Methodological Perspectives for Register-Based Health System Performance Assessment

Developing a Hip Fracture Monitoring System in Finland
Editorial board
Matti Heikkilä, chairman
Marjatta Bardy
Marko Elovainio
Mika Gissler
Riitta Haverinen
Timo Tuori
Matti Virtanen

The research presented in this series has been approved for publication after undergoing a formal referee evaluation process.

© Author and STAKES

Translation of the Swedish abstract: Done Information Oy

Cover design: Harri Heikkilä
Layout: Christine Strid

ISBN 978-951-33-2114-7 (nid.)
ISSN 1236-0732
ISBN 978-951-33-2132-1 (PDF)

Gummerus Printing
Vaajakoski 2008
Finland
Dedicated to the memory of my father
Acknowledgements

First of all, I want to praise my beloved wife Anne, the queen of my heart and the goddess of my soul, for giving a purpose to my life. I would also like to express my gratitude to my mother Irja and my sister Paula for their unconditional love, and to my goddaughter Saana-Emilia for bringing so much sunshine that this world has become a happier place.

The preparation of this thesis would not have been possible without a research grant for the project ’Register-based data in hospital outcomes research’ from the Academy of Finland. I am grateful to Professor Pekka Rissanen and Adjunct Professor Ilmo Keskimäki for guiding me to the world of health services research. I also thank the collaborators within the ’Performance, effectiveness and cost of treatment episodes’ -project led by Research Professor Unto Häkkinen. In addition, I want to acknowledge the role of my other co-authors and the colleagues I have worked with. Thank you; you know who you are.

I express my sincere appreciation to Professor Emeritus Seppo Mustonen. His extraordinary creativity, talent, and experience combined with the ability to say the right things at the right time have offered an invaluable source of inspiration and support during my whole statistical career. Moreover, the general computing environment provided by Survo, developed by Professor Mustonen, offered an excellent platform for register-based data analysis.

I also thank Professor Risto Lehtonen, Professor Hannu Niemi and Professor Esa Lääärä for comments and discussions concerning the manuscript of this thesis. I am thankful to the official reviewers of the dissertation, Adjunct Professor Timo Alanko and Adjunct Professor Helka Hytti. Their comments helped me to further simplify and clarify the message.

Finally, one thing has become clear to me during the preparation of this thesis. It is extremely time-consuming to conduct and publish multidisciplinary scientific research that aims to offer alternative solutions to pragmatic problems. Although I had to make some compromises, I still hope that this thesis is able to reflect my personal way of thinking as well as my enthusiasm for scientific research.
Abstract


The resources of health systems are limited. There is a need for information concerning the performance of the health system for the purposes of decision-making. This study is about utilization of administrative registers in the context of health system performance evaluation.

In order to address this issue, a multidisciplinary methodological framework for register-based data analysis is defined. Because the fixed structure of register-based data indirectly determines constraints on the theoretical constructs, it is essential to elaborate the whole analytic process with respect to the data. The fundamental methodological concepts and theories are synthesized into a data sensitive approach which helps to understand and overcome the problems that are likely to be encountered during a register-based data analyzing process.

A pragmatically useful health system performance monitoring should produce valid information about the volume of the problems, about the use of services and about the effectiveness of provided services. A conceptual model for hip fracture performance assessment is constructed and the validity of Finnish registers as a data source for the purposes of performance assessment of hip fracture treatment is confirmed. Solutions to several pragmatic problems related to the development of a register-based hip fracture incidence surveillance system are proposed. The monitoring of effectiveness of treatment is shown to be possible in terms of care episodes. Finally, an example on the justification of a more detailed performance indicator to be used in the profiling of providers is given.

In conclusion, it is possible to produce useful and valid information on health system performance by using Finnish register-based data. However, that seems to be far more complicated than is typically assumed. The perspectives given in this study introduce a necessary basis for further work and help in the routine implementation of a hip fracture monitoring system in Finland.

Keywords: administrative registers, methodology, secondary data, hip fractures, health services, performance, effectiveness, incidence, profiling, care episodes
Abstract in Finnish


Terveydenhuoltojärjestelmän resurssit ovat rajallisia. Jotta näitä resurseja pystyttäisiin hyödyntämään mahdollisimman tarkoituksennakaisällä tavalla, tarvitaan tietoa terveydenhuoltojärjestelmän vaikuttavuudesta. Tässä tutkimuksessa tarkastellaan kuinka suomalaisia rekisteriaineistoja voidaan käyttää päätoimen tekovälineen tukevan terveydenhuollon vaikuttavuustiedon tuottamiseen.

Tutkimuksessa kehitetään menetelmällinen viitekehyks rekisteriperusteisen datan analyysin tukea yhdistämällä eri tieteenaloilta periytyviä menetelmällisiä ideoita. Lähtökohtana on, että rekistereitä hyödyntävän tutkimusprosessin tulee edetä myös aineiston ehdoin, koska hallinnollisiin tarkoituksiin kerättyjen rekisteriaineistojen tietosisältö rajoittaa käyttömahdollisuuksia. Kehitettynä viitekehyksen puitteissa näiden välttämättömien rajoitusten syitä ja seurauksia on mahdollista käsitellä systemaattisella tavalla, joka auttaa ymmärtämään ja ratkaisemaan rekisteriperusteisessa tutkimuksessa tyyppillisesti kohdattavia ongelmia.


Johtopäätöksenä voidaan todeta, että suomalaiset rekisteriaineistot tarjoavat erinomaiset mahdollisuudet terveydenhuollon vaikuttavuutta koskevan tiedon tuottamiseen, vaikka hyödyllisen ja tieteellisyyden kriteerit täyttävän tiedon tuottaminen onkin varsin haasteellista. Tässä tutkimuksessa esitetyt näkökulmat muodostavat perustellun lähestymistavan rekisteritietojen entistä tehokkaampaan hyödyntämiseen ja auttavat lonkkamurtuman rutiniluontiseeman seurantajärjestelmän kehittämisessä.

Avainsanat: hallinnolliset rekisterit, metodologia, toisen käden aineistot, lonkkamurtumat, terveydenhuolto, vaikuttavuus, ilmaantuvuus, laatu, hoitoketjut
Abstract in Swedish


Vid bedömnaden av hälso- och sjukvårdens effektivitet krävs uppgifter om hälsoproblemens incidens, användningen av tjänsterna och vården kvalitet. För dessa ändamål utvecklas i denna undersökning en begreppsmodell för databehovet när det gäller uppföljningsystem för höftfrakturer och bekräftas att validiteten hos data i registermaterialet till dessa delar är tillräcklig. En metod som är förenlig med registermaterialen och som löser flera praktiska problem som är förknippade med uppföljningen av incidensen av höftfrakturer presenteras. Dessutom visas hur genomförandet av vården kan utvärderas med hjälp av registren för hela vårdkedjans del. Slutligen granskas hur vårdens kvalitet kan jämföras mellan olika sjukhus med hjälp av en mer detaljerat underbyggd effektivitetsindikator.

Som slutsats kan man konstatera, att de finska registermaterialen erbjuder utmärkta möjligheter att producera uppgifter om hälso- och sjukvårdens effektivitet, även om det innebär en ganska stor utmaning att försöka ta fram uppgifter som uppfyller kriterierna för användbarhet och vetenskaplighet. De synpunkter som presenteras i den här undersökningsbildar en underbyggd angreppsmetod som gör det lättare än tidigare att utnyttja registeruppgifter på ett effektivt sätt och hjälper vid utvecklandet av ett med rutinmässigt system för uppföljning av höftfrakturer.

Nyckelord: administrativa register, metodologi, andrahandsmaterial, höftfrakturer, hälso- och sjukvård, effektivitet, incidens, kvalitet, vårdkedjor
Contents

Acknowledgements

Abstract
Abstract in Finnish
Abstract in Swedish

List of original publications .............................................................................................................. 13

1 Introduction ................................................................................................................................... 15
   1.1 Aim of the study .................................................................................................................. 16
   1.2 Structure of the study ......................................................................................................... 17

2 Register-based data analysis ....................................................................................................... 19
   2.1 From technical data processing to a knowledge discovery process ................................ 19
   2.2 Secondary data as a methodological problem ................................................................... 20
   2.3 Prerequisites in the register-based data analysis ............................................................... 21
      2.3.1 Principles of measurement ....................................................................................... 22
      2.3.2 Information science ................................................................................................. 23
      2.3.3 Statistical computing ............................................................................................... 25
      2.3.4 Statistics .................................................................................................................. 26
      2.3.5 Theory ..................................................................................................................... 27

3 Register-based health system performance assessment ............................................................. 28
   3.1 Finnish health system .......................................................................................................... 28
   3.2 Information production in the Finnish context ............................................................... 29
   3.3 Measurement of the performance of a health system ....................................................... 29
   3.4 Assumptions for the production of concrete health system performance information .......... 30
      3.4.1 Register-based health system performance monitoring ............................................ 31

4 Register-based data on hip fractures – sources and validity ...................................................... 33
   4.1 Hip fracture ......................................................................................................................... 33
      4.1.1 Hip fracture treatment in Finland ............................................................................. 33
      4.1.2 Register data on hip fractures in Finland ................................................................. 34
      4.1.3 Previous register-based hip fracture studies in Finland ............................................. 35
      4.1.4 Register data for the current study ............................................................................ 35
   4.2 Validity of register data in the case of hip fracture .............................................................. 36
      4.2.1 Completeness of registration .................................................................................. 37
      4.2.2 Conceptual model for hip fracture performance monitoring ..................................... 37
      4.2.3 Consistency between prospective and register-based data ...................................... 40
      4.2.4 Conclusions on validity ......................................................................................... 41

5 Hip fracture incidence .................................................................................................................. 43
   5.1 Aging-related hip fractures ............................................................................................... 44
      5.1.1 A method for identifying a first aging-related hip fracture ....................................... 45
   5.2 Risk factor extraction ......................................................................................................... 47
   5.3 Risk population data ......................................................................................................... 49
   5.4 Hip fracture incidence between 1998 and 2002 in Finland .............................................. 51
   5.5 Conclusions on hip fracture incidence monitoring ............................................................. 54
6 Hip fracture treatment ................................................................. 55
6.1 State space for a hip fracture care episode .................................. 55
6.2 Effectiveness in terms of care process ......................................... 57
6.3 Multivariate responses in outcomes research ............................... 57
   6.3.1 Traceplots for care episode visualization ............................... 58
6.4 Comparing the care episode profiles of subpopulations ............... 61
   6.4.1 Risk adjustment .................................................................. 61
   6.4.2 Summarization of the follow-up data .................................... 62
6.5 Conclusions on monitoring hip fracture treatment ...................... 64

7 Operative delay as a performance indicator ................................. 66
7.1 Effects of different operative delays on mortality ...................... 66
7.2 Profiling of providers ................................................................ 67
7.3 Adjusted effect of operative delay on mortality ......................... 72
7.4 Provider-level hypotheses ....................................................... 73
7.5 Conclusions on operative delay as a performance indicator ........ 76

8 Discussion .................................................................................. 77
References .................................................................................... 80

Original publications
List of original publications


Articles are included in the thesis with permissions from the publishers.
1 Introduction

The resources available to health systems are limited. The equitable, efficient, and effective use of these resources requires administrative planning and political willingness. Information is needed concerning the performance of the health system for the purposes of decision-making [10].

Recently, a continuous evidence-based quality improvement aiming to “do the right thing right” has become very popular [11]. Essentially, this approach can be seen as an application of organizational learning and knowledge management ideas to the health care context [12]. To fulfill the requirements of this paradigm, information should be evidence-based, well-organized, pragmatic, task-specific, and available when and where it is needed [13].

It is obvious that information systems have a central role in this kind of information production. A comprehensive health information system should be suitable for clinical work as well as nationwide health policy purposes [14, 15]. The idea of such population data-based health information systems is not new [16], and some good examples do exist [17]. Finland has had a long tradition of maintaining administrative information systems, and extensive nationwide registration—using an individual’s single personal identification number for all systems—is exceptional also from the international point of view [18]. However, even though the potential advantages of Finnish information systems are well recognized, the utilization of these data sources has been rather limited.

This controversy is not specific to Finnish health information systems. In fact, advances in information technology have made it possible to produce and store all kinds of data effectively, but the emphasis has been on technical aspects and not on the information itself [19]. For as long as there have been such information systems, the problem of giving too much data in an unusable form has been a common complain [20]. In fact, the predominant belief that secondary data stored in the health information system consist of autonomous, atom-like building blocks is fundamentally erroneous. Therefore, it is also erroneous to believe that the production of more detailed data would solve the basic problems. On the contrary, the more detailed and complex are the variables to be recorded, the more background information and tacit knowledge is required for secondary utilization. Such a fact is known as the law of medical information [21]. In this sense, it can be expected that as the amount of data increases with the introduction of electronic patient records and correspondingly complex information systems, it will be even more difficult to transform data into useful information.

Traditionally, the problems in analyzing data have been considered as statistical issues, but statistical research focusing on probabilistic inference based on mathematics has not been able to offer enough concrete help in the analysis
of growing amounts of data [22]. The practical need for information has led to the development of alternative ways of analyzing data, such as data mining [23]. However, the development of new tools for data analysis has not solved the actual problem of transforming raw data into useful information.

As a matter of fact, it is naïve to believe that there would be some magic trick to overcome problems arising from the philosophical and methodological limitations of empirical research. Rather than trying to squeeze the data into a predefined model or saying too much on what can and cannot be done, data analysis should work to achieve an appropriate compromise between the practical problems and the data [24]. This kind of activity has been characterized as “greater statistics”, which tends to be inclusive, and eclectic with respect to methodology, while being closely associated with other disciplines and also practiced by many non-statisticians [25]. Because formal statistical expertise provides an excellent basis for the understanding and evaluation of methodological ideas, statisticians ought to take advantage of the situation, get involved in interdisciplinary activities, learn from the experience, expand their own minds—and thereby also their field—and act as catalysts for the dissemination of insights and methodologies [26].

1.1 Aim of the study

This thesis is about scientific research methodology in the context of register-based health system performance assessment. The application field is health services research. Health services research differs from most areas of applied research, since it does not exist in isolation from the decision-making procedures needed in health policy. A multidisciplinary approach and knowledge is required in health services research, including perspectives from at least biological, medical, social, clinical, management, economic, statistical and information sciences [27].

Another special aspect here is the use of register-based data. Since register-based data have been originally produced for other purposes than for this specific research, the traditional methodological assumptions concerning the nature of data are not valid.

In addition, information that allows for the monitoring of health system performance should be available if a health information system is to be adequate. This means that the perspectives of routine statistics production and the technical possibilities of information systems must also be considered simultaneously by means of a more research-oriented approach. Unfortunately, there are no widely accepted principles for the production of register-based statistics, which makes the development of register-statistical methodology challenging [28].

