	1290422000
638	Porvoo	01	904260	0190426000
615	Pudasjärvi	18	904290	1890429000
624	Pyhtää	08	904380	0890438000
626	Pyhäjärvi	18	904400	1890440000
698	Rovaniemi	21	904680	2190468000
700	Ruokolahti	09	904700	0990470000
702	Ruovesi	06	904720	0690472000
754	Anjalankoski	08	904980	0890498000
758	Sodankylä	21	905010	2190501000
762	Sonkajärvi	13	905050	1390505000
765	Sotkamo	19	905060	1990506000
837	Tampere	06	905320	0690532000
851	Tornio	20	905430	2090543000
853	Turku	03	905450	0390545000
858	Tuusula	01	905500	0190550000
859	Tyrnävä	18	905510	1890551000
864	Toijala	06	905540	0690554000
905	Vaasa	16	905650	1690565000
908	Valkeakoski	06	905670	0690567000
927	Vihti	01	905840	0190584000
936	Virrat	06	905910	0690591000
980	Ylöjärvi	06	906070	0690607000
422	Lieksa	12	906120	1290612000
111	Heinola	07	906150	0790615000
761	Somero	03	906310	0390631000
893	Uusikaarlepyy	16	906590	1690659000
895	Uusikaupunki	03	906860	0390686000
835	Tammisaari	01	906920	0190692000
004	Alahärmä	15	991000	1599100000
052	Evijärvi	15	991000	1599100000

Appendix 4...

971	Ylihärmä	15	991000	1599100000
016	Asikkala	07	991090	0799109000
576	Padasjoki	07	991090	0799109000
071	Haapavesi	18	991340	1899134000
682	Rantsila	18	991340	1899134000
108	Hämeenkyrö	06	991630	0699163000
932	Viljakkala	06	991630	0699163000
139	Ii	18	991650	1899165000
972	Yli-Ii	18	991650	1899165000
148	Inari	21	991710	2199171000
890	Utsjoki	21	991710	2199171000
153	Imatra	09	991760	0999176000
689	Rautjärvi	09	991760	0999176000
167	Joensuu	12	991810	1299181000
856	Tuupovaara	12	991810	1299181000
244	Kempele	18	992360	1899236000
297	Kuopio	13	992800	1399280000
163	Jaala	08	992870	0899287000
306	Kuusankoski	08	992870	0899287000
402	Lapinlahti	13	993030	1399303000
916	Varpaisjärvi	13	993030	1399303000
418	Lempäälä	06	993160	0699316000
922	Vesilahti	06	993160	0699316000
425	Liminka	18	993230	1899323000
436	Lumijoki	18	993230	1899323000
223	Karjalohja	01	993250	0199325000
444	Lohja	01	993250	0199325000
737	Sammatti	01	993250	0199325000
254	Kiikoinen	06	993560	0699356000
413	Lavia	06	993560	0699356000
493	Mouhijärvi	06	993560	0699356000
772	Suodenniemi	06	993560	0699356000
534	Nilsiä	13	993730	1399373000
687	Rautavaara	13	993730	1399373000
072	Hailuoto	18	993910	1899391000
567	Oulunsalo	18	993910	1899391000
101	Houtskari	03	993920	0399392000
150	Iniö	03	993920	0399392000
279	Korppoo	03	993920	0399392000
533	Nauvo	03	993920	0399392000
573	Parainen	03	993920	0399392000
442	Luvia	04	994240	0499424008
609	Pori	04	994240	0499424000
680	Raisio	03	994511	0399451100
704	Rusko	03	994511	0399451100

Appendix 4...

920	Velkua	03	994511	0399451100
683	Ranua	21	994691	2199469100
699	Rovaniemi rural municipality	21	994691	2199469100
739	Savitaipale	10	994851	1099485100
775	Suomenniemi	10	994851	1099485100
081	Hartola	07	995191	0799519100
781	Sysmä	07	995191	0799519100
175	Jurva	15	995381	1599538100
846	Teuva	15	995381	1599538100
848	Tohmajärvi	12	995401	1299540100
943	Värtsilä	12	995401	1299540100
095	Himanka	17	995411	1799541100
421	Lestijärvi	17	995411	1799541100
849	Toholampi	17	995411	1799541100
889	Utajärvi	18	996001	1899600100
973	Ylikiiminki	18	996001	1899600100
854	Pello	21	996031	2199603100
976	Ylitornio	21	996031	2199603100
541	Nurmes	12	996171	1299617100
911	Valtimo	12	996171	1299617100
046	Enonkoski	11	996201	1199620100
246	Kerimäki	11	996201	1199620100
741	Savonranta	11	996201	1199620100
507	Mäntyharju	10	996221	1099622100
588	Pertunmaa	10	996221	1099622100
586	Perniö	03	996231	0399623100
784	Särkisalo	03	996231	0399623100
580	Parikkala	11	996281	1199628100
728	Saari	11	996281	1199628100
891	Uukuniemi	11	996281	1199628100
476	Maaninka	13	996321	1399632100
749	Siilinjärvi	13	996321	1399632100
401	Lammi	05	996361	0599636100
855	Tuulos	05	996361	0599636100
414	Lehtimäki	15	996401	1599640100
759	Soini	15	996401	1599640100
989	Ähtäri	15	996401	1599640100
239	Keitele	13	996421	1399642100
595	Pielavesi	13	996421	1399642100
577	Paimio	03	996441	0399644100
738	Sauvo	03	996441	0399644100
483	Merijärvi	18	996471	1899647100
563	Oulainen	18	996471	1899647100
926	Vihanti	18	996471	1899647100

Appendix 4...