These issues make things rather complex. It is not possible to rely solely on a particularly well-developed paradigm and so the only reasonable way is to accept the diversity of the approaches and learn to deal with them simultaneously. This
can be done by “raising” the ideas from the application level to the philosophical–methodological level. However, the abstract ideas may become real only through the applications. In other words, the innovative results can be achieved only by rethinking the methodology separately for each specific application.

In short, the aim of this thesis is to give multidisciplinary methodological perspectives for scientific register-based information production for the purposes of health system performance assessment. More specifically, the goals are 1) to develop a methodological framework that helps to utilize register-based data effectively, 2) to demonstrate how the register-based data can be used as a data source for the performance monitoring of a health system. The specific application area is hip fracture from the public health point of view.

1.2 Structure of the study

This thesis consists of this summary and the following six original contributions.


This summary synthesizes the methodological ideas from these contributions. The second chapter is about register-based data analysis in general. It begins with the database-oriented approach to data analysis and then extends the perspective to a more comprehensive research-process approach which is able to deal with secondary data, which is the topic of article 1. The latter part of chapter two gives a brief review of certain multidisciplinary ideas that help in understanding the problems likely to be encountered during a register-based data analyzing process. The general methodological ideas that I have presented mainly in articles 1 and

1 Articles are included in the thesis with permissions from the publishers.
Introduction

3 and developed further elsewhere [7, 8], are in this chapter reflected against the existing literature. As a whole, chapter two defines a methodological framework for register-based data analysis and describes fundamental assumptions that have been applied in the later chapters.

In chapter three, some fundamental background information on health system performance assessment in Finland is given and a perspective for the applied research problem is fixed with certain focusing assumptions. This chapter extends and focuses the introduction given in article 2 on the basis of my other publications [7–9]. The fourth chapter includes a brief introduction to hip fracture, register-based sources of information on hip fractures, and a review on previous register-based research on hip fractures. The latter part of the fourth chapter summarizes the contents of article 2, which defined a conceptual model for the hip fracture performance assessment and examined the validity of Finnish registers as a data source for the purposes of a performance assessment of hip fracture treatment.

Chapter five, based on articles 3–5, deals with the pragmatic problems related to the development of a surveillance system for register-based hip fracture incidence. The sixth chapter updates the conceptual model on the hip fracture treatment process presented in article 1 so as to be compatible with the conceptual model on hip fracture performance monitoring reported in article 2 and gives a very basic example of an application of the conceptual model. I have already presented the ideas summarizing the approach elsewhere [9], and prepared several manuscripts that deal with the application in more detail. Chapter seven gives a more detailed example on the justification of a performance indicator, and elaborates on the methodological ideas presented in article 6, including a basic model for the profiling of providers first described in article 1.

Finally, in the eighth chapter, the motives for this study are once more discussed, the idea of the methodological approach is briefly summarized, and whether the aims of the study have been fulfilled is evaluated.
2 Register-based data analysis

A register is an information system which continuously records subject-based data for a particular complete set of subjects. A register contains a logically coherent collection of related data with some inherent meaning, typically reflecting some events occurring in the observational reality. A register is implemented for a specific purpose and has an intended group of users. These properties also fulfill the definition of a database and hence a register can be considered as a database [29]. Typically a register is a large collection of data and is maintained using computers. As long as there have been (computerized) databases, there has been an increasing interest to provide information at the correct level of detail to support decision-making [30, 31].

2.1 From technical data processing to a knowledge discovery process

In principle, it is quite straightforward to implement procedures that take register data as input and produce a sensible summarization of them as an output. For example, intensive cross-tabulations and basic visualization tools can be used for the comprehensive summarization of stored data. In addition, implementations of many traditional statistical techniques are suitable for such processing. However, the massive size of a database may impose certain computational difficulties in applying methods initially developed for smaller data sets [32]. Therefore, more recent and computationally efficient methods and algorithms offer an invaluable enhancement to data analysis [33, 34].

In practice, the description of data analysis as a finite series of precisely encoded rules needed to transform raw data in a database into interesting information gives an overly simplistic impression of data analysis, because it remains unclear what should actually be done in order to achieve information in data analysis. In this sense, it is convenient to consider data analysis as a collection of tasks, such as exploratory data analysis, descriptive and predictive modeling, pattern discovery, and identification [35]. In fact, it has been suggested that reasonable results can be achieved by considering the whole series of actions required to transform data into information as a knowledge discovery process [36]. A descriptive model of a standard process of data mining—which gives detailed pragmatic guidelines for performing a knowledge discovery process—has been developed [37]. In this business-oriented tradition, the goals of the process are evaluated in terms of its ability to produce interesting information. Suggested criteria for interesting
information are, for instance, evidence, non-redundancy, simplicity, novelty, and usefulness [38].

The problem with such a pragmatically useful description of a business-oriented knowledge discovery process is that certain important aspects of traditional empirical research and scientific method have been disregarded. Therefore I have complemented the process model to give a more comprehensive and detailed description of the phases commonly encountered during an empirical research process that incorporates register-based data [1]. The improved schema for a knowledge discovery process is presented in Figure 1. The main phases in such process are: understanding the phenomenon, understanding the problem, understanding data, data preprocessing, modeling, evaluation and reporting [1]. The case study of this thesis represents an application of this schema.

2.2 Secondary data as a methodological problem

If the schema presented in Figure 1 is compared to a standard scientific inquiry, the most important difference is the connection between the problem and the data. In fact, typically a theory of a given phenomenon of interest should drive the primary data collection. This makes it possible to concentrate on just those parts of observable reality that are considered most relevant for the current theoretical purpose. A similar approach is not possible with register-data, as the formulation of a problem in terms of register-based data is opportunistic, given that the measurement can only be based on existing secondary data originally produced for some other purposes than the research problem at hand [39]. In fact, the limitations of secondary data are determined by the choices made in relation to the production of the data, such as a decision to collect only easily available data and using fixed categorizations in the data production, with consequent variation in the production practices and categorizations used, and a lack of flexibility in the information systems to take into account changes in phenomena and society [40].
I have proposed that the methodological challenges related to the utilization of register-based data can be illustrated using ideas from communication theory [1]. A schematic diagram illustrating information communication via registers is presented in Figure 2. First, it is assumed that some phenomenon exists that can be observed. Since it is impossible to completely observe all details or perform exact measurements, some kind of coding is used to describe things. This coded signal is then stored in a database. The noise and bias can be interpreted as an explanation for measurement compromises, possible inconsistencies and coding errors, and coding practices existing in the stored signal. When this signal is then utilized, it must be decoded into a suitable form. This phase is also subject to noise and bias caused by incompatibility of choices and interpretations made by the data producer and the data user. Even the decoded signal (data) is not a final phase in the research process, because further analysis and processing is needed in order to transform the data into information. Even though this is a very simple and technical representation of communication, it seems to contain the essential elements needed in the common-sense understanding of secondary data.

2.3 Prerequisites in the register-based data analysis

The schemas in Figure 1 and 2 give a basic overview of the process of register-based data analysis, but leave many pragmatic details unanswered. Even though the actual realization of the process is determined by the research problem and the available register data, it has been pointed out that the effective use of register-data presumes skills in at least four areas: in computer science, in statistics, in the principles of measurement, and in the theory of the subject matter [41]. In the following, I briefly review and synthesize the most important ideas, assumptions and approaches that were prerequisites for the case study of this thesis.
2.3.1 Principles of measurement

Typically it is assumed that reality can be confronted by recording observations that reflect the phenomenon of interest. Measurement aims to create data as symbolic representations of the observations. The two main aspects for measurement are the representational and the practical approach [42]. Representational measurements are quantifications of attributes of objects existing in reality, while the practical measurements are operationalizations of the phenomena of interest. All representational measurements resulting in concrete data can be formulated in terms of a practical approach.

The operationalization used in the practical approach for measurement determines how the phenomenon $P$ that becomes visible via observations $O$ is mapped to data $D$. In the pragmatic sense, operationalization can be taken as successful if it becomes possible to make valid interpretations $I$ of symbolic data $D$ in regard to the phenomenon $P$, i.e. that a satisfactory saturation between the phenomenon, conceptualization and data is achieved [1]. The key point is that three recognized levels of contextual dependencies accompanying the empirical research are explicitly separated: a theory driven observation ($O$), the operationalized data ($D$), and a theoretical interpretation of the data ($I$) [43].

It is particularly important to notice that the actual information is not assumed to be contained in the data, but is something that has to be produced from the data and the pre-knowledge. Such an idea has been presented more formally in terms of an infological equation

$$ I = i(D,S,t), $$

which states that the information $I$ is produced from the data $D$ and the pre-knowledge $S$ by the interpretation process $i$ during time $t$ [44]. From this point of view it is obvious that any sharing of data can only be a proxy for the process of sharing of information, because the unbiased sharing of information would require the background knowledge $S$ to be identical with the producer and the users of data. In this sense it is not surprising that data as such are of no value and become interesting only if there is also a meaning and a context for them [45].

The range of possible interpretations of the data can be reduced by increasing the common background knowledge of the potential data users by offering descriptive data about data, that is, metadata [46]. While it is true that the increased amount of metadata increases the proportion of shared information, the more diverse contexts the data have to be usable in, the more work is required to disentangle the data from the context of its production [21].

These are also starting points in register-based data analysis, and the problem is to find some shared perspective between the original and intended data utilization purposes in terms of available data and (more or less tacit) metadata [3]. Following
the ideas of a cognitive fit approach [47], there could be a problem-solving task related to phenomenon \( P \), but only data \( D_P \) are available that tell about some other phenomenon \( P' \). A “true” task solution \( I_P \) can only be approximated with a conditional task solution \( I_{P|P'} \). The conditional task solution is based on conditional data \( D_{P|P'} \), which are a result of some intelligent transformation of available data \( D_P \) and are based on the cognitive fit between the internal representation of hypothetical “true” data \( D_P \) and available data \( D_P' \). In practice this means that an extra interpretation-operationalization phase is required in the research process if secondary data are to be utilized [1], i.e. that data in a register need to be decoded to become suitable research data (Figure 2: register \( \rightarrow \) decoding \( \rightarrow \) data). This important idea has been applied throughout this dissertation.

The problem with the approach is that phenomena are different in terms of measurement possibilities. As the law of medical information states, it is more difficult to utilize more complex variables for purposes other than those originally intended [21]. In this sense, it is in practice fruitful to evaluate the amount of subjectivity in the variables. One criterion is the separation between representational and practical measurement: the measurement is representational only if it is reasonable to assume that the data \( D \) are direct reflections of observations \( O \) that refer to some entity existing in reality [42], i.e. that the phenomenon of interest is directly observable. Another useful criterion is related to the amount of additional background knowledge \( S \) required in the interpretation of data \( D \). I have suggested that if there is no need for knowledge beyond the shared human common sense, the measurement can be characterized as stable [8]. In fact, it is known that space and time provide a shared biological-development basis for ordering human common sense as our world of sensory observations is arranged to observe things changing in space over time [48]. If measurement is representational and stable, it is factual. These ideas have an important role in this thesis, because the purpose has been to minimize unnecessary subjectivity by maximizing the use of factual data. For example, the hip fracture care episodes are reconstructed in terms of factual data.

### 2.3.2 Information science

A concrete data management requires some kind of data structures. In fact, the raw data are nothing but a sequence of bits stored in a database. There must be some a priori fixed rules for the interpretation and handling of bit sequences, which define the basic data objects, such as integers, floating point numbers and character strings. A data model specifies which kinds of data and data manipulations are permissible [49]. New types of data objects and associated operations can be easily constructed by combining the existing types and operations [50]. This idea makes it possible to provide a conceptual representation of data which hides storage and implementation details which are of no interest to most database users [29].
More generally, data modeling can be used to construct (computer-based) symbol structures which capture the meaning of data and organize it in ways that make it understandable and useful [51]. Data modeling is restricted in the sense that only what is (or can be) represented is considered to exist [52]. A recently introduced unified meta-information architecture of statistics (UMAS) provides a comprehensive conceptual framework for data modeling [53]. The aspects of the complicated framework that are useful in pragmatic register-based data analysis can be illustrated with the help of a pyramid framework that is based on the cognitive principles of how humans store everyday knowledge [48]. A slightly extended pyramid framework is presented in Figure 3. Following the ideas that I have presented independently [7], but which the UMAS also captures, the framework had to be complemented with an additional logical component.

A semantic object (conceptual entity), reflecting some phenomenon in reality, is a basic element in the framework. Taxonomy and partonomy reflect the principles of a cognitive categorization of the object and represent also the schemas needed to interpret the observational data. These parts form a knowledge component of the framework, i.e. reflect the required pre-knowledge that cannot be extracted from the data. The object must be logically defined by identifying an observable host for a concept and by determining which attributes related to the concept are to be observed. The logical component also includes theoretical measurement properties of variables given in terms of random variables that represent the (assumed) operationalization of the variables. The data component of the framework represents the actual data by means of three distinct data perspectives: place (where), time (when), and realized observation (what). The whole path from a knowledge component to the realized observation is needed to
describe what is actually observed. The (extended) pyramid framework provides a human–computer environment for (register-based) data analysis that incorporates models of human cognition so that the utilization of large databases becomes more effective than without the framework [54]. A generalized event-sequence approach that I have presented in article 1 is obviously a special case of this more general framework. The framework also had an important role in the development of the conceptual model in article 2.

2.3.3 Statistical computing

Information sciences also deal with more pragmatic tasks concerning the acquisition and preparation of data, which can also be classified as statistical tasks. The acquisition of relevant data in practice may be a difficult and time-consuming task, because the sources of data need first to be identified and the use of register-based individual-level data for scientific research requires permissions to be applied for [55]. Technically the problem is similar to the one of extracting and integrating data from multiple sources into a new data view (data warehouse) [56]. An important special case aims to identify the instances which include data about the same real-world entity [57]. Such duty is better known as record linkage [58]. The record linkage becomes problematic, however, if the entity identifiers in the data sources are not identical and probabilistic techniques are needed to predict the equivalence [59]. More generally, data integration may require schema matching, which aims to find compatible interpretations between multiple database structures [60]. For example, typical mortality data have a different structure to hospital discharge data, but in practice it is useful to define a data structure that allows incorporation of data from both sources.

There are also some commonly encountered data preparation tasks [1]. Data cleaning task involves detecting and removing errors and inconsistencies from data in order to improve the quality of data [61]. In statistics literature, data cleaning is known as editing and imputation [62], and in applied mathematics as error correction [63]. Data cleaning deals with logical errors, such as violations against integrity constraints or duplicate observations. The nature of data cleaning is technical, meaning that such corrective manipulation of data could be done no matter what the actual research problem under investigation is. Data reduction aims to produce a reduced representation of data which is much smaller in volume than the original data set, yet produces (almost) the same answers to the research problem [64]. This kind of data preparation can consist of anything from simple dropping of unimportant variables, combination of several variables or observations into single one, or more radical changes in aggregate levels to even more complicated analyses. For example, in some cases it may be reasonable to assume a multivariate normal distribution for variables of interest, which means
that only the covariance matrix and means (sufficient statistics) are needed for subsequent analyses. Data abstraction tries to embed an intelligent interpretation (enrichment) of “raw” data into analyses so that the resulting derived data set would be at the level of abstraction corresponding to the current problem [65]. In practice this kind of abstraction may be feasible by processing data with explicit algorithms so that important a priori schemas that are not directly available in data become formalized in terms of the data. For example, register-based data are usually patient-specific, while medical knowledge is patient-independent and consists of generalizations that apply across patients. A complication after a surgical operation is a medical concept, but from individual-based hospital discharge data it must be abstracted by using some rules, such as a list of particular diagnosis codes with appropriate time stamps recorded in the data. Data abstraction always results in a problem-specific derived data set.

In summary, the data preparation has been characterized as a process that requires human participation, which means that as much art as science is needed in good data preparation [66]. Whatever preprocessing tasks are applied, it is obvious that incorporated explanatory analyses offer insights and realistic perspectives into the data. It can be further stated that a sophisticated preprocessing—which is full of ideologically dependent qualitative choices—in order to scale matters down to a size more suitable for specific analyses is the most important and time-consuming part of register-based data analysis [1]. Data sensitive preprocessing had a key role in the empirical part of this thesis.

2.3.4 Statistics

Statistics is a science devoted to the production, analysis, modeling, and presentation of data. Statistical models help to distinguish systematic patterns from random fluctuations, measurement errors, and confounding biases occurring in the data. However, the traditional mathematically oriented statistical paradigm has had some difficulties in adapting to a new situation that requires concrete analyses of massive (register-based) datasets [22]. One reason is that the basic assumptions about the independent observations and the sampling error as the main source of uncertainty are often violated in complex register-based data sets which possibly include total populations [67]. Further, the traditional statistical significance has become as an issue, because even the practically unimportant differences easily become statistically significant within the large data sets. In addition, with massive datasets it can be expected that there are some distortions and errors in the data, and typically it is unfeasible to manually check all of them. On the other hand, the reasonable use of data typically requires the use of problem-specific data abstraction rules which transform the data to a useful form. This kind of data analysis is typical for “soft” qualitative analysis, but with massive register-based
datasets, strong technical skills in quantitative data management are also required. In summary, a flexible approach for statistical inference is typically beneficial with register-based data. Moreover, statistics offers not only a set of tools for problem-solving, but also a formal way of thinking about the modeling of the actual problem [68]. In this sense, statistical thinking also has a key role in the process of register-based data analysis.

In fact, one useful perspective for register-based data analysis is to consider the whole process of information communication (Figure 2), and try to explicate the sources of noise and bias that may distort the communication [69]. The basic idea is—in addition to a traditional modeling of the phenomenon of interest—to take into account suspected sources of bias, possibly with the help of a priori information [70]. Such is one way to deal with the particular restrictions of secondary data by using theory driven statistical modeling. Chapter 7 in this thesis also incorporates this kind of statistical thinking.

Another option for statistical modeling is to apply statistical algorithms directly to data and let the data speak for themselves [71]. Also this approach is often useful with massive secondary data sets, because data exploration and description plays a more general role than it does in the case of small data sets. The screening of data for significant associations without having specific hypotheses may be questionable [72], but it may also be erroneous to assume that data are automatically in concordance with the theory. Within this approach it becomes possible to be data sensitive so that empirical justification is given to the proposed operationalizations. Most analyses in this thesis are of this type.

2.3.5 Theory

Theory of the subject matter is needed for the formulation of the research problem. As shown in Figure 1, theory is also the driving force in the generation of the question and in the choice of the perspective, as well as in the interpretation of the data and analyses. In other words, the theory determines the framework within which justified data analysis becomes possible.

The special feature in register-based data analysis is that more than one theory must be simultaneously dealt with [3]. For meaningful results there is a need to find some communality in terms of realized data between problem-driven subject matter theory and the (more or less unknown) theories that have been used during the production of the register-based data. The extended pyramid framework helps to structure the data in a concrete and methodologically sound way that makes the finding of suitable compromises much easier. For example, the components of data that have the most stable measurement properties are obviously the ones that can be most easily reused in various contexts and the basics of many useful theoretical approaches can be built on these factual main elements.
3 Register-based health system performance assessment

In the previous chapter it was argued that the register-based data analysis should be considered as a research process. Obviously that is true also in the case of scientific health system performance assessment. In this chapter some essential background information needed for the understanding of the phenomena of interest are given. The applied research problem is also introduced and the perspective is fixed with certain focusing assumptions.

3.1 Finnish health system

A (public) health system is a creation of the human community, and the health system of any society can only be understood in the light of its societal operating principles and policies [73]. The Finnish societal system is typically characterized as a Nordic welfare state, which is internationally rather exceptional. The Finnish health care system is very decentralized, and the country’s numerous (>400) local authorities (municipalities) are responsible for arranging services [74]. Each municipality is a member of one of the 21 hospital district joint authorities, which are responsible for organizing specialized medical services and coordinating hospital treatment in its own district. Secondary and tertiary level medical care is provided by a hierarchy of regional, central and university teaching hospitals. Services for older people are provided in both social and health care, both being incorporated into the same national planning and financing system [75]. The organization and financing of health care has long been considered a public responsibility [76]. There are also many recent reforms to the Finnish health system, such as the setting of thresholds for admission onto waiting lists for elective surgical procedures, the introduction of a set of maximum waiting-time targets for non-urgent examinations and treatments, a national electronic patient record and a project aiming to restructure municipalities and services [77].
3.2 Information production in the Finnish context

One important tool in the maintenance of a Nordic welfare state has been to produce information which supports the (nationwide) decision-making. The main emphasis has been on producing official statistics and indicators that reflect important aspects of health and social welfare, and therefore offer information for decision-making and controlling purposes. However, the national-level control mechanisms were decomposed in 1993 due to the change from direct authoritative control (imperatives, rules, norms, earmarked funds) into indirect control based on information (instructions, guidelines, hopes, total funds) [78]. This new nationwide policy on “steering by information” has aimed to produce information for meso-level decision-makers (such as municipalities and hospital districts) [79]. The problem is that now there should be information to suit the purposes of hundreds of local policies instead of one global policy as in the past [80]. Moreover, the potential users of information may easily refuse to act on the given information if they consider it to be irrelevant for their purposes [81]. In this sense, there is an obvious need for methodological studies that aim to make policy relevant information production more efficient.

3.3 Measurement of the performance of a health system

Even though there does not exist a universal value base to all health care systems, most health systems in developed countries aim to promote, restore and maintain health [82]. The common goal is the optimization of the health of individual patients and populations in an equitable, efficient, and effective manner that is acceptable to patients, providers and administrators [83]. The growing need for appropriate services with limited resources and the concern about the continuing inequities in health and in access to health care means that there is much interest among decision-makers in improving the performance of health systems [84]. However, performance measurement has proved to be very difficult in practice [85]. The main problem in performance measurement seems to be that there is little agreement on the philosophy of measurement and on what to measure [86]. For example, at least 15 dimensions for health care performance can be identified from the existing frameworks [87].

In fact, the problem seems to be related to a more general change in societies and in information production [88]. The evaluation of performance has become a more powerful organizing concept in activities of societies and there is a real need for information about performance. However, it has been claimed that the
abstract rhetoric of administration blurs the actual meaning of performance: there are no more norms about what and how things should be done, but the political responsibility is implicitly included by using vague formulations of strategic objectives which delegate the actual implementation to “actors” [89]. In this sense, performance assessment as such is an empty concept [88]. Only if the goals are fixed and operationalized to a measurable form can the concept of performance become interesting and potentially useful.

Concrete quantitative measurement of performance has mainly been conducted using various performance indicators [90–92]. Recently the whole idea of performance measurement seems to be focused on the sophisticated development and reporting of indicators [93-95]. Furthermore, methods for evidence synthesis and assessment in the context of multiple indicators have been suggested [96, 97]. As the indicators certainly reflect something that can be measured, it is typically useful to compare indicators between the appropriate subpopulations, because descriptive data on relative performance may help to identify “best practices” or “what works”. However, the causal mechanism underlying differences between the indicators typically remains unknown in such an approach [98], which “invalidates” the appealing analogy to the experimental research design [99]. In this sense, register-based indicators without detailed justification of their theoretical-methodological basis do not represent scientifically valid information.

3.4 Assumptions for the production of concrete health system performance information

In this thesis, I focus on the methodology of producing health system performance information. The key assumption is that the routinely collected register data are to be used for performance assessment purposes. The use of register-data in health research is known to be prone to several problems, such as the perceived lack of value of administrative data, privacy and confidentiality, data availability, population coverage, registration period, record linkage possibilities, lack of clinical data, data format, coding systems, coding practices, completeness of registration, accuracy of registered data, data processing, size of data, and discovery of chance occurrences [100–104]. I do not claim that the problems are not real, but adopt the view that proponents of certain approaches have been more interested in advocating their ready-made mechanical procedures than in understanding alternative logics of interpretation [105]. In fact, many of the mentioned problems are related to particular study designs or properties of existing data sets and therefore relevant only for certain types of research questions or data. For the most fruitful results it seems to be essential to elaborate the whole analytic process with respect to the data, because the fixed structure of secondary data indirectly determines constraints on
3 Register-based health system performance assessment

the theoretical constructs [106]. In this sense, the methodological approach for register-based data analysis presented in Chapter 2 also provides a basic framework for the register-based performance assessment of the health system. I focus on the data sensitive analyzing approach, and the aim is a step-by-step transformation of data into as objective information as possible by maximizing the use of the most factual register data.

I assume that the interesting dimensions of health system performance assessment can be examined by using the framework which conceptualizes health system performance assessment in terms of structure, process and outcomes [107]. In short, the structure comprises resources that are devoted to producing actions whose primary purpose is to improve health, while the process means the realized utilization of these resources, and the outcomes reflect the effects of resource utilization on health. More recent formulations of the framework emphasize also the importance of a context and health policy goals that have led to certain health system implementations [10, 108].

The definition of what kind of expectations the commonly accepted abstract health policy goals truly reflect is difficult [109]. However, it can be defined that effectiveness in this context measures how the health system (or resources spent on it) affects the health of the target population [10]. In this thesis I concentrate on the effectiveness dimension of health system performance, because effectiveness seems to be a key dimension in several performance frameworks [87]. It is also known that the production of health benefits (effectiveness) plays a central role in assessing the cost of producing health benefits (efficiency) as well as the distribution of these benefits and costs across groups (equity) [10]. In addition, there is an acute need for information on effectiveness for the purposes of information steering in Finland [110].

3.4.1 Register-based health system performance monitoring

The Finnish registers offer an exceptional coverage of health and social welfare data [18]. In principle, a separate scientific study utilizing these data could be conducted to answer some specific question. However, from the information steering point of view, the available register-based data should be routinely converted to useful information. As it is unfeasible to repeat separate studies manually over and over again, a better alternative is to incorporate new information production pipelines to the health information system that can be used in the production of routine statistics [14]. In fact, the implementation of a nationwide health system performance monitoring system is one of the main development goals of for the information systems of the health and social welfare services in the near future [111].
In the Finnish proposals for the health information system, register data (raw data) and evidence-based decision support (know-how of transforming the raw data into information about issues that should be known) are essential ingredients for the continuously improving epidemiological and quality components that routinely produce new scientifically valid information [14, 15, 112]. Following these proposals, health system performance monitoring should consist of two components. The epidemiological component is needed for monitoring the volume and incidence of health problems, which are essential information for the evaluation of prevention strategies and which also help to prepare for changes in the need for care caused by changes in the population structure. The quality component concentrates on producing information on the effectiveness and quality of treatment, which are essential information for the evaluation of health system performance and for the purposes of finding ways to improve the health system. For the purposes of this thesis, I assume that pragmatically useful health system performance monitoring should be able to produce valid information about the volume and incidence of the problems, use (and costs) of services as well as about the effectiveness (and quality) of provided services.

It is well known that the performance assessment is practically feasible in a reasonable way only if it is carefully developed and tailored for each specific health problem separately [106, 113]. Therefore, I concentrate only on one health problem, hip fracture.
Register-based data on hip fractures – sources and validity

The assumptions and theoretical considerations in the chapters above create adequate starting points for more concrete health system performance assessment. The case study in this dissertation is intended to offer perspectives and methods for the implementation of a nationwide system for monitoring health system performance. Hip fracture is used as a concrete example. A brief review of essential background knowledge is given first. Thereafter, the feasibility of using Finnish register data in the case of hip fracture treatment performance assessment is discussed based on article 2.

4.1 Hip fracture

Hip fractures are common injuries among older people, and associated with substantial morbidity and mortality [114]. The term hip fracture refers to a fracture of the upper end of the thigh bone (femur). Most hip fractures in persons aged 50 years and over result from moderate low-energy trauma, usually a fall from a standing height or lower [115]. For younger persons it is more likely that a case of hip fracture results from a high-energy trauma, such as traffic accidents or a fall from a height [116]. Prevention has focused on minimizing the risk of falls and on reducing the injury potential of those falls [117]. About 7000 (of which more than 95% occur for patients aged 50 years and over) hip fractures per year occur in Finland currently [5].

Ageing among populations is increasing hip fracture patients’ mean age and the number and severity of their pre-existing co-morbidities, which is likely to cause additional problems in patients’ treatment and rehabilitation in the future [118]. Sometimes the hip fracture can be interpreted as an indication of the “beginning of the end” (patients were doing well until they broke a hip and went downhill quickly) and sometimes as an “end of the beginning” (hip fracture signals that the cumulative effect of small declines has reached a critical level) [119].

4.1.1 Hip fracture treatment in Finland

Virtually all suspected hip fracture patients are first referred for examination and treatment to the nearest hospital with orthopedic services. The main objective in hip fracture treatment is to return the patient to his or her level of function before
the fracture [115]. The diagnosis of fracture of the hip is straightforward, using x-ray examination. A surgical operation is performed on the majority of patients. The main methods used in treatment are reduction of the fracture using internal fixation and hip replacement arthroplasty. The care pathway for a hip fracture patient is rather complex with several phases such as surgical management and rehabilitation [120], and is known to result in diverse episode profiles in Finland [121]. Typically a patient is transferred for rehabilitation to the health center serving the patient’s resident municipality after a short postoperative hospital treatment [122]. Finnish health centers are local primary health care units, which also contain inpatient wards. Other institutional environments of care include residential homes and service housing with 24-hour assistance, which both correspond to the nursing home type of care. Non-institutional services utilized by hip fracture patients include outpatient health services, home nursing, ordinary service housing, home-help services, and support for informal care [75].

For six Finnish hospitals, patients aged 50 and over had an average mortality at 30 days after the fracture of about 7%, 17% at four months, 26% at one year, and about 50% at four years [123]. At four months, about 40% of patients lived at home, about 15% were unable to walk, and about 8% had a lot of pain in the injured hip [123]. The functional capacity of the patients does not typically revert to the level prior to the fracture [124]. Hip fractures are also costly to the society. The average patient-specific costs during the first post-fracture year in Finland were estimated to be around €14 410 and more than €35 000 in case of a previously home-dwelling individual who becomes a long-term care patient following the fracture [125].