485	Merimasku	03	996481	0399648100
529	Naantali	03	996481	0399648100
705	Rymättylä	03	996481	0399648100
249	Keuruu	14	996511	1499651100
495	Multia	14	996511	1499651100
400	Laitila	03	996561	0399656100
631	Pyhäranta	03	996561	0399656100
490	Mietoinen	03	996581	0399658100
503	Mynämäki	03	996581	0399658100
251	Kiihtelysvaara	12	996621	1299662100
632	Pyhäselkä	12	996621	1299662100
319	Köyliö	04	996631	0499663100
783	Säkylä	04	996631	0499663100
172	Joutsa	14	996721	1499672100
415	Leivonmäki	14	996721	1499672100
435	Luhanka	14	996721	1499672100
280	Korsnäs	16	996751	1699675100
475	Maalahti	16	996751	1699675100
912	Vammala	06	996811	0699681100
988	Äetsä	06	996811	0699681100
140	Iisalmi	13	996821	1399682100
925	Vieremä	13	996821	1399682100
233	Kauhava	15	996911	1599691100
281	Kortesjärvi	15	996911	1599691100
177	Juupajoki	06	996941	0699694100
443	Längelmäki	06	996941	0699694100
562	Orivesi	06	996941	0699694100
231	Kaskinen	16	996991	1699699100
545	Närpiö	16	996991	1699699100
257	Kirkkonummi	01	997021	0199702100
755	Siuntio	01	997021	0199702100
292	Kuivaniemi	18	997111	1899711100
751	Simo	18	997111	1899711100
151	Isojoki	16	997121	1699712100
218	Karijoki	16	997121	1699712100
287	Kristiinankaupunki	16	997121	1699712100
293	Kullaa	04	997161	0499716100
886	Ulvila	04	997161	0499716100
040	Dragsfjärd	03	997181	0399718100
243	Kemiö	03	997181	0399718100
923	Västanfjärd	03	997181	0399718100
247	Kestilä	18	997231	1899723100
603	Piippola	18	997231	1899723100
617	Pulkkila	18	997231	1899723100
630	Pyhäntä	18	997231	1899723100

Appendix 4...

479	Maksamaa	16	997251	1699725100
559	Oravainen	16	997251	1699725100
944	Vöyri	16	997251	1699725100
304	Kustavi	03	997281	0399728100
833	Taivassalo	03	997281	0399728100
918	Vehmaa	03	997281	0399728100
305	Kuusamo	18	990861	1899086100
614	Posio	18	990862	1899086200
832	Taivalkoski	18	990862	1899086200
005	Alajärvi	15	991011	1599101100
403	Lappajärvi	15	991012	1599101200
934	Vimpeli	15	991012	1599101200
069	Haapajärvi	18	991321	1899132100
317	Kärsämäki	18	991322	1899132200
691	Reisjärvi	18	991322	1899132200
205	Kajaani	19	992031	1999203100
578	Paltamo	19	992032	1999203200
697	Ristijärvi	19	992032	1999203200
785	Vaala	19	992033	1999203300
940	Vuolijoki	19	992033	1999203300
240	Kemi	20	992321	2099232100
241	Keminmaa	20	992322	2099232200
845	Tervola	20	992322	2099232200
261	Kittilä	21	992521	2199252100
047	Enontekiö	21	992522	2199252200
273	Kolari	21	992522	2199252200
498	Muonio	21	992522	2199252200
301	Kurikka	15	992831	1599283100
777	Suomussalmi	19	992832	1999283200
105	Hyrynsalmi	19	992833	1999283300
620	Puolanka	19	992833	1999283300
320	Kemijärvi	21	992981	2199298100
583	Pelkosenniemi	21	992982	2199298200
732	Salla	21	992982	2199298200
742	Savukoski	21	992982	2199298200
405	Lappeenranta	09	993051	0999305100
416	Lemi	09	993052	0999305200
831	Taipalsaari	09	993052	0999305200
978	Ylämaa	09	993052	0999305200
581	Parkano	06	993991	0699399100
250	Kihniö	06	993992	0699399200
303	Kuru	06	993992	0699399200
478	Maarianhamina	00	994341	0099434100
060	Finström	00	994342	0099434200
170	Jomala	00	994342	0099434200

Appendix 4...