Treatment processes as well as the outcomes vary considerably between areas and hospitals, and improved auditing of hip fracture treatment has been suggested [126, 127]. A recent Finnish current care guideline on the management of hip fracture patients proposes that a nationwide hip fracture register allowing continuous auditing should be established in Finland [128]. In this sense, there is a pragmatic justification for the methodological studies aimed at transforming routinely collected register data into relevant hip-fracture-specific information about the performance of the health system. In addition, hip fracture is a good choice for a pilot study on performance assessment, because it can also be viewed as a tracer condition in health systems, testing how well health and social services are integrated in the provision of acute care, rehabilitation, and continuing support for a large and vulnerable group of patients [129].

4.1.2 Register data on hip fractures in Finland

Finland has a long history of collecting data on health and social services. At the end of 20th century, there were about twenty different routinely collected national administrative health registers [18, 55]. A unique national personal identification
number is used in all Finnish registers which can be utilized as a linkage key in order to combine data from multiple sources. A particularly important register for the purposes of health services research is the Finnish Health and Social Welfare Care Registers (including the Finnish hospital discharge register). These registers contain data on all inpatient care periods in hospitals, in health centers, in residential homes and in service housing with 24-hour assistance. The data warehouse of the Finnish Hospital Benchmarking Project (nowadays a part of the Finnish Health Care Register) is the corresponding register for hospitals, but also incorporates data on outpatient visits. In general, the complete registration combined with easily linkable registers makes large, longitudinal population-based studies feasible in Finland.

4.1.3 Previous register-based hip fracture studies in Finland

The first nationwide register data based on hospital discharges in Finland are available for the year 1960, although the continuous hospital discharge data collection began no earlier than 1967. Hip fractures were not reported separately, but were combined with other fractures of limbs [130]. The register data from 1968 included hip fractures as a separate group [131]. The hospital discharge data from 1968 is also a data source for the first hip-fracture-specific register-based study in Finland [132]. Most of the register-based studies concerning hip fractures have had an epidemiological perspective [133–137]. In the late 1980s and early 1990s, a project aimed at improving the reporting of treatment and costs data based on registers had hip fractures as a separate group [138]. At the same time, the first register-based small-area analyses examining treatment practices in terms of surgical operations reported data on hip fractures treated with arthroplasty [139]. The next step in the utilization of register-based data on the description of hip fracture treatment practices was the reconstruction of the care episodes of the patients [121, 140]. There followed a pilot study examining the effectiveness of hip fracture treatment using register-based data [141, 142].

4.1.4 Register data for the current study

The data used in this study were also based on Finnish registers. As virtually all hip fracture patients are treated at the hospital inpatient ward and given that hip fracture is easy to diagnose, it is very likely that the patient population can be easily identified from the hospital discharge data by using a simple database query with a list of diagnosis codes. To make sure that all hip fracture patients were included in the study population, a total population of fracture of femur (corresponding to
ICD-10: S72) in the period 1998–2002 was identified in the Finnish Health Care register.

In order to capture the medical histories of these patients, data on all inpatient (1987–2002) and outpatient hospital care (1998–2002), residential home care and care in service houses with 24-hour assistance (1997–2002), and deaths for this population were obtained from the Finnish hospital discharge register, the Finnish Health and Social Welfare Care Register, the data warehouse of the Finnish Hospital Benchmarking Project and the National Causes of Death statistics using the unique national identification numbers of the patient population. It was first time that the hospital outpatient visits from the benchmarking project data and the data on inpatient care in social institutions from the Social Welfare Care Register were included in a hip fracture specific register-study.

The results of simple database queries were integrated into a new data set containing 988,762 records for 39,041 patients. Each record in this data set corresponded to one care period in inpatient institutional care or outpatient visit in hospital (or death), and included variables such as patient and provider ID-numbers, age, sex, area codes, and diagnosis and operation codes, as well as dates of admission, operation and discharge (or death).

4.2 Validity of register data in the case of hip fracture

Valid data is a prerequisite for any information system aiming to produce practically useful information on health system performance. Several studies have reported a high completeness of registration and a high validity for most variables in the Finnish registers [143–146]. However, there have been some suspicions about the quality of register data in the case of hip fracture [147], and recent a Finnish current care guideline has requested more validity analyses [128]. Moreover, there have not been studies where the validity of data is judged against the health system performance assessment purposes.

Therefore, an investigation aimed at clarifying these issues was conducted [2]. Prospectively collected data for 106 consecutive hip fracture cases in the Kuusankoski Regional Hospital between January 1, 1999 and January 31, 2000 and the register data for patients treated during the same period in the surgical ward of the Kuusankoski Regional Hospital were compared and the observed differences were checked from the medical records.
4.2.1 Completeness of registration

It was rather straightforward to check whether the patients in the register data having a recorded hip fracture diagnosis (ICD-10: S72.0, S72.1, S72.2) in a particular hospital period were found in the prospective data and vice versa. As it was very likely that all hip fracture patients were identified from the register or prospective data, and as the hip fracture diagnosis could be confirmed from the medical records, it was possible to evaluate whether all true hip fracture cases were identifiable from the data sources. It turned out that there were in total 105 patients with a confirmed hip fracture diagnosis. The audit data missed one patient and the register data two patients. That indicated very good completeness of registration in both data sources. Three extra hip fracture candidates clearly had a false positive diagnosis in the register, and a further five candidates had been excluded from the prospective data on the basis of additional information, i.e. register data may slightly overestimate the number of fresh hip fractures unless appropriate data abstraction rules are used for the exclusion of extra cases. In summary, this analysis confirmed that the validity of the data seems to be acceptable for simple incidence monitoring purposes.

4.2.2 Conceptual model for hip fracture performance monitoring

It was more difficult to evaluate the quality of register data for performance assessment purposes, because data validity must be judged against the intended utilization purposes [148]. However, there is no obvious golden-standard that would have told which aspects and issues of performance monitoring should have been evaluated between the data sources [149]. In addition, it was obvious that the quality criteria could change with time. For instance, the detailed diagnosis of hip fracture is more important in the operating room than during the final stages of rehabilitation, i.e. the requirements for correct diagnosis in data may be more important in a surgical ward than in a nursing home. Therefore, in order to make any statements concerning the quality of data, it was essential to first outline the properties of required measures carefully under a fixed conceptual model for performance assessment (determine how data result from the theory), and then evaluate the observed data against these requirements (examine how well the theory can be reconstructed using the properties of actual observed data). The same idea was expressed more explicitly in chapter 2.3.1.

In the case of hip fracture treatment, the importance of practical and reliable information on evidence-based performance assessment has been recognized [126], and good experiences of the Swedish Rikshöst-registry have led to the formalization of data production procedures for the purposes of a standardized audit of hip
fractures in Europe (SAHFE) [127]. The suggested prospective collection of clinical data represents the practical “state-of-the-art” consensus for data requirements for assessing the treatment of hip fracture. This work has produced a list of SAHFE-variables which reflect the operationalizations of patient-level concepts that have been considered important in the case of hip fracture. However, the more or less randomly ordered list of variables is not a useful conceptual model.

In order to define a conceptual model, the SAHFE-variables were chosen to represent an adequate base for data requirements in terms of patient level concepts. The extended pyramid framework presented in Chapter 2 was used in the systematic characterization of these concepts. Each phenomenon represented by a concept was assumed to be attributable to an individual on a continuous time scale. It was also assumed that each concept can be described using a systems approach with a (limited or unlimited) number of states so that the system is always in some of these states at any time point. All concepts were classified on the basis of a subjective evaluation of potentially suitable measurement scales and types of statistical distributions (theoretical measurement properties). The second classification dimension was the temporal stability of the concept, i.e. it was considered, for example, at what time intervals the concept makes sense and how often the concept should be measured. In addition to these statistical classification principles, a contextual interpretation was linked to each concept in order to make sure that the resulting operationalization offers desirable insights on the phenomenon of interest. The linkage was mainly based on the conceptualization used in SAHFE-metadata, but it was complemented using theoretical models of performance assessment [10, 150, 151]. After this initial classification phase, the concepts were further classified into more general groups in terms of their contextual interpretation, temporal stability and theoretical measurement properties. In the third phase selected measures from the SAHFE-list, from the prospective Kuusankoski data, and from the register data were mapped to the conceptual model and the goodness of the fit in terms of comprehensibility and simplicity of the model was evaluated. These three qualitative phases were repeated until the resulting conceptual model was considered satisfactory.

The final conceptual model presented in Table 1 gave a pragmatic mapping between concepts from a performance assessment theory, SAHFE metadata and structured data entry in prospective data and secondary register-based data in terms of contextual interpretation, temporal stability and theoretical measurement properties. Of the model dimensions, the biological constants do not change in time and therefore one measurement is generalizable to all times. Biological events, accident/fall history, and hip fracture event represent dimensions for which actual values of measures are recorded in the proximity of some observable event. All other dimensions relate to phenomena that potentially change in time and should be continuously monitored. In practice continuous monitoring is impossible, and actual measurements of continuously changing phenomena can be made only
Table 1. Conceptual model dimensions and their realizations in different data sources

<table>
<thead>
<tr>
<th>Dimension</th>
<th>SAHFE measure</th>
<th>Prospective data measure</th>
<th>Register data measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological facts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biological events</td>
<td>date of birth, death, year of menopause, menarche</td>
<td>date of birth and death</td>
<td>date of birth and death</td>
</tr>
<tr>
<td>Biological constants (DNA)</td>
<td>sex</td>
<td></td>
<td>sex</td>
</tr>
<tr>
<td>Biological measures</td>
<td>height, weight hemoglobin, creatine, albumin, bone density</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographic history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household composition</td>
<td>living alone</td>
<td>marital status, living alone</td>
<td>area of living, level of care</td>
</tr>
<tr>
<td>Place of living</td>
<td>residential status</td>
<td>residential status, area of living</td>
<td></td>
</tr>
<tr>
<td>Socioeconomic history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education, occupation, economic resources</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health related behavior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk behavior</td>
<td>smoking, alcohol intake</td>
<td>alcohol usage</td>
<td>alcohol-related inpatient care</td>
</tr>
<tr>
<td>Diet, activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective quality of life</td>
<td>pain, psychological state, fear of fall</td>
<td>pain</td>
<td></td>
</tr>
<tr>
<td>Objective need for care</td>
<td>comorbidity, complications walking</td>
<td>complications walking, activities of daily living</td>
<td>recorded diagnoses, use of care</td>
</tr>
<tr>
<td>Diseases/symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive functioning</td>
<td>walking</td>
<td>abbreviated mental test score</td>
<td></td>
</tr>
<tr>
<td>Use of care</td>
<td>walking aids</td>
<td>use of assistive technology, changes in living environment</td>
<td></td>
</tr>
<tr>
<td>Technical aid</td>
<td>provider, type of stay, length of stay</td>
<td>use of painkillers, type of care, rehabilitation,</td>
<td></td>
</tr>
<tr>
<td>Medication</td>
<td>provider, type of stay</td>
<td>dates for hospital visits</td>
<td></td>
</tr>
<tr>
<td>Health care utilization</td>
<td>provider, type of care, admission and discharge</td>
<td>provider, type of care, admission and discharge</td>
<td></td>
</tr>
<tr>
<td>Accident/fall history</td>
<td>date of fall, place of fall</td>
<td>date of fall, place of fall</td>
<td>place of and reason for hospitalized injury</td>
</tr>
<tr>
<td>Properties of fall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hip fracture event</td>
<td>time of fracture, occurrence place, fracture type, fracture side</td>
<td>day of fracture, reason for fracture, occurrence place, fracture type, fracture side</td>
<td>place of and reason for hospitalized injury</td>
</tr>
<tr>
<td>Properties of fracture</td>
<td>hospital ID, admission time, admission source, start of operation, reason for operative delay, type of surgeon and anesthetist, type of anesthetic, operative method, length of surgery, type of prophylaxis, complications in operation, reoperations, time to mobilization, discharge time, discharge destination</td>
<td>admission day, admission source, day of operation, reason for operative delay, operative method, complications in operation, reoperations, discharge day, discharge destination</td>
<td>hospital ID, admission day, admission source, diagnoses, day of operation, operation codes, discharge day, discharge destination</td>
</tr>
<tr>
<td>Initial treatment process</td>
<td>day of fracture, reason for fracture, occurrence place, fracture type, fracture side</td>
<td>day of fracture, reason for fracture, occurrence place, fracture type, fracture side</td>
<td>day of fracture, reason for fracture, occurrence place, fracture type, fracture side</td>
</tr>
<tr>
<td>Clinical stability</td>
<td>ASA grade, comorbidities, coexisting fractures, reason for operative delay in medically unfit patients</td>
<td>ASA grade, reason for operative delay in medically unfit patients</td>
<td>recorded diagnoses, use of care</td>
</tr>
</tbody>
</table>

1) Indirect measurement requiring record linkage.
at selected time points. In the prospective data, the measurement was done in connection with a hip fracture event and for certain measures also at two weeks, at four months, and one year following the fracture. In the register data, the recording took place at each discharge.

The constructed conceptual model explicitly revealed commonalities between the prospective and register data. As expected, the content of prospective data was richer in clinical detail than the register data, even after using indirect measures of certain concepts. The conceptual model also showed which dimensions were comparable between the available data sources. In the actual empirical analyses these comparable dimensions were examined by using the available complete prospective data for 106 hip fractures and corresponding register-based data.

4.2.3 Consistency between prospective and register-based data

Three types of comparisons between the data sources were used depending on the theoretical measurement properties of the variables. In the type I comparisons, the same rather easily measurable variable was available in both data sources, and detected differences could be further checked from the medical records of the patients. The percentage of cases with a recorded value for the variable (completeness) and proportion of correctly recorded values for cases with some recorded value (accuracy) were calculated for such variables. Completeness and accuracy were very good for age, sex, area of living, admission and discharge information as well as for hip fracture diagnosis and operation day and type. Side of fracture was seldom recorded and accuracy of information on place of accident was poor in the register.

In the type II comparisons, conceptually similar measures which had substantially different operationalizations in both data sources were examined. From this category, follow-up information was a particularly important issue for performance assessment purposes. Using the register-based data it was straightforward to extract information on the use of inpatient health services and mortality on a daily basis. The derivation of corresponding information from the prospective data was much more complicated and required extra assumptions and interpolation. The overall follow-up patterns were very similar indicating that the register-based data suits better for the follow-up of use of inpatient health services than prospective data. However, some extra work needed for the prospective collection of follow-up data may be useful in order to produce data on issues that are not available in the register-based data, but offer clinically interesting follow-up information, such as the use of technical aids, pain, and validated measures of functional disability and quality of life.
Type III comparisons were actually a mixture of type I and type II comparisons. This approach was used to evaluate the more detailed diagnosis of hip fracture. Using a detailed hip fracture classification of femoral neck fracture, trochanteric fracture or subtrochanteric fracture, the overall agreement between prospective and register data—calculated as a sum of equivalent classifications in data sources divided by sample size—was 86.3%. The 95% confidence interval for overall agreement—calculated with a nonparametric bootstrapping technique [152]—was 79.4–92.2%. The proportion of agreements that were not due to chance was measured using a delta coefficient, which is based on the model of multiple-choice tests and avoids certain known deficiencies of the traditionally used kappa coefficient [153]. Kappa may be considered as an approximation of delta, as both yield values that are very similar when at least one marginal distribution is balanced or when both marginal distributions are moderately unbalanced in the same direction [154]. According to the estimated delta coefficient a high proportion 81.3% (95% CI: 71.3–91.3%) of agreements were not due to chance. The delta coefficient was somewhat higher than the kappa coefficient of 73.5% (95% CI: 61.2–85.7%) indicating that the assumptions of kappa were not satisfied. In addition, the accuracy of diagnoses was addressed by considering bone anatomy as a biological entity identifying true fracture status [155]. This made it possible to assume that the trait underlying different fracture types was continuous, and a simple latent trait model in the form of a polychoric correlation coefficient was estimated [156], yielding a value of 0.92 (95% CI: 0.86–0.98) which also reflected a very high consistency between the register and prospective data.

4.2.4 Conclusions on validity

In order to analyze the quality of register data for the purposes of performance assessment, a conceptual model had to be constructed. There were several novel aspects in the model. The statistical properties of the concepts had a very important role in the classifications and the special features of different data sources were simultaneously taken into account during the construction process. As a result, the presented conceptual model revealed the most important assumptions affecting the actual results of analyses, and showed how it was possible to produce a pragmatically useful mapping between problem-oriented and data-driven theories.

As expected, the register data lacked clinical detail, but had a data structure that allowed the complete observation of inpatient care history, and therefore outperformed prospective data in this respect. The empirical validity analyses revealed that the completeness of registration was very good, but appropriate data abstraction rules should be used in order to avoid overestimation of hip fracture cases. Also the accuracy of the most easily measurable variables and the quality of follow-up information in the register was found to be very good. In addition,
detailed hip fracture classification showed a high consistency between the register and prospective data. As a conclusion, register-based data seemed to be a valid data source for the purposes of the performance assessment of hip fracture treatment.

However, there was room for improvements in the register. An inflexible data structure for the performed operations causes problems if a patient has several operations during a single hospitalization. There was also a need for additional data, while especially the validated functional disability or quality of life measure would have been useful. The content of these register data could also be improved by giving standardized instructions that guide practitioners to record the relevant additional operation codes available in the classification, such as on which side the fracture was. Also a voluntary input of additional hip fracture event data to the centralized register should be made possible, because the best option for practical performance assessment purposes would be to combine SAHFE-type of data with the register data.
5 HIP FRACTURE INCIDENCE

The epidemiological component of the hip fracture surveillance system has an important role, because without continuous monitoring of hip fracture incidence it would be difficult to assess the effectiveness of prevention interventions and also to prepare for changes in the need of care caused by changes in the population structure. This chapter is based on articles 3–5.

In principle, the Finnish register data are well suited to the hip fracture incidence calculation. Hospital admission and discharge days are easily observable facts (i.e. have stable measurement properties in the terminology presented in chapter 2), which are correctly and completely recorded in the Finnish data. However, in practice it is not straightforward to determine the true number of hip fracture events from the register data. In fact, there are several recognized potential pitfalls in using hospital discharge data for calculating the number of injuries [157, 158].

Even though all hip fractures are identifiable from the hospital discharge data (virtually all hip fracture patients require inpatient hospital treatment and hip fracture is easy to diagnose), there are several discussable details, such as the exact diagnosis codes to be used in the identification, the use of secondary diagnoses in the hip fracture identification, and exclusion of certain nonstandard cases. However, from the methodological perspective a much more important source of bias (which is independent of the use of diagnosis codes) is related to the detection of multiple hospitalizations of a single patient. At the conceptual level this means that all events related to the same underlying disorder should be recognized. A so-called care episode approach provides a sound methodological framework for dealing with multiple hospitalizations in terms of discharge data [159]. In the case of hip fracture, it is known that a typical treatment process starts with an emergency admission to a hospital after the hip fracture event followed by an operation at the surgical ward and continues with rehabilitation at other inpatient wards.

With a careful development of data abstractions rules to exclude multiple hospitalizations it was possible to demonstrate that several previous Finnish studies on hip fracture incidence had overestimated the numbers of hip fractures [5]. More specifically, the decision to use the calendar year boundaries to combine multiple hospitalizations in these studies was artificial and had no epidemiological justification, because the underlying disorder had nothing to do with the beginning or end of the year. In fact, it was likely that many patients having their fracture during the final months of each year were erroneously counted as separate cases for two consecutive years [3].

However, these considerations, combined with the problem that even the clinical criterion of a fresh hip fracture may vary, point out that it is impossible to
obtain a unique estimate for the number of hip fractures using register-based data. This ambiguity reflects the undesirable incompatibility between the theoretical model and data. As it was not possible to change data, there was a need for an alternative theoretical interpretation that would be more compatible with the restrictions of data.

5.1 Aging-related hip fractures

It is known that hip fracture incidence increases exponentially with age [160]. This kind of population level functional dependency means that the condition is related to aging [161]. Biological aging can be seen as a complex process occurring stochastically in organs and tissues, which results in irreversible damage accumulation and vulnerability to the failures in maintaining the integrity of tissues and organs [162], and the components of a biological system are prone to aging-related failures in the system [163]. Most aging-related conditions are closely connected to alterations in certain tissue or organ, although a well-known exception is death, which is a result of fatal failure(s) in any vital components of human body.

Following this line of thought, it became obvious that a hip fracture was also a special case in aging-related conditions. In fact, hip fracture is not just an aging-related failure in some tissue or organ, but typically a result of an accidental event (such as fall). Cumulative damage in some tissues (such as bone and muscle) increases the probability of (accidental or pathological) fracture, but it is also likely that some aging-related conditions increase the risk of accident. In this sense, hip fracture seemed to capture the effects of non-fatal failures in components of the human body, which most likely appear in the form of some kind of disabilities. Therefore, by drawing a methodological analogy to the event of death, it could be hypothesized that the event of hip fracture actually represents a multidimensional disability event (“disability death”).

From this point of view, it was clear that the occurrence of first aging-related (low energetic) hip fracture would be the most interesting one, because it gives the upper limit for the time of development of critical hip fracture risk factors (for a hip-fracture-related chronic disability condition). Therefore, the actual problem of identifying the interesting hip fracture episodes could be transformed to the identification of the first aging-related hip fractures of the patients. This theoretically justifiable approach turned out to be in better concordance with the register data than the identification of all hip fracture events [3]. In terms of chapter 2, an extra interpretation–operationalization phase was needed.
5.1.1 A method for identifying a first aging-related hip fracture

The pragmatic problem in identifying the first hip fractures was that only a limited backward follow-up was possible with the register data, i.e. some fractures could have occurred outside the observation window. A straightforward solution would have been to use all available data for the determination of first hip fractures, but it was considered more convenient to use the same “hip fracture free” clearance period, that is shorter than the minimum backward follow-up time in the data, for all individuals.

In order to justify the length of clearance period, it was assumed that there may be a cut-off point after which the probability of having a new hip fracture is reduced to the level it would be even without a preceding hip fracture, i.e. that the preceding hip fracture would be an “uninformative” predictor of the new hip fracture [3]. In the concrete analyses, the first admissions (index points) with the diagnosis of hip fracture in 1998–2002 were identified for each patient. For each index point, backward time to the previous hip fracture admission of the same patient was calculated. Time was measured in months. There were follow-up data from the period 1987–2002, so the minimum (backward) follow-up time was 11 years. If there was no previous hip fracture admission, the observation was considered to be censored at the beginning of 1987.

The discrete hazard probabilities of the form

\[ P(T = t \mid T \geq t) \]

reflecting the conditional probability for an event occurrence at time \( t \) on condition that no event had occurred until \( t \), were estimated for each follow-up time \( 0 < t \leq 160 \) as

\[ h(t) = \frac{c(t)}{n(t)}, \]

where \( c(t) \) was the number of events at time \( t \) and \( n(t) \) was the number of persons at risk at time \( t \). To diminish the degree of random variation, estimated hazard rates were smoothed using a simple moving average technique with polynomial weights [164]. The smoothed hazard rates were referred as observed probabilities.

For the calculation of corresponding probabilities in the case that no preceding hip fractures had occurred (expected probabilities), there was first a need for reasonable estimates for the incidence of first hip fractures. It was assumed that clearance periods between one and ten years are reasonable a priori, and that the average incidence between the years 1998 and 2002 can be interpolated to the whole observation period. After the preparation of risk population data (described
below in chapter 5.3) it became possible to calculate age- and sex-group specific incidence rates by using standard methods [165]. Sexes were analyzed separately and clearance periods of one and ten years were used. A logarithmic transformation was made for the estimated age-group specific incidence rates, which linearized the exponential increase of rates with age. After that, a linear regression model was fitted to the logarithmic incidence rates so that an interpolated incidence rate could be predicted for any age by using the estimated model [166]. For ages below 40 years, a constant incidence rate was used in the prediction, because an assumption of the exponential increase in incidence with age was not satisfied for young ages [167].

As age and sex were known for each index period identified from the data, it was possible to subtract one month from each age and then predict the expected probability of a hip fracture after a one-month (backward) follow-up for each case in the risk population by using the estimated models. As the incidence rates were person-year based, they had to be divided by 12 so that they corresponded to a time period of approximately one month. An average expected probability of hip fracture could then be calculated by summing the expected probabilities across the risk population and finally dividing by the size of the risk population. By subtracting one month more from each age and making the predictions again, an average expected probability of hip fracture after a two-months (backward) follow-up could be derived. Following the same procedure, expected probabilities up till a 160-month (backward) follow-up were calculated.

The observed (solid line) and expected probabilities (dotted lines) are presented in Figure 4. The curve of observed probabilities showed that there was a non-stabilized risk for a hip fracture admission up to about 120 months. The difference between expected probabilities based on one- and ten-year clearance periods was quite small, and the observed risk also remained higher than expected during the first 7–10 years. Based on these analyses, a conservative ten-year clearance period was selected to make sure that the risk was reduced to the same level as that without a preceding hip fracture. The ten-year criterion has also been used in other studies [168].
5.2 Risk factor extraction

The use of a clearance period in the determination of first hip fractures identified the persons with first hip fracture in addition to the determination of the number of first hip fractures. This made it possible to extract the status of certain hip fracture risk factors available in the data in relation to these patients. Three different risk factor extraction techniques, internal, external, and empirical extraction, were used for this purpose.

Internal extraction corresponded to the use of data abstraction rules within the register data. For example, using the Finnish data it was not enough to determine the index hip fracture period, but the actual day of hip fracture had to be inferred using appropriate rules. The basic idea in the rule construction was to find one simple rule at a time that was applicable to the majority of the population, and each additional rule was then developed and applied only for the remaining population. The index period could be a very long stay in residential care, while the diagnosis referred to an accident that happened near the discharge day, or the index period could be a surgical period, while the actual accident had happened.
before the admission. Therefore, the first rule aimed to identify hospitalizations with an operation code for a hip fracture procedure in the proximity of the candidate period, and the second rule identified hospitalizations at the surgical ward. The third rule was developed to identify patients who had died shortly after the first diagnosis. With these rules a reasonable index period could be found for about 96.9% of patients, while the last three percent required more complicated identification strategies.

After the detection of the most reasonable index period, the basic covariates (such as the patient’s age, sex, area of living) could be extracted from that period. Type of fracture was determined on the basis of diagnoses from the operative surgical period if possible, because those were most likely the first ones based on x-ray examination. Further abstraction rules were required for the definition of more complex covariates. For example, a person was classified as a long-term care patient if there were at least 90 days of inpatient care or a recorded administrative decision for long-term care immediately before the fracture.

In external extraction, variables linking individuals to aggregate levels (such as area codes) were first internally extracted and then external data describing aggregate units were linked back to each individual with the help of these variables. For example, the area code (municipality) at the index period was used to classify each patient as rural or urban (including semi-urban areas) using the official grouping defined by Statistics Finland (rural municipalities are those municipalities in which less than 60 per cent of the population lives in urban settlements).

\[\text{FIGURE 5. Daily numbers of the hip fractures in Finland 1998–2002}\]
In empirical extraction, the observed data were used to justify the definitions. For example, the Finnish register data allowed accurate calculation of the daily numbers of hip fractures, also separately for long-term care patients. After smoothing with a moving average technique, a small but clear seasonal variation attributable to non-institutionalized persons was detected (Figure 5). Without these kinds of preliminary analyses it would have been more difficult to justify that winter season should last from November to April and correspondingly the summer season from May to October. The method developed above for the definition of the length of clearance period was another example of empirical extraction.

5.3 Risk population data

For incidence calculations, data on risk population were also needed, and official population figures (taken on the last day of the years 1997–2002) in 5-year groups were obtained for each municipality. Municipality worked as an aggregate unit in external extraction, and also allowed easy determination of population figures for any combination of these basic units.

It was more difficult to determine the risk population for internally extracted risk factors. There were no data available in the official statistics on the numbers of people in long-term institutional care with the required, rather detailed stratification. Therefore, data on long-term institutional care for the period 1997–2002 were extracted from the Finnish Social and Welfare Care register using the recorded decision for long-term care or more than 90 days of continuous care in a single institution as the extraction criteria (n = 532 169). In principle, these data allowed the calculation of the total numbers of clients in long-term institutional care with the required stratification, as the register included individual-level data on all inpatient hospital and nursing home care in Finland.

However, it turned out that the calculation technique used in the derivation of official statistics utilized only census data (collected at the final day of each year), while the number of long-term care patients seemed to be much higher if discharge data were also taken into account [4]. This was examined more closely by cross-tabulating both numbers with the stratification of sex, year, age, and type of institution. The proportion \( p \) of census data clients \( (O) \) in combined data clients \( (E) \) corresponded to the prevalence proportion \( p = O / E \). In order to describe the differences in the proportions and to estimate the missing \( E \) for the latest year, it was assumed that the counts \( O \) were Poisson distributed and a Poisson regression model was formulated of the type

\[
\log(O / E) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k,
\]
where $x_i$ are the covariates and the model parameters $\beta_i$ are log relative risks [169].

The final model estimated with the years 1997–2001 included main effects for age, sex, and type of institution as well as the year-effect for each age-institution combination. The fit of the model was very good in terms of $\chi^2$-test, and no oddities were detected in residual plots. As an assumption of binomial distribution instead of Poisson distribution was theoretically more elegant, the model with a log-binomial formulation was also estimated [170]. The point estimates were practically identical, but the variances of the coefficients were somewhat smaller, and turned out to be almost equivalent to the robust variance estimates in Poisson regression. As a conclusion, an easily estimable Poisson regression model resulted in good point estimates, but gave more conservative (wider) confidence intervals.