035	Brändö	00	994343	0099434300
043	Eckerö	00	994343	0099434300
062	Föglö	00	994343	0099434300
065	Geta	00	994343	0099434300
076	Hammarland	00	994343	0099434300
295	Kumlinge	00	994343	0099434300
318	Kökar	00	994343	0099434300
417	Lemland	00	994343	0099434300
438	Lumparland	00	994343	0099434300
736	Saltvik	00	994343	0099434300
766	Sottunga	00	994343	0099434300
771	Sund	00	994343	0099434300
941	Vårdö	00	994343	0099434300
753	Sipoo	01	994971	0199497100
018	Askola	01	994972	0199497200
611	Pornainen	01	994972	0199497200
915	Varkaus	13	995721	1399572100
090	Heinävesi	12	995722	1299572200
212	Kangaslampi	12	995722	1299572200
061	Forssa	05	996141	0599614100
834	Tammela	05	996142	0599614200
169	Jokioinen	05	996143	0599614300
103	Humppila	05	996144	0599614400
981	Ypäjä	05	996144	0599614400
211	Kangasala	06	996161	0699616100
289	Kuhmalahti	06	996162	0699616200
439	Luopioinen	06	996162	0699616200
635	Pälkäne	06	996162	0699616200
730	Sahalahti	06	996162	0699616200
992	Äänekoski	14	996181	1499618100
275	Konnevesi	14	996182	1499618200
770	Sumiainen	14	996182	1499618200
774	Suolahti	14	996182	1499618200
560	Orimattila	07	996241	0799624100
015	Artjärvi	07	996242	0799624200
504	Myrskylä	07	996242	0799624200
616	Pukkila	07	996242	0799624200
729	Saarijärvi	14	996261	1499626100
216	Kannonkoski	14	996262	1499626200
226	Karstula	14	996262	1499626200
265	Kivijärvi	14	996262	1499626200
312	Kyyjärvi	14	996262	1499626200
633	Pylkönmäki	14	996262	1499626200
977	Ylivieska	18	996271	1899627100

Appendix 4...

009	Alavieska	18	996272	1899627200
746	Sievi	18	996272	1899627200
214	Kankaanpää	04	996291	0499629100
099	Honkajoki	04	996292	0499629200
181	Jämijärvi	04	996292	0499629200
230	Karvia	04	996292	0499629200
272	Kokkola	17	996341	1799634100
315	Kälviä	17	996342	1799634200
429	Lohtaja	17	996342	1799634200
885	Ullava	17	996342	1799634200
109	Hämeenlinna	05	996351	0599635100
082	Hattula	05	996352	0599635200
083	Hauho	05	996353	0599635300
210	Kalvola	05	996353	0599635300
692	Renko	05	996353	0599635300
434	Loviisa	01	996371	0199637100
424	Liljendal	01	996372	0199637200
585	Pernaja	01	996372	0199637200
407	Lapinjärvi	08	996373	0899637300
701	Ruotsinpyhtää	08	996373	0899637300
933	Vilppula	06	996390	0699639000
506	Mänttä	06	996391	0699639100
260	Kitee	12	996411	1299641100
248	Kesälahti	11	996412	1199641200
707	Rääkkylä	11	996412	1199641200
678	Raahe	18	996431	1899643100
625	Pyhäjoki	18	996432	1899643200
708	Ruukki	18	996432	1899643200
748	Siikajoki	18	996432	1899643200
636	Pöytyä	03	996491	0399649100
979	Yläne	03	996491	0399649100
019	Aura	03	996492	0399649200
219	Karinainen	03	996492	0399649200
074	Halsua	17	996521	1799652100
236	Kaustinen	17	996521	1799652100
584	Perho	17	996522	1799652200
924	Veteli	17	996522	1799652200
481	Masku	03	996571	0399657100
017	Askainen	03	996572	0399657200
419	Lemu	03	996572	0399657200
538	Nousiainen	03	996572	0399657200
906	Vahto	03	996572	0399657200
098	Hollola	07	996601	0799660100
283	Hämeenkoski	07	996602	0799660200

Appendix 4...

316	Kärkölä	07	996602	0799660200
180	Jyväskylän mlk	14	996611	1499661100
592	Petäjävesi	14	996612	1499661200
850	Toivakka	14	996612	1499661200
892	Uurainen	14	996612	1499661200
931	Viitasaari	14	996651	1499665100
256	Kinnula	14	996652	1499665200
601	Pihtipudas	14	996652	1499665200
593	Pieksämäki	10	996671	1099667100
085	Haukivuori	10	996672	1099667200
640	Pieksänmaa	10	996672	1099667200
010	Alavus	15	996681	1599668100
300	Kuortane	15	996682	1599668200
863	Töysä	15	996682	1599668200
887	Urjala	06	996691	0699669100
310	Kylmäkoski	06	996692	0699669200
928	Viiala	06	996692	0699669200
500	Muurame	14	996701	1499670100
277	Korpilahti	14	996702	1499670200
778	Suonenjoki	13	996731	1399673100
686	Rautalampi	13	996732	1399673200
844	Tervo	13	996732	1399673200
227	Karttula	13	996733	1399673300
921	Vesanto	13	996733	1399673300
491	Mikkeli	10	996741	1099674100
097	Hirvensalmi	10	996742	1099674200
696	Ristiina	10	996742	1099674200
598	Pietarsaari	16	996761	1699676100
440	Luoto	16	996762	1699676200
599	Pedersöre	16	996762	1699676200
684	Rauma	04	996771	0499677100
051	Eurajoki	04	996772	0499677200
266	Kodisjoki	04	996772	0499677200
406	Lappi	04	996772	0499677200
537	Noormarkku	04	996801	0499680100
484	Merikarvia	04	996802	0499680200
608	Pomarkku	04	996802	0499680200
747	Siikainen	04	996802	0499680200
743	Seinäjoki	15	996831	1599683100
544	Nurmo	15	996832	1599683200
589	Peräseinäjoki	15	996833	1599683300
975	Ylistaro	15	996833	1599683300
079	Harjavalta	04	996881	0499688100
262	Kiukainen	04	996882	0499688200
531	Nakkila	04	996882	0499688200

Appendix 4...