The proportions increased with age and also with year in social institutions, but were smaller among men and in social institutions in general. Reasons for most differences could be traced back to the definition of long-term care, but also to the properties of the register data especially in the Social Welfare Care register. It was concluded that the approach where both census and discharge data had been taken into account was better suited for epidemiological purposes [4].

The practical problem from the incidence calculation point of view was then to calculate the numbers of clients on the final days of the years 1997–2002. That was straightforward for all but the most recent year which required prediction using the estimated model, because no discharge data concerning the year to follow were available.

The procedures described above resulted finally in simultaneously observed risk populations on the final day of the years 1997-2002 with stratification by sex, age, urbanity, and institutionalization. A common method has been to interpolate average population at risk by using the mean of observed census data on two consecutive years. In order to generalize the interpolation so as to also incorporate the seasons, the population figures were plotted against the years in each group with stratification by sex, age, institutionalization, and urbanity. Visual inspections suggested that the associations between the years and population figures were rather simple, and that simple curve-linear models could be used to approximate the relationship. Therefore, linear regression models with constant, year, and squared year as exploratory variables were used to summarize the curve-linear trend in the population figures. These models allowed an interpolation of population figures to any day during 1997-2002. These models were then used in the calculation of approximate follow-up times with the required stratification including seasons and years.

These denominators were desirable in terms of accuracy, but not perfect, because an assumption of a stable population (no short-term fluctuations in migration or mortality) was needed for interpolating census day population figures to appropriate mean follow-up times [171]. In addition, it was not possible to exclude persons who had had hip fracture during the preceding ten years. Bias
resulting from keeping the prevalent pool in the risk population was estimated to be small because of low incidence and high mortality, and to be towards conservative estimates (i.e. to underestimation rather than overestimation of group specific incidences).

5.4 Hip fracture incidence between 1998 and 2002 in Finland

After the adequate data were finally assembled, it was straightforward to examine incidence rates with standard methods [165]. Standardized incidence was higher among women and long-term care patients, and a long increase in standardized incidence in Finland was found to be stabilized [5].

From the hip fracture surveillance system point of view, it was particularly important to be able to monitor the small-area variation in hip fracture incidence. The estimation of the hospital-district-specific standardized incidence rates was performed by using a hierarchical estimator, because the small sex- and age-group-specific numbers of events were prone to extensive random variation. More specifically, it was assumed that better estimates (in terms of mean squared error) could be obtained by a weighted sum of area specific estimates and an appropriate guess [172]. In an empirical Bayes-approach the guesses are chosen to be the average estimates obtained without area stratification [173]. The hospital-district-specific weights and group-specific rates were estimated for each year by using an iterative algorithm [172]. According to the results [5], the incidence had been somewhat higher in Helsinki, but otherwise the differences were rather small (Figure 6).

The constructed population-based data also allowed the examination of the simultaneous impact of several risk factors on hip fracture incidence [3]. Year was not considered as an interesting factor, so the data from years 1999–2002 (during which the average incidence had remained constant) were combined. A Poisson regression model was used in the analysis [169]. There were a lot of significant interactions between risk factors, and it seemed to be more fruitful to model the cross-tabulated data in the style of log-linear analysis [174] instead of aiming at determining the independent impacts of the risk factors on the incidence by means of epidemiological modeling. In this sense, the modeling goal was to compress the representation of the cross-tabulated data, and all main effects and statistically significant interactions up to the third degree were included in the model. The fit of the final model was reasonable in terms of \( \chi^2 \)-test, and residuals did not reveal any exceptional observations. The relaxation of the Poisson assumption by using a negative binomial regression model [175] turned out to be unnecessary as it was after estimation almost equal to the Poisson model, indicating that no significant overdispersion was apparent.
Figure 6. Age- and sex-adjusted hip fracture incidence in Finland 1998–2002 by hospital district, rates per 100 000 persons years. Year-specific rates with confidence intervals are plotted for each hospital district, from 1998 (uppermost) to 2002 (lowest). The short vertical line at each hospital district corresponds to the average rate in 1998–2002 in that hospital district. The dashed line gives the overall mean incidence during 1998–2002 in Finland.
As there were quite a lot of parameter estimates, it was assumed that the simplified cross-tabulation based on the model predictions had captured the systematic characteristics of the original data and removed the random noise, and the actual effects were investigated in the terms of group-specific predicted rates. The main results illustrated in Figure 7 indicated that there was a higher hip fracture incidence among older and institutionalized persons. The incidence among women managing at home increased more rapidly that among men, but in institutions the risks were rather similar. The incidence was higher during the winter time for non-institutionalized persons. The effect of urbanization was more complicated suggesting that the effect was different between men and women, and that there was a tendency for higher incidence among institutionalized persons, the youngest of the men and oldest of the women in urban areas.

The results could be interpreted from the population aging point of view, and seemed to reflect aging-related disability [3, 5]: Institutionalization is a sign of reduced coping with daily activities [176], while women tend to have more disabilities than men as age increases [177], and the additional hip fractures during winter occur mostly among persons with a good enough functional status to allow walking outdoors [178]. As reduced functional capacity and coping with physical activities of daily living are known to be important risk factors for hip fracture [179], it was hypothesized that hip fracture incidence would be proportional to the prevalence of disability related hip fracture risk factors in the underlying population. The hypothesis was not directly falsified and requires further examination, as a

![Figure 7. Hip fracture incidence in Finland in 1999-2001 by sex, age, institutionalization, season, and urbanity. Rates are per 100 000 person years, logarithmic scale. Dotted lines refer to rates for urban areas.](image-url)
high correlation between the prevalence of outdoor walking ability and hip fracture incidence was detected [3].

5.5 Conclusions on hip fracture incidence monitoring

The feasibility of producing information for the purposes of monitoring hip fracture incidence based on Finnish administrative register data was demonstrated in the current chapter. It was pointed out that unwary use of Finnish register data had resulted in biased estimates of the numbers of hip fractures. As there were problems in the reliable identification of all hip fractures, the definition of hip fracture was linked to population aging. Within the renewed theory only the first aging-related hip fracture of each individual had to be found from the data, and a new method was developed to support this purpose. A systematic account of pragmatic problems in the extraction of risk factor and risk population data and in the data analyses was given with proposed solutions. The empirical results suggested that hip fracture incidence would be linked to the prevalence of population level disability. This connection may be important from the monitoring point of view, because under this hypothesis, hip fracture incidence trends would also tell about more general disability trends of the population.
6 Hip fracture treatment

Not all hip fractures can be prevented so it is important that the effectiveness and quality of treatment following hip fracture are continuously evaluated. The basic idea in this kind of outcomes research is to elucidate cause-and-effect relationships between health actions (treatments, interventions) and health gains (outcomes) [150]. Usually large scale empirical investigations that aim to do that are observational studies [180], as it is not possible to assign subjects at random to treatment or control groups, as would be done in a controlled experiment. It is quite common that administrative register data are used in outcomes research [103]. Good examples exist, but the actual assumptions that would justify the use of register data in outcomes research are seldom reported. Such practice makes it impossible to evaluate whether the key assumptions have been uncritically taken as given or whether the assumptions have been carefully tailored to the case under study. Therefore, it was considered important to provide detailed justification for the register-based monitoring of hip fracture treatment, i.e. to aim at finding an adequate compromise between problem-oriented and data-driven theories as proposed in chapter 2.

6.1 State space for a hip fracture care episode

The available data consisted of individual-level data on the use of health services for hip fracture patients. Such data reported directly observable events that determine whether an individual had used certain health service at a particular moment in time. The events describing the use of health services connected data to an idea of record linkage, where such observable events are assembled to the “book of life” of an individual [181]. In this sense, data provided descriptions of patient-specific paths through the observable components of the health system. As the aim was not to analyze complete health service usage, but to examine hip fracture related treatment, an episode-of-care approach was adopted for operationalization purposes [182]. A care episode refers to a series of health-related events linked to a particular health problem [183]. That is, the episodes-of-care approach helps to define health-problem-specific boundaries for the health system structure and to identify interesting elements from the care processes [159].

The first task in the reconstruction of hip fracture care episodes was to identify the beginning of the episode. A natural choice for an index event was the surgical hospitalization after the first aging-dependent hip fracture as described in subchapter 5.1. In addition to being compatible with incidence monitoring, a
restriction to first aging-dependent hip fracture homogenized the study population appropriately.

The next task was to categorize the usage data so that the observed events would have a relevant interpretation from the hip fracture point of view. The idea was to construct a state space that could be used to describe possible transitions across a set of care regions in time [184, 185]. Many different ad hoc classifications of health services had been utilized in the analyses of hip fracture treatment [120, 186, 187]. In addition, theories and models describing the processes of care for the elderly were also applicable to the case of hip fracture [188]. More specifically, it has been suggested that care and services intended for older persons could be categorized in terms of the level of care and the need for care [189]. The level of care is a description of the intensity and type of service, while the need for care is attributable to the patient’s characteristics (such as health status, severity of medical condition and functional disabilities).

Following this idea, the actual care regions were considered to consist of four different levels of care that were identifiable from the data (in increasing order of intensity): 1) home (including home care, ordinary service houses and outpatient care), 2) nursing home (service houses with 24-hour assistance and residential homes), 3) health center (inpatient ward of local primary care unit), 4) hospital. In addition, an absorbing death state was included to the state space. These levels of care are common to the whole country in the sense that all the more specialized components of local structures can be classified into these main categories. Validation analyses reported in article 2 showed that this categorization was compatible with the one used in a standardized hip fracture audit, and that the register data had excellent properties in this respect. The constructed state space is presented graphically in Figure 8.

**FIGURE 8.** State space for the hip fracture care episode
6.2 Effectiveness in terms of care process

The constructed state space was well suited to the description of care processes. In fact, by using only data that were based on directly observable events, it became possible to produce hard facts about the care episodes. The problem was the lack of theoretical links between the care episode data and the outcomes of care. However, by assuming that the outcome of interest was effectiveness—operationalized as a change in health of the target population—and that the patients were not treated ambiguously, it was possible to hypothesize that the decision to treat a patient in a given care region is associated with the patient’s health status [190-192]. In other words, it was assumed that the transitions between care regions reflected the appraised need for care, which is also a sign of a qualitative change in health status. For instance, an acute admission to a hospital typically reflects a sudden drop in health status of a patient, while a discharge to lower levels of care would be expected to reflect a somewhat improved health status. This assumption was supported by empirical evidence [75, 176] showing that at the population level there had been a strong positive correlation between the intensity-ordered levels of care and the need for care in Finland.

Using the notation from chapter 2.3.1, there was a problem-solving task related to effectiveness, where a mental representation for the task solution

\[ I_{\text{user(effectiveness)}} | \text{producer(health service usage)}} \]

was based on the data

\[ D_{\text{user(effectiveness)}} | \text{producer(health service usage)}} \]

that had been derived from the externally available data

\[ D_{\text{producer(health service usage)}} \]

and that had an acceptable cognitive fit with the internal representation of suitable data

\[ D_{\text{user(effectiveness)}} \].

As a conclusion, an appropriate compromise between the data-driven and the problem-driven theories could be reached in the terms of available data.

6.3 Multivariate responses in outcomes research

By following the episodes-of-care approach it became clear that the whole realization of the care process was important in performance measurement. This was somewhat conflicting with the traditional outcomes research models that typically deal with univariate response variables or try to combine information from several indicators [96]. In the methodological sense it was obviously more proper to deal with a multivariate response variable that described the whole care process. This posed certain challenges, because no well-established methods were available for concrete analyses.

More specifically, the available register data allowed complete observation of realized levels of care on a daily basis for each individual. The care process of an individual \( j \) consisted of a sequence of multinomial observations, i.e. corresponded to the repeated categorical response profile...
$Y_j = (Y_{ij}, \ldots, Y_{kj})$,

where $k$ is the length of a sequence. One option for data analysis would have been to use a multivariate probability distribution for the responses [193]. However, from a pragmatic point of view it was considered more important to be able to summarize the multivariate responses in a comprehensive, yet intuitively meaningful way. This kind of exploratory perspective for the data analysis of categorical response profiles has been developed within the sequence methods framework, where the basic idea is to analyze whether useful patterns can be found among the sequences [194]. However, no methods matching the targeted purpose were found in the literature. The most relevant studies concentrated on the clustering of sequences to appropriate groups [195, 196], but did not offer much help on the visual summarization of sequences within the groups.

### 6.3.1 Traceplots for care episode visualization

I show here that it was possible to achieve the required comprehensive summarization by simply plotting the actual realizations of patient-specific care episodes. First, it was assumed that some symbol is linked to each possible category. It was obvious that each care episode could then be expressed with a sequence of these symbols. For example with the hip fracture case, if 3 stands for hospital use, 0 for being home, and 4 for death, the sequence

\[3333330000000000033444444444\]

tells that the individual was followed for 30 days (one symbol for each day); first she was seven days at the hospital, and then after twelve days at home she had a readmission to hospital where she died.

Matters became more complicated when there were sequences for more than one individual. Instead of using numerical symbols, it was beneficial to use colors for telling the realized state. With the hip fractures it was natural to use a gray scale where a darker color reflected the more intensive level of care. By using the colors to indicate the level of care, it was easy to plot diagrams in which each row corresponded to a care episode of one individual. This idea is illustrated in the Figure 9, where the register-based hip fracture care episodes with one-year follow-up were plotted for the randomly ordered cohort of patients living in the town of Espoo during 1998–2001 ($n = 516$).
The presentation of care episodes in a random order resulted in a diagram without any simple interpretation. Intuitively, it was clear that the sorting of episodes could reveal some patterns among the sequences. As there were no known solutions to this kind of problem, a heuristic algorithm for the sorting was developed. The basic idea in the algorithm was to measure similarity between the care episodes and to find an optimal location for each episode in the figure. As the aim was to find a visually simple presentation, common sense logic suggested that the episodes with a lot of common realized states had to be similar. This idea was formalized by measuring similarity between two episodes in terms of the number of equal daily states, i.e. as a sum of the daily specific Hamming-distances (zero if the realized states were the same, otherwise one).

In the actual algorithm two linked lists were used. The first singly linked list (unsorted list) was initialized to contain all episodes in a random order, and the second doubly circular linked list (sorted list) with a single seed episode consisting of only death statuses. The fundamental loop in the algorithm aimed to move one episode from the unsorted list to an optimal place in the sorted list. More specifically, the unsorted list was looped over one episode at a time, and distances
between this current candidate episode and the preceding and following episodes at each possible place in the sorted list were calculated. The episode with the smallest distance in the unsorted list was then moved to the place in the sorted list that had given this smallest distance. The fundamental loop was repeated until all episodes from the unsorted list had been moved to the sorted list.