423	Lieto	03	996891	0399689100
284	Koski TL	03	996892	0399689200
480	Marttila	03	996892	0399689200
838	Tarvasjoki	03	996892	0399689200
202	Kaarina	03	996971	0399697100
602	Piikkiö	03	996972	0399697200
286	Kouvola	08	997001	0899700100
909	Valkeala	08	997002	0899700200
102	Huittinen	04	997011	0499701100
619	Punkalaidun	03	997012	0399701200
913	Vampula	03	997012	0399701200
430	Loimaa	03	997031	0399703100
431	Loimaa municipality	03	997032	0399703200
006	Alastaro	03	997033	0399703300
482	Mellilä	03	997033	0399703300
561	Oripää	03	997033	0399703300
178	Juva	10	997051	1099705100
623	Puumala	11	997052	1199705200
768	Sulkava	11	997052	1199705200
740	Savonlinna	11	997061	1199706100
618	Punkaharju	11	997062	1199706200
681	Rantasalmi	11	997062	1199706200
182	Jämsä	14	997071	1499707100
183	Jämsänkoski	14	997072	1499707200
291	Kuhmoinen	14	997072	1499707200
075	Hamina	08	997091	0899709100
489	Miehikkälä	08	997092	0899709200
935	Virolahti	08	997092	0899709200
734	Salo	03	997191	0399719100
073	Halikko	03	997192	0399719200
259	Kisko	03	997193	0399719300
587	Pertteli	03	997193	0399719300
252	Kiikala	03	997194	0399719400
308	Kuusjoki	03	997194	0399719400
501	Muurla	03	997194	0399719400
776	Suomusjärvi	03	997194	0399719400
174	Juankoski	13	997201	1399720100
204	Kaavi	13	997202	1399720200
857	Tuusniemi	13	997202	1399720200
399	Laihia	16	997221	1699722100
152	Isokyrö	16	997222	1699722200
942	Vähäkyrö	16	997222	1699722200
694	Riihimäki	05	997261	0599726100
086	Hausjärvi	05	997262	0599726200
433	Loppi	05	997263	0599726300

Appendix 5. Small area codes in the large cities

Municipality code	Name of municipality	Hospital district	Health centre area	Helsinki code	Small area code	New area code
297	Kuopio					
297	Zone 1	13	992800	0	1	1399280001
	City centre					
	Niirala					
	Haapaniemi					
	Rönö and islands near it					
	Siikalahti					
	Puijonlaakso-Taivaanpankko					
	Puijonsarvi (Päiväranta)					
	Puijo (Inkilänmäki)					
	Jukula					
	Niuva-Länsipuijo-Rypysuo					
	Laivonsaari					
	Itkonniemi-Männistö-Linnanpelto					
	Saarijärvi-Rahusenkangas					
	Kettulanlahti					
	Kelloniemi-Likolahti					
	Savilahti					
	Neulaniemi peak					
	Kolmisoppi-Vuorilampi					
	Neulamäki					
	Särkiniemi-Särkilahti					
	Sorsasalo					
297	**Zone 2**	13	992800	0	2	1399280002
	Small Neulamäki					
	Jynkkä					
	Levänen					
	Pitkälahti					
	Litmanen					
	Pirtti					
	Lehtoniemi-Keilankanta					
	Rautaniemi					
	Kiviharju					
	Vanuvuori					
	Hiltulanlahti					
297	**Zone 3**	13	992800	0	3	1399280003
	West Riistavesi					
	East Riistavesi					
	Melalahti					
	Vehmersalmi	13	992800	0	3	1399280003

Appendix 5...

297	**Zone 4** Puutossalmi Pellesmäki Kurkimäki Haminalahti Hirvilahti South Kallavesi islands North Kallavesi islands Ranta-Toivala Jännevirta Kurkiharju	13	992800	0	4	1399280004
049	Espoo					
049	Greater Leppävaara	01	901240	0	1	0190124001
049	Greater Tapiola	01	901240	0	2	0190124002
049	Greater Matinkylä	01	901240	0	3	0190124003
049	Greater Espoonlahti	01	901240	0	4	0190124004
049	Greater Kauklahti	01	901240	0	5	0190124005
049	Old Espoo	01	901240	0	6	0190124006
049	North Espoo	01	901240	0	7	0190124007
091	Helsinki					
091	Southern greater adm. area	01	901520	1	0	0190152010
091	Vironniemi primary adm. area	01	901520	1	1	0190152011
091	Ullanlinna primary adm. area	01	901520	1	2	0190152012
091	Kampinmalmi primary adm. area	01	901520	1	3	0190152013
091	Takatöölö primary adm. area	01	901520	1	4	0190152014
091	Lauttasaari primary adm. area	01	901520	1	5	0190152015
091	Western greater adm. area	01	901520	2	0	0190152020
091	Reijola primary adm. area	01	901520	2	1	0190152021
091	Munkkiniemi primary adm. area	01	901520	2	2	0190152022
091	Haaga primary adm. area	01	901520	2	3	0190152023
091	Pitäjämäki primary adm. area	01	901520	2	4	0190152024
091	Kaarela primary adm. area	01	901520	2	5	0190152025
091	Central greater adm. area	01	901520	3	0	0190152030
091	Kallio primary adm. area	01	901520	3	1	0190152031
091	Alppiharju primary adm. area	01	901520	3	2	0190152032
091	Vallila primary adm. area	01	901520	3	3	0190152033
091	Pasila primary adm. area	01	901520	3	4	0190152034
091	Vanhakaupunki primary adm. area	01	901520	3	5	0190152035
091	Northern primary adm. area	01	901520	4	0	0190152040
091	Maunula primary adm. area	01	901520	4	1	0190152041
091	West Pakila primary adm. area	01	901520	4	2	0190152042
091	Tuomarinkylä primary adm. area	01	901520	4	3	0190152043
091	Oulunkylä primary adm. area	01	901520	4	4	0190152044

Appendix 5...

091	Eastern Pakila primary adm. area	01	901520	4	5	0190152045
091	North-eastern greater adm. area	01	901520	5	0	0190152050
091	Latokartano primary adm. area	01	901520	5	1	0190152051
091	Pukinmäki primary adm. area	01	901520	5	2	0190152052
091	Malmi primary adm. area	01	901520	5	3	0190152053
091	Suutarila primary adm. area	01	901520	5	4	0190152054
091	Puistola primary adm. area	01	901520	5	5	0190152055
091	Jakomäki primary adm. area	01	901520	5	6	0190152056
091	South-eastern greater adm. area	01	901520	6	0	0190152060
091	Kulosaai primary adm. area	01	901520	6	1	0190152061
091	Herttoniemi primary adm. area	01	901520	6	2	0190152062
091	Laajasalo primary adm. area	01	901520	6	3	0190152063
091	Eastern greater adm. area	01	901520	7	0	0190152070
091	Vartiokylä primary adm. area	01	901520	7	1	0190152071
091	Myllypuro primary adm. area	01	901520	7	2	0190152072
091	Mellunkylä primary adm. area	01	901520	7	3	0190152073
091	Vuosaari primary adm. area	01	901520	7	4	0190152074
092	Vantaa					
092	Myyrmäki service area	01	901530	0	1	0190153001
092	Martinlaakso service area	01	901530	0	2	0190153002
092	Tikkurila service area	01	901530	0	3	0190153003
092	Korso-Koivukylä service area	01	901530	0	4	0190153004
092	Hakunila seervice area	01	901530	0	5	0190153005
179	Jyväskylä					
179	Regular city	14	901920	0	1	1490192001
179	Kypärämäki-Kortepohja	14	901920	0	2	1490192002
179	Lohikoski-Heinälampi	14	901920	0	3	1490192003
179	Huhtasuo	14	901920	0	4	1490192004
179	Kuokkala	14	901920	0	5	1490192005
179	Keltinmäki-Myllyjärvi	14	901920	0	6	1490192006
179	Keljo	14	901920	0	7	1490192007
179	Halssila	14	901920	0	8	1490192008
179	Säynätsalo	14	901920	0	9	1490192009
398	Lahti					
398	City centre (S1)	07	902990	0	1	0790299001
398	Mukkula (S2)	07	902990	0	2	0790299002
398	Kivimaa-Kiveriö-Joutjärvi (S3)	07	902990	0	3	0790299003
398	Ahtiala (S4)	07	902990	0	4	0790299004
398	Kolava-Kujala (S5)	07	902990	0	5	0790299005
398	Laune (S6)	07	902990	0	6	0790299006
398	Kärpänen (S7)	07	902990	0	7	0790299007
398	Jalkaranta (S8)	07	902990	0	8	0790299008

Appendix 5...

564	Oulu					
564	Höyhtyä health station	18	903900	0	1	1890390001
564	Kaijonharju health station	18	903900	0	2	1890390002
564	City centre health station	18	903900	0	3	1890390003
564	Kontinkangas health station	18	903900	0	4	1890390004
564	Myllyoja health station	18	903900	0	5	1890390005
564	Tuira health station	18	903900	0	6	1890390006
564	Rajakylä health station	18	903900	0	7	1890390007
837	Tampere					
837	City centre	06	905320	0	1	0690532001
837	North-eastern area	06	905320	0	2	0690532002
837	South-eastern area	06	905320	0	3	0690532003
837	Sourthern area	06	905320	0	4	0690532004
837	South-western area	06	905320	0	5	0690532005
837	North-western area	06	905320	0	6	0690532006
837	Northern area	06	905320	0	7	0690532007
853	Turku					
853	City centre	03	905450	0	1	0390192001
853	Hirvensalo-Kakskerta	03	905450	0	2	0390545002
853	Uittamo-Skanssi	03	905450	0	3	0390545003
853	Itäharju-Varissuo	03	905450	0	4	0390545004
853	Koroinen	03	905450	0	5	0390545005
853	Tampereentie	03	905450	0	6	0390545006
853	Kuninkoja	03	905450	0	7	0390545007
853	Naantalintie	03	905450	0	8	0390545008
853	Maaria-Paattinen	03	905450	0	9	0390545009
609	Pori					
609	City centre	04	994240	0	1	0499424001
609	Northern Terrestrial Pori	04	994240	0	2	0499424002
609	Eastern Terrestrial Pori	04	994240	0	3	0499424003
609	Western Terrestrial Pori	04	994240	0	4	0499424004
609	Marine Pori I	04	994240	0	5	0499424005
609	Marine Pori II	04	994240	0	6	0499424006
609	North Pori	04	994240	0	7	0499424007
609	South-West Pori	04	994240	0	8	0499424008
442	Luvia	04	994240	0	8	0499424008