A sorted traceplot for the Espoo cohort is presented in the Figure 10. The sorted episodes had a simple structure which identified several qualitatively different types of episodes. In this sense, the face validity of the developed algorithm was pretty good. From the methodological point of view the aim of achieving an easier interpretation from the sorting reminds one about the use of rotation in factor analysis: there was no unique solution and the interpretability of the resulting simple structure was the criterion that finally determined the utility of the solution.
6.4 Comparing the care episode profiles of subpopulations

Even though the illustrations of the realized care processes of total populations based on the factual data were useful as such, there was also a need for comparisons of multivariate responses in different subpopulations as especially the comparisons may tell something about the effects that are of interest. In order to illustrate such comparisons with the hip fracture data, patients who were institutionalized at the time of fracture were selected as a base group. This was an interesting group from the health services point of view, because it could be interpreted as representing the upper boundary of the expected use of resources for hip fracture patients. In addition, the recent Finnish clinical guideline for hip fracture treatment had requested further research on the outcomes of those patients who already were in long-term care at the time of fracture.

In order to examine how institutionalization changes the treatment to be received, the patients managing at home (at the time of fracture) were considered an interesting group to be compared with the institutionalized patients. More specifically, the patient was defined as home-dwelling at the time of hip fracture if there had not been care during the 30 days before the fracture, and no long-term care in the year before the fracture. This population had a low initial need for care, and it represented the lower boundary for the expected resource usage.

After the exclusion of patients aged less than 65 at the time of fracture, there were 4325 patients identified as institutionalized at the time of hip fracture in 1998–2001 in Finland. This was about 21.2% of all patients aged 65 or more with the fracture of a proximal femur. The population of home-dwelling patients was larger and consisted of 10 347 patients.

6.4.1 Risk adjustment

An observational study is considered biased, if the populations to be compared have differences in observed or unobserved background factors in ways that confound the outcomes of interest [180]. In the strict sense, the comparison between institutionalized and home-dwelling patients was not an observational study, because it was not reasonable to conceive of the groups as interventions and the interest was not on the causal effects of any intervention but on the observed differences between the groups. Anyhow, there were background factors, such as age and sex, whose impact had to be controlled, and an actual method that could be used to control for these confounding variables was needed [197].

Regression models have been widely used for the adjustment [151]. However, in the case of current multivariate responses, that was not a realistic approach. A
more attractive and pragmatic choice for adjustment was the matching framework [198]. In addition to the fact that matching works with any response (even if the response was still unobserved at the time of matching), it can ensure that the characteristics of the compared populations truly overlap, and that each patient remains intact as a single patient and observed differences directly reflect disparities between real patients and not between any artificial predictions based on estimated model parameters [199].

As the number of covariates to be adjusted was small, a pair directly matched according to sex, fracture type, and age (in years) was chosen for each patient in the institutionalized population \( n = 4325 \) from the randomly ordered set of home-dwelling patients \( n = 10347 \) without replacement. Perfect matches were found for 98.7% of the patients, while the rest had minimal differences in age. If the number of factors to be matched would have been larger, other matching techniques (such as the propensity score approach) could have been utilized [198].

### 6.4.2 Summarization of the follow-up data

In order to get a comprehensive explorative view of the data, the traceplots describing the care episodes were considered to be potentially useful. As the subpopulation to be compared had a different initial need for care, the main interest was on the summarization of the need for care after the hip fracture. It was also known that the

![Institutionalized patients vs. Home-dwelling patients](image)

**FIGURE 11.** Traceplots for the hip fracture care episodes of institutionalized patients in Finland in 1998–2001 and for home-dwelling patients matched for age, sex, and fracture type. Each row corresponds to a care episode of one patient, which consists of 120 daily realizations (four-month follow-up) of the level of care, where the darker color reflects a more intensive level of care (white = home, light gray = nursing home, middle gray = health center, dark gray = hospital, black = death).
levels of care described the intensity of care in an ordinal fashion, and that one step in the levels of care corresponds to approximately a constant step in the functional dependency measures [75, 176]. Therefore, it was assumed that an average level of care over the follow-up period would work as a proxy variable for the need for care of an individual. These averages were calculated and the traceplots were sorted according to them. As Figure 11 reveals, even the straightforward mean sorting resulted in rather simple and interpretable structures. For both subpopulations the sorted order could be interpreted as an approximate continuum from good health (top) to bad health (bottom). As expected, there were huge differences between the institutionalized and home-dwelling patients.

Another way to summarize follow-up data in the case of hip fracture treatment was developed in Sweden in the middle-1980s [200]. In these state diagrams, the proportion of patients in each state is calculated for each day, and the daily proportions are then plotted. A logarithmic scale on the time dimension is typically used in order to emphasize the beginning of the follow-up time in relation to its importance in the hip fracture case. It turned out that the state diagrams corresponded to a kind of margin of the traceplot, i.e. could be derived by summarizing the vertical dimension of the traceplot with proportions of patients in different states.

The state diagrams in Figure 12 clearly demonstrate the most obvious differences between the institutionalized and home-dwelling patients, such as the higher mortality among the institutionalized patients and the greater proportions of lower levels of care for the home-dwelling patients. In fact, the state diagrams gave an easily interpretable yet comprehensive view of the multivariate response by reporting simultaneously a huge number of measures that could be used as traditional performance indicators, such as day-specific mortalities.

A. Institutionalized patients

B. Home-dwelling patients (matched pairs of institutionalized patients)

![State diagrams for hip fracture patients aged 65 or more](image)
The information in the state diagram could be further compressed. An obvious way was to compress the multivariate response to the state indicator that reported the mean proportions of different levels of care during the follow-up after the fracture for both subpopulations. Such state indicators are reported in Table 2. These percentages would correspond to areas in the Figure 12, if a standard time scale (instead of a logarithmic one) had been used in the figure.

TABLE 2. Percentages of days spent at different levels of care during the one-year follow-up of hip fracture patients

<table>
<thead>
<tr>
<th></th>
<th>Home</th>
<th>Nursing home</th>
<th>Health centre</th>
<th>Hospital</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institutionalized patients</td>
<td>6.3</td>
<td>32.3</td>
<td>26</td>
<td>3.7</td>
<td>31.7</td>
</tr>
<tr>
<td>Home-dwelling patients*</td>
<td>54.0</td>
<td>6.6</td>
<td>17.1</td>
<td>4.8</td>
<td>17.6</td>
</tr>
</tbody>
</table>

* Age, sex, and fracture type matched pairs of institutionalized patients.

6.5 Conclusions on monitoring hip fracture treatment

In summary, this chapter aimed to provide a justification for the register-based monitoring of hip fracture treatment. First a state-space model for hip fracture treatment that was compatible with the register-based data was developed. Then assumptions which were used to link the developed data-driven model to the theory of effectiveness measurement were discussed. It was pointed out that the combination of theory and data naturally leads to the use of multivariate responses, and an illustration of summarization of multivariate response with novel methodology were given. In addition, an example of adjusted comparisons between multivariable responses was given. Finally it was demonstrated how complicated responses could be reduced to simpler summarizations and further to more traditional indicators.

In conclusion, this chapter demonstrated how it was possible to find an adequate compromise between problem-oriented and data-driven theories. This idea can be illustrated within the general framework for register-based health services research that I have proposed [9]. Figure 13 visualizes the general structure of the conceptual framework. The figure should be read from left to right. The components that are closer to each other in the figure are more directly connected in the framework. The two central paths from left to right in the figure correspond to theory-oriented modeling (solid line) and to data-driven modeling (dashed line). It must also be remembered that the selected health problem gives a third and most concrete theory to be utilized in the actual application of the framework.
FIGURE 13. Visualization of the conceptual framework for register-based health services research. A solid line reflects the theory-driven approach, and the dashed line the data-driven approach. The approaches are most compatible under the oval.
7 Operative delay as a performance indicator

In the previous chapter it was demonstrated what kind of assumptions were needed to transform register-based data into factual information about hip fracture treatment. In this chapter, it is illustrated what kind of issues may be related to the justification of the use of a performance indicator. This chapter is mainly based on article 6, but the presentation of a model for the provider profiling originates from article 1.

A surgical operation is performed for most patients during the acute management of hip fracture. A typical hip fracture patient is confined to hospital bed-rest before the surgical operation. A fairly short delay between hospital admission and surgical operation has been advised in the clinical guidelines [128, 129], but considerable variation in the operative delays between the providers has been reported [201–203]. This discrepancy, which points out a potential area for improvements, makes the operative delay an attractive performance indicator [204].

However, the evidence concerning the timing of surgery in relation to the overall outcomes of hip fracture treatment has been quite mixed. This is unfortunate, because the operative delay should have confirmed effects on outcomes, otherwise improvements may possibly be targeted at issues which do not improve health [205]. Therefore, it was considered important to examine the association between operative delay and mortality more carefully. The data were restricted to patients aged 65 or more and operated on using an internal fixation, a prosthesis, or a total hip replacement during 1998–2001 in Finland. The study population consisted of 16 881 patients, which was 83.3% of all first-time hip fracture patients aged 65 or older in Finland 1998–2001.

7.1 Effects of different operative delays on mortality

The first open question was the definition of the time period for delayed surgery [204]. The register data included admission and operation days, so the operative delay could be calculated only with limited accuracy, as the number of nights spent in a hospital before the operation. Time from admission to death was measured in days. The cumulative probabilities of death for different operative delays were calculated using Kaplan-Meier estimators. As can be seen from Figure 14,
waiting times from 0–2 nights had the same effect on mortality, but there was a clear increase in mortality with longer waiting times. Therefore, early surgery was defined as consisting of patients with a 0–2 nights waiting time and late surgery as consisting of patients with at least 3 nights waiting time.

7.2 Profiling of providers

Next, it was interesting to compare whether the potential performance indicator (proportion of late surgery patients) varied by hospital. This kind of activity has been commonly referred to as the profiling of providers [206]. It has been pointed out that the results of provider profiling have been commonly interpreted as a ranking of providers, even though the extent of variation attributable to the quality of provider may have been negligible [207], and an appropriate statistical modeling has been considered as a necessary prerequisite for justified conclusions in provider profiling [208]. In fact, it has been suggested that there exist variations between providers for at least three reasons: 1) differences may be attributable to a random variation related to the size of the provider, 2) the patient case-mix varies from provider to provider, and 3) providers may differ in the performance of their care [209].
There were several commonly applied risk-adjustment methods available for the case of a binary response variable [210]. Following the standard approach, a model for the probability of late surgery was constructed by using a logistic regression model

\[
\log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k ,
\]

where \( x_i \) are the predictor variables and the model parameters \( \beta_i \) are log odds ratios. Age, sex, fracture type, comorbidities, pre-admission residence, inpatient care preceding fracture, admission weekday, and year were used as predictors. The estimated probability of late surgery was calculated for each individual \( j \) as

\[
p_j = \logit^{-1}(\hat{\beta}_0 + \hat{\beta}_1 x_{1j} + \hat{\beta}_2 x_{2j} + \cdots + \hat{\beta}_k x_{kj}),
\]

where \( \hat{\beta}_j \) are estimated parameters, and

\[
\logit^{-1}(z) = \frac{\exp(z)}{1 + \exp(z)}.
\]

The fit of the model was acceptable yielding a c-statistics value 0.65, indicating a reasonable ability for discrimination.

Since the focus in profiling was on providers and not on individuals, the observed (\( Y \)) and expected outcomes (\( p \)) were then aggregated to the provider level [207], i.e.

\[
O_i = \sum Y_j
\]

and

\[
E_i = \sum p_j,
\]

where \( Y_j \) refers to the individual’s observed status of late surgery and the sums are over patients treated by provider \( i \). The ratio of observed to expected numbers of late surgery \( O / E \) was then used to quantify the risk-adjusted differences between providers. This quantity can be interpreted as the relative risk of a given provider to have the same share of late surgery patients as would have been expected on average for the similar patient population [210].

However, as the adjustment is practically always only partial, there remains residual variability between the providers that may result in overestimated differences in profiling analyses [211], and it has been recommended to use of hierarchical (random effect) models in the profiling [208]. The advantages of hierarchical modeling include the automatic removal of regression to the mean.
Operative delay as a performance indicator

The common approach in the modeling of the $O/E$-ratio has been to assume that the observed counts $O_i$ being non-negative integers describing frequencies of rare events have a Poisson distribution with mean $\gamma_i$, that is

$$O_i \sim \text{Poisson} \left( \gamma_i \right).$$

Two types of hierarchical models based on the Poisson assumption were constructed, a gamma-Poisson model [213] and a Poisson-lognormal model [214].

The gamma-Poisson case was modeled as

$$\log \gamma_i = \log E_i + \log (\lambda_i),$$

where the expected outcomes $E_i$ adjust the patient-level variation and

$$\lambda_i \sim \text{Gamma} \left( \alpha, \beta \right)$$

describe the provider-level variation of interest. A vague exponential prior was used for $\alpha$ and a gamma prior for $\beta$. The posterior distributions of $\lambda_i$ corresponded to the shrunken $O/E$-ratios and were of primary interest. In the Poisson-lognormal case the model took the form

$$\log \gamma_i = \log E_i + \theta_i,$$

where

$$\theta_i \sim \text{N} \left( \mu, \sigma^2 \right),$$

which could have been easily extended to incorporate provider-level covariates. A normal prior was used for $\mu$ and a gamma prior for precision $1/\sigma^2$. Here the posterior distributions for $\exp(\theta)$ were the interesting ones.

In principle, the conditions justifying the Poisson approximation for the distribution of $O_i$ were not very well satisfied, because there was a fixed number of patients at each provider and the proportions of late surgery patients were quite large. Therefore, a beta-binomial model [213] and a logistic model [207] were also formulated. For both models it was assumed that

$$O_i \sim \text{Binomial} \left( p_i, n_i \right),$$

where $n_i$ stands for the number of patients at provider $i$. The proportion $p_i$ was modeled in the beta-binomial case as

$$p_i \sim \text{Beta} \left( \alpha, \beta \right),$$

and in the logistic case as
Operative delay as a performance indicator

\[
\text{logit}(p_i) = \theta_i,
\]

where

\[
\theta_i \sim N(\mu, \sigma^2).
\]

The gamma priors were used for \( \alpha \) and \( \beta \), and the priors for \( \mu \) and \( 1/\sigma^2 \) were the same as in the Poisson-lognormal model. The posterior distributions for \( (n_i/E_i)p_i \) gave the shrunken \( O/E \)-ratios in both cases.

The actual estimation was carried out by using Markov chain Monte Carlo (MCMC) simulation [215]. The mixing of the estimation procedures was examined by using two chains in each estimation, and the convergence was evaluated on the basis of Gelman-Rubin convergence plots [216]. After an initial 1000 burn-in iterations, a further 10 000 iterations were used in the actual estimation of the parameters. The complexity and relative fit of the models were assessed with the deviance information criterion (DIC) [217]. The obtained shrunken \( O/E \)-ratios were further multiplied by the average share of late surgery patients so that the reporting could be done in terms of easily interpretable risk-adjusted proportions for providers [207].

The actual results were rather similar for all models. In terms of DIC, the fit was somewhat better with binomial assumption than with the Poisson assumption. With the binomial assumption variance tend to be smaller if the provider was small or the proportion of late surgery patients was large. The shrinkage was a bit stronger for models with non-normal random effects distribution. As the purpose was not to rank the providers or identify unusual performance, but to indicate the potentiality for improvements, the most conservative gamma-Poisson model was selected in order to avoid too strong interpretations.