Appendix 6. Data from the Finnish Care Register (HILMO) and Benchmarking project

Note! Because HILMO has undergone major changes in years 1994 (HILMO), 1996 (ICD10) and 1997 (NOMESCO classification of procedures) the data is often delivered in three corresponding parts (90–93, 94–95 and 96–). Changes in variables and lables in time can be detected from the variable list. Psychiatric patient's extra sheet was only used in Psychiatric hospital care sub-study.

	Variable name	Years	Content
90-93			
1	TUTNRO	90-03	Identification number (Statistics Finland)
2	VUOSI	90-03	Year of the hospital admission
3	PALTU	90-03	Service provider (hospital)
4	EALA	90-93	Doctors speciality
5	KOKU	90-03	Municipality of residence (not included in the datasets)
6	TPVM	90-03	Date of admission
7	PPVM	90-03	Date of discharge
8	STAP	90-93	Mode of admission
9	MLUO	90-93	Payment category
10	ONKOL	90-93	Previously operated
11	TMP1	90-96	Surgical procedure 1 (Hospital League)
12	TMP2	90-96	Surgical procedure 2(Hospital League)
13	TAPPAI	90-95	Place of accident
14	EDIA	90-95	External cause of accident
15	PDG	90-95	Main diagnosis (ICD9)
16	SDG1	90-95	1st subsidiary diagnosis (ICD9)
17	SDG2	90-95	2nd subsidiary diagnosis (ICD9)
18	DG4	90-93	3rd subsidiary diagnosis (ICD9)
19	MISTA	90-93	Where the patient came from
20	MIHIN	90-93	Where the patient goes at discharge
21	HOITOPV	90-03	Number of bed days
22	IKA	90-03	Age
23	SUKUP	90-03	Sex
24	KNTSHP	90-03	Hospital district of the municipality of residence
25	JPVM	92-03	Date of admittance to the queue to surgery
26	JOPV	92-93	Days in the queue to surgery
94-95			
27	PALTUTAR	94-03	Specification of the provider code
28	ULASU	95-03	Code for a patient living abroad
29	PALA	94-03	Service sector
30	EA	94-03	Doctors speciality
31	SATAP	94-03	Mode of admission
32	TULI	94-03	Where the patient came from
33	LPKOD	94-03	Code for the unit where the patient came from
34	LPKODTAR	95-03	Extension for LPKOD
35	LANT	94-03	Refering doctor

Appendix 6...

36	LANTKO	94-03	Code for the refering unit
37	LANTTAR	95-03	Extension for LANTKO
38	TUSYY1	94-03	Main reason for seeking treatment
39	TUSYY2	94-03	Other reason for seeking treatment 1
40	TUSYY3	95-03	Other reason for seeking treatment 2
41	HOITOITU	95-03	Need of treatment at arrival to hospital
42	HOITOI	94-03	Need of treatment when discharged from hospital
43	TMP3	94-96	Surgical procedure 3 (Hospital League)
44	OPVM	94-03	Date of procedure
45	PITK	94-03	Decision of long-term care
46	SYP	94-03	Demanding heart patient (if yes, the extra sheet is opened)
47	PSYKP	94-03	Psychiatric patient (if yes, the extra sheet is opened)
48	ERITKO	94-95	Code for specialised hospital care
49	LOMAPV	95-03	Days of sickleave
50	JATKOH	94-03	Extended care
51	JATKOOD	94-03	Code for the unit of extended care
52	JATTAR	95-03	Extension for JATKOOD
53	AS	94-03	Hospital charges payable to the patient
54	KU	94-03	Hospital charges payable to the municipalitie
55	MAPER	94-03	Grounds for hospital charges
56	TO	94-03	Hospital charges payable to an other hospital district
57	MU	94-03	Hospital charges payable to other institutions
96-03			
58	PDGO	96-03	Main diagnosis, symptomatic (ICD10)
59	PDGE	96-03	Main diagnosis, etiological (ICD10)
60	SDG1O	96-03	Subsidiary diagnosis 1, symptomatic (ICD10)
61	SDG1E	96-03	Subsidiary diagnosis 1, etiological (ICD10)
62	SDG2O	96-03	Subsidiary diagnosis 2, symptomatic (ICD10)
63	SDG2E	96-03	Subsidiary diagnosis 2, etiological (ICD10)
64	ULKSYY	97-03	External cause of accident
65	TAPTYYPPI	98-03	Type of accident
66	PTMPK1	97-03	Main procedure of the main operation 1 (NOMESCO)
67	PTMPK2	97-03	Main procedure of the main operation 2 (NOMESCO)
68	PTMPK3	97-03	Main procedure of the main operation 3 (NOMESCO)
69	MTMP1K1	97-03	Other procedures 1 (NOMESCO)
70	MTMP2K1	97-03	Other procedures 2 (NOMESCO)
Demanding heart patients extra sheet			
71	TMPTYP1	95-03	Type of cardiac procedure 1
72	TMPTYP2	95-03	Type of cardiac procedure 2
73	TMPTYP3	95-03	Type of cardiac procedure 3
74	TMPLAJ	94-03	Procedure class
75	NYHA	94-03	Capability score (NYHA)
76	RISKPI	94-03	Risk score for coronary surgery (Euroscore)
77	TYTILE	94-03	Employment status before hospital admission

Appendix 6...

78	TMPPRI	94-03	Primary outcome of the cardiac procedure
79	TMPTYP	94	Type of cardiac procedure
80	TMPTAR	94-95	Extension for TMPTYP
81	TYSTAT	94-95	Employment status before hospital admission
82-92	TMPC1-TMPC11	03	Procedure types 1-11
93	TMPLAJC	03	Procedure class
94	TMPKIIR	03	Urgency of the procedure
95	RISKEUR	03	Risk score for coronary surgery (Euroscore)
96	RISKEURL	03	Logistic risk score for coronary surgery (Euroscore)
98-101	KOMPL1-KOMPL5	03	Complications 1-5
Psychiatric patient's extra sheet*			
102	TUTAP	94-03	Manner of arrival at psychiatric ward
103	TRPV	94-03	Number of days in hospital care against will
104	HOIKER	94-03	Number of psychiatric admissions
105	TUGAS	94-03	GAS assessment at admission
106	POGAS	94-03	GAS assessment at discharge/at the moment of patient census
107	LHOI	94-03	Drug therapy
108	PAKKTOI1	94-03	Coersive measure 1
109	PAKKTOI2	94-03	Coersive measure 2
110	PAKKTOI3	94-03	Coersive measure 3
111	PAKKTOI4	94-03	Coersive measure 4

* Operational from 1995

Variables used for the Helsinki City sub-study from the Benchmarking data base are mainly the same as the HILMO variables. Benchmarking data is available also for the city districts in Helsinki area.

Appendix 7. Data from the Social Insurance Institution

1. Special reimbursement right register

1	TUTNRO	Identification code (Statistics Finland)
2	SK1	Code of chronic disease
3	APVM6	Date for the beginning of reimbursement right (1st of the month)
4	LPVM6	Expiration date for the reimbursement right (999912=permanent)

2. Prescription register

1	TUTNRO	Identification code (Statistics Finland)
2	ATC	Medicine code (ATC)
3	VNRO	Nordic serial number for the medicine
4	OTPV	Date of purchase
5	PLKM	Number of pacakges purchased
6	KUST	Expences (->2001 penni and 2002 -> cent)
7	KORV	Reimbursement (->2001 penni and 2002 -> cent)
8	SAIR	Disease code for special reimbursement right (SII code)
9	LAJI	Reimbursement type (50% / 75% / 100%)
10	DDD	Defined Daily Dose (derived from SII registers by serial number)
11	SVNO	Doctor’s health insurance code
12	ERIK1	Doctor’s speciality field 1
13	ERIK2	Doctor’s speciality field 2

3. Pension register

1	TUTNRO	Identification code (Statistics Finland)
2	ELAJI	Type of pension
3	VAIKA	Effective date
4	ZEA	Original effective date
5	ZPO	Advance date of expiry
6	SAIR	Main cause of disability (ICD-code)
7	LSA1	Additional cause of disability 1 (ICD-code)
8	LSA2	Additional cause of disability 2 (ICD-code)

Appendix 8. Data from the Finnish Centre for Pensions

1	TUTNRO	Identification number (Statistics Finland)
2	EA00KID	Technical identifier for the aggregate of pensions
3	EA01EKID	Technical identifier for pension
4	PE10LAJI	Type of pension
5	PE10APPV	Date of admission
6	PE10AALK	Effective date
7	PE10LAKI	Pension law
8	PE10PPVM	Date of expiry
9	TKSYY1	Cause of disability to work 1 (ICD)
10	TKSYY2	Cause of disability to work 2 (ICD)