The potentiality for improvements was assessed by assuming that the percentage of late surgery patients can be reduced to some potentially achievable level [218]. The level was determined empirically from the provider-specific shrunken estimates by assuming that if 20% of providers have achieved that level or better, the others should also be able to do it. In addition, the 20th centile is about one standard deviation from the mean and can be interpreted to reflect the best practice proportion [213]. As can be seen from the Figure 15, there was extensive variation in the proportions of late surgery between the providers, and the 95% credible intervals crossed the potentially achievable share of 92.7% of early surgery patients only for less than half of the providers.
Operative delay as a performance indicator

As the difference between unadjusted mortalities in late and early surgery groups was considerable, it would have been straightforward to infer that the reduction of the operative delay would also diminish the mortality. However, there were no such huge variations in provider-specific mortalities, and the linkage between operative delay and mortality had to be examined more carefully.
7.3 Adjusted effect of operative delay on mortality

In fact, because of the potential biasing effects of different patient characteristics between the early and late surgery group, there was first a need to adjust for these factors. As the data included times from admission to death or end of follow-up, a Cox proportional hazards model were used to control for the observed patient characteristics.

A proportional hazard model specifies the hazard for individual \( j \) as

\[
\lambda_j(t) = \lambda_0(t) \exp(\beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj}),
\]

where \( \lambda_0(t) \) is a nonnegative function of time \( t \) known as baseline hazard, \( x_i \) are the predictor variables and the model parameters \( \beta \) are log hazard ratios [219]. In the Cox model, the baseline hazard function \( \lambda_0(t) \) is considered as a nuisance parameter, and parameters \( \beta \) are estimated semi-parametrically without fixing the form of \( \lambda_0(t) \). The estimated hazard ratio between the late surgery and early surgery was 1.24 (95% CI: 1.15 to 1.34) as no other covariates were included. The hazard ratio for the late surgery reduced to 1.18 (95% CI: 1.09 to 1.28) after the inclusion of age, sex, fracture type, comorbidities, pre-admission residence, inpatient care preceding fracture, admission weekday, year, and hospital dummies as covariates.

Since there were no detailed clinical data available, it was likely that certain relevant covariates were missing from the model. In addition, the patients were known to be nested to hospitals. In the case of survival data, the concept of frailty offers a suitable way to incorporate random effects in the proportional hazards model to account for association and unobserved heterogeneity. A frailty may be considered as an unobserved random factor \( Z \) that has a multiplicative effect on the hazard function of an individual or a group of individuals. More specifically, a constant shared frailty model is a conditional independence model having the form

\[
\lambda_j(t | Z_i) = Z_i \mu_i(t) \exp(\beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj}),
\]

which corresponds to the hazard function of the individual \( j \) belonging to group \( i \) given the group specific frailty \( Z_i \), where \( \mu_i(t) \) is a baseline hazard rate that may or may not depend on \( i \), and \( \beta \) and \( x \) are as above [220]. The frailty \( Z \) was assumed to have a form of gamma distribution and to vary between hospitals \( i \), but a common baseline hazard \( \lambda_0(t) \) was assumed. This model resulted in a slightly smaller hazard ratio 1.17 (95% CI: 1.08 to 1.27) than with the fixed effects model.

However, the fixed and random effect models failed to fulfill the assumption of proportional hazards [221]. Therefore, a stratified Cox model, which allowed a
different baseline hazards for groups of interest but common values for coefficients $\beta_i$, was estimated with robust sandwich variance estimators which took the clustering by hospitals into account [219]. Stratification was done in terms of age, sex, cancer, dementia, and hospital. The resulting model seemed to fulfill the proportional hazards assumption and gave a hazard ratio estimate of 1.15 (95% CI: 1.03 to 1.29), still indicating a higher mortality among late surgery patients.

The plotting of smoothed hazard functions revealed that there was clearly increased risk of death during the four months following the hip fracture in early and late surgery groups. The plotting of excess cumulative mortality (late surgery mortality – early surgery mortality) showed that the difference between the groups first rapidly increased to 3% and then continued to increase until a maximum of about 5% at 1–3 years. Based on these considerations, a binary variable indicating one-year mortality was selected to summarize information in the actual time-to-death variable. It was further hypothesized that logistic regression analyses could approximately replicate the actual results obtained with the Cox model by using a simpler binary response variable. The concrete analyses confirmed this hypothesis to be true.

With the binary response it was straightforward to apply a pseudo-randomizing instrumental variable approach to the current problem. The basic idea with the approach is to define an instrument variable that causes variation in the treatment variable of interest (early or late surgery in this case), but has no direct effect on the response variable (one-year mortality), because it then becomes possible to estimate how much the variation in the treatment variable induced by the instrument affects the response variable [222]. The pseudo-randomized assignment of patients into the early and late surgery groups was done by using the admission day of the week as an instrumental variable [223, 224]. This variable had an effect on operative delay, but no direct effect on mortality in the case of hip fracture [225]. In practice, the approach took a simple form: to test whether the risk-adjusted one-year mortality was equal for each admission day of the week. It turned out that, according to a $\chi^2$-test with null hypothesis of equal mortalities for each admission day of the week, the difference in adjusted one-year mortality was not attributable to the operative delay. This result was in contrast with the results from the Cox models above.

### 7.4 Provider-level hypotheses

The clear difference of the estimates of the operative delay effect on mortality between the methods indicated that there was a need for additional methods to explain such differences. Fortunately, the results pointing out the huge variation between providers in the shares of late surgery patients made it possible to formulate novel hypotheses to address the problem. First, it was assumed that
an acceptable delay in hip fracture surgery corresponds to the clinical decision to postpone the operation and that the clinical decisions remain nearly constant within providers. Under this assumption, the potentially achievable lower limit in the proportion of late surgery patients could also be interpreted as the upper limit for the proportion of acceptable delayed patients. Correspondingly, the expected proportion of unacceptable delayed patients was the proportion of late surgery patients exceeding this upper limit. This inference resulted in a hypothesis stating that the overall mortality of hip fracture patients should increase with the rising share of the late surgery patients given that the longer operative delay would have an adverse effect on mortality. Another novel provider-level hypothesis was that the long-term mortality of the late surgery patients is higher if only the patients who are unfit for surgery are delayed, since the unfit condition for the surgery is also a risk factor for one-year mortality.

In order to test these hypotheses, the provider-level heterogeneity was further examined in terms of simultaneous provider specific shares of the late surgery patients and one-year mortality. The provider-specific overall mortalities were decomposed into mortalities of early and late surgery patients by recognizing the fact that the weighted average of these mortalities gives the overall mortality, i.e.

\[ m_{\text{overall}}(p) = p \cdot m_{\text{late}}(p) + (1 - p) \cdot m_{\text{early}}(p), \]

where \( p \) corresponds to the proportion of late surgery patients and \( m(p) \) to mortality with proportion \( p \) of late surgery patients.

It was straightforward to calculate these three (adjusted) mortalities for each provider, and use a scatter plot between the share of late surgery patients and mortalities to describe the associations. In order to extract trends from these associations, it was, after preliminary examinations, assumed that the trends between the share of late surgery patients and overall as well as early surgery mortalities are linear, i.e. that they have the forms

\[ m_{\text{overall}}(p) = a + bp, \]

and

\[ m_{\text{early}}(p) = c + dp, \]

where \( p \) is the proportion of late surgery patients, and \( a, b, c \) and \( d \) are regression coefficients to be estimated from the observed data. Late surgery mortality could then be calculated deterministically from the weighted average formula given above.

Due to the extensive random variation in mortality figures among small providers, there was a need for hierarchical modeling giving shrunken estimates. More specifically, it was assumed that better estimates (in terms of mean squared error) could be obtained by a weighted sum of observed mortalities and appropriate guesses, i.e. that
where $s_i$ is the shrunken mortality estimate for provider $i$, $o_i$ corresponds to observed and $e_i$ to guessed mortality, and $w_i$ is the (probability) weight to be estimated from the data [173]. The guesses were chosen to be the trends extracted from the data by using the formulas given above, and the weights and shrunken estimates were calculated by using Stein estimation in the form of an iterative algorithm [172].

This simultaneous examination of the provider-level proportions of the late surgery patients and one-year mortality illustrated in Figure 16 revealed that there was an almost flat (positive) trend between the greater share of late surgery patients and one-year overall mortality. This confirmed that the actual independent effect of operative delay on mortality is very small. Another result indicated that the smaller proportion of late surgery patients was clearly (nonlinearly) associated with a higher mortality for these patients. In other words, the providers with low proportions of late surgery patients had been able to select the patients with such severity that immediate operation would not have been clinically justified.

![Figure 16. A scatter plot of adjusted mortality for the operated hip fracture patients and the share of hip fracture patients with late surgery for different providers. (The triangles describe the overall mortality of the operated hip fracture patients for each provider, and the crosses the mortality of the late surgery patients for each provider. The line with a positive slope is the trend of overall mortality, and the dotted curve describes the association between the share of late surgery patients and mortality across the providers.)](image-url)
The provider-level associations were further analyzed for five severity groups of patients by using the same methodology. The severity was measured as predicted one-year mortality for observed patient characteristics, and groups were formed by dividing the patients into five equivalently sized classes on the basis of sorted severity score. Intuitively speaking, the least severe patients were young, coming from home and were without severe medical conditions. Correspondingly the most severe patients were older, with much comorbidity and coming from residential care. For the patients in the least severe group, mortality was higher for the late surgery patients for all providers. Probably this indicated that in this group, the hip-fracture-related mortality was caused by the medical problems which necessitated late surgery and the prolonged surgical delay did not itself increase the mortality. For the patients in the most severe group, mortality was also higher for the late surgery patients for all providers. However, for this group the late surgery mortality was lower in the providers with a small share of late surgery patients and there were no differences in early surgery mortalities between the providers. It seems that in this group, it was essential to perform early surgery for all patients who could bear it, since the significantly prolonged surgical delay makes the patient’s condition worse and increases mortality. For groups 2–4 the interpretation was more difficult than for extreme patients. Since the mortality of early surgery patients increased significantly and was even higher than the mortality of the late surgery patients for the providers with a large share of late surgery patients, it seems that in these groups an operation too early may have caused more harm than the prolonged waiting time. In summary, these results actually gave reasonable explanations for the mixed results in the existing literature.

7.5 Conclusions on operative delay as a performance indicator

The analyses presented above demonstrated that methods that have been traditionally used may result in biased results if used uncritically in the case of register-based data, with such a bias being difficult to detect. The problem was that the register data did not include detailed clinical data about the individual-specific factors that may have postponed the operation, and that the actual selection mechanism related to becoming a late surgery patient varied between the providers. In this case the formulation of novel provider-level hypotheses helped to reveal these associations indirectly from the register-based data. In conclusion, the average effect of operative delay on mortality was small, but the provider-level association between the share of late surgery patients and non-optimal treatment was obvious, and the proportion of patients with a prolonged waiting time for a hip fracture operation seems to be an effective quality indicator, since it clearly indicates an area for improvement.
8 Discussion

As a statistician and a scientist I was quite enthusiastic when I first participated in a study that used Finnish administrative registers as a data source. The massive amount of data had been shown to be of good quality and with an internationally exceptional coverage as a result of unique individual identification numbers allowing easy linkages, offering excellent opportunities for valuable health services research. There were many good examples of such research, and it seemed that the administrative data could be used effectively as long as the potential pitfalls were avoided. However, practical data analyses quickly revealed that the meaningful use of data was very challenging, and that the most fundamental parts of the analysis corresponded more closely to a qualitative than to a quantitative methodological approach. This troubled me, because—even though it was possible to successfully use administrative data in health services research—the links between scientific theories and secondary data were mostly extremely fuzzy. It was also clear that the data sources and therefore the methodological problems abroad were not identical to Finnish ones, and the optimal solution for the problems in a Finnish context obviously requires that the methodology for a solution is developed from the Finnish context point of view. In this sense, the main question in the utilization of register-based data turned out to be: How the use of secondary register data could be justified by using strong methodological arguments tailored to the Finnish context?

In this thesis I have tried to give an answer to that question. The first aim of this study was to develop a methodological framework that helps in the effective utilization of register-based data. The problem was approached from a multidisciplinary methodological perspective. Chapter 2 gave a brief review to the methodological approaches, issues, and assumptions that had a fundamental role in the more applied work of later chapters. In this sense, Chapter 2 could also be interpreted as representing summarizations of those philosophic-methodological ideas that are likely to be fruitful in the characterization and solving of the concrete problems in register-based data analysis in general, and thereby sketching the core of the framework.

Two main ideas of the framework should be noted. First is the aim to find suitable compromises between the problem-oriented and data-driven theories, and the second is the empirical justification of theory-driven operationalized definitions, i.e. the confirmation that the theoretical constructs are in concordance with the register-based data. Both of these ideas were used throughout this study, combined with the idea of minimizing the overall bias via the maximization of the use of the most factual data, and reflected the need for data sensitive methodological
thinking. In this sense, the framework is not just about the methods, but about the use of the right methods in the right situations.

There are two potential problems with the framework: 1) the framework is not a sufficient condition for the register-based data analysis, because a strong and more or less tacit background knowledge on multidisciplinary methodological issues (including application experience) is a prerequisite for a full scale understanding and evaluation of the approach, and 2) the framework is not a necessary condition for a register-based data analysis, because it may be possible to end up with similar results without any knowledge of the framework. However, as the presented framework included a multidisciplinary collection of extensions of well-grounded methodological approaches, it is likely that a skilled methodologist would find at least some ideas helpful in the efficient utilization of register-based data analysis.

The second aim was to demonstrate how register-based data can be used as a data source for the performance monitoring of a health system, i.e. to connect the methodological developments to the practical needs of information and further to concrete health policy goals, such as the “steering by information” policy regime. The actual demonstration included four components. The first was the validity of register data, the second concentrated on the epidemiological perspective of the monitoring system, the third on the monitoring of treatment in terms of care episodes, and the fourth gave an example on the justification of a more detailed performance indicator. At the end of each chapter the main conclusions were summarized. From the methodological point of view, each component aimed to find an acceptable compromise between the theory and the data. Especially the first three components also included the tailoring of conceptual definitions and assumptions. Data sensitive empirical analyses had a central role in each component, and throughout these chapters the empirical results were novel contributions to the medical literature.

The development of a comprehensive hip fracture monitoring system was such a huge task, that it was not possible to include everything into this thesis. In fact, several manuscripts developing issues further are already under preparation. These include a review article about prerequisites in register-based research, several papers that extend the care episode ideas presented in Chapter 6 or develop more detailed risk-adjustment methods, and working papers that compare the costs of treatment [226] and model the lengths of stay in treatment processes [227]. In this sense, the practical work is still in progress, but the perspectives given in this thesis introduce a necessary basis for further work and help in the routine implementation of a hip fracture monitoring system [228].

It should be noted that this study concentrated on methodological issues in the wide