Appendix 9. Data from the City of Helsinki

	Variable name	Years	Content
1	TUTNRO	1998-03	Identification number (Statistics Finland)
2	D1-8	1998-03	1.-4 main symptomatic and etiological causes
3	EALAY	1998-01	Speciality of the service provider
4	EALYH	1998-01	Sector of the service provider
5	KAY_KLO	1998-03	Time of visit
6	KAY_PVM	1998-03	Date of visit
7	KUNTA	1998-03	Municipality
8	MIHIN	1998-03	Where the patient goes at discharge
9	MIHIN_YKS	1998-03	The unit patient goes to at dicharge
10	MISTA	1998-03	Where the patient came from
11	MIS_YKS	1998-03	The unit patient came from
12	PH_KOODI	1998-03	The code of the unit
13	PH_NIMI	1998-03	The name of the unit
14	PMU	1998-03	Service sector
15	PPIIRI	1998-03	The small area code of the unit
16	SAIR	1998-03	Hospital number
17	SH_TH	1998-03	1=preventive care, 2=sickness care
18	T1-T5	1998-03	Procedure codes
19	TAPTYYP	1998-01	Type of accident (ICD10 codes Y940-Y960)
20	TTPAIKK	1998-03	Place of accident
21	VORYHMA	1998-01	Profession of the attending health personnel
22	KTYYPPI	1998-03	Type of visit
23	TYYPPI	2002-03	Health centre, hospital policlinic, inpatient
24	ALUEKOODI	2002-03	The small area code of the unit
25	D9-D20	2002-03	5.-7 main symptomatic and etiological causes
26	EA_LYH	2002-03	Sector of the service provider
27	EK_UK	2002-03	First or subsequent visit
28	ERIKOISALA	2002-03	Speciality of the service provider
29	KAUNTITAPA	2002-03	Type of health care contact
30	PALVELU_ID	2002-03	Procedures
31	PMU2	2002-03	Service sector 2
32	SAIR_KT	2002-03	Type of visit to the hospital
33	SAIR_TILAST	2002-03	Hospital visit entering the statistics
34	VUOSI	2002-03	The year entering the data
35	SUOR_PA_ID	2002-03	The ID of the healt care unit
36	SAIR_KTYYP	2002-03	Type of visit in specialist care
37	KAYNTITYYP	2002-03	Type of visit
38	LISATIEDOT	2002-03	Visit entering the statistics
39	RES_RYHMA	2002-03	Type of health care resource
40	SAIR_TYYPP	2002-03	Type of hospital

Appendix 10. Area level variables used in REDD

Note! Information is requested by health centre area for all relevant years

SOTKA and Statistics Finland		
1	TYKAZ	Employment rate (% of the work force)
2	TYPIX	Long-term unemployed (% of the unemployed)
3	TYNUZ	Juvenile unemployment rate (% of the 15–24 year old work force)
4	TKELX	Disability pensions (/1000)
5	RIKOX	Crimes against property (/1000 inhabitants)
6	RIKVX	Violent crimes (/1000 inhabitants)
7	SIJOX	Children taken into custody (/1000 inhabitants)
8	KOKEX	Secondary education (% of over 15 years)
9	KOKOX	Higher education (% of over 15 years)
10	TOITX	Income support (% of all residents)
11	AKAHX	Confined accommodation (% of all housholds)
12	MUKIX	Foreign language speakers (/1000 native language not Finnish or Swedish)
13	PPOTX	Intoxicant users (/1000)
14	PSAIX	Number of psychiatric hospital days among 15–64 -year olds (/1000)
15	PSLAX	Number of psychiatric hospital days among 0–14 -year olds (/1000)
16	LKORX	Reimbursed medicine costs (persons/1000)
17	LERIX	Special reimbursement of medicine costs (persons/1000)
18	KUOLX	Mortality rate
19	HUOLTZ	Demographic dependency ratio
20	ELSUZ	Economic dependency ratio
21	KUNEX	Net migration/1000
Social Insurance Institution		
22		Medicine reimbursements by ATC code
23		Reimbursements for visits to private doctors
24		Reimbursements for care and examinations
25		Number of entitlements for special reimbursements by disease code
26		Number of new disability pensions (5 disease groups)*
27		Number all inhabitants on disability pension (5 disease groups)*

* Includes Centre for Pensions data

Appendix 11. Procedure codes for elective surgery sub-study

Procedure	Remarks	Procedure code (NOMESCO)	Procedure code (League of Hospitals)
Coronary revascularisation (CABG; PTCA)	CABG	FNA-E	5311-5319
	PTCA	FN1AT	
	NB: PTCA codes	FN1BT	
	n.a. before 1994	FN1YT	
		FNAG	
		TFN40	
		TFN50	
	demanding heart patients extra sheet		
	CABG	11, 25,	
	PTCA	82, 83, 84	
	1994 HILMO		
	PTCA	tmptyp=1,	
	CABG	tmptyp=2	
Hip replacement Exclusion criteria: ICD10 S72	first operation reoperation	NFB30-99, NFC00-99	9293- 9294, 9321-9329
Knee replacement	first operation reoperation	NGB10-99 NGC00-99	9313, 9314, 9331-9339
Lumbar disc operation	NB: Check whether the ICD10 dg is M51.1	ABC16, ABC26	9211
Cataract operation		CJC00-99, CJD00-99, CJE00-99	3623, 3624, 3625, 3626, 3627, 3628, 3630, 3631, 3633, 3639
Hysterectomy		LCC00-01, LCC10-97, LCD00-97, LEF13	8305, 8306, 8307, 8308, 8310, 8319, 8436
Prostatectomy		KEC00-20,	7451, 7452,
		KED00-10,	7456, 7459
		KED22, KED52,	
		KED58, KED62,	
		KED76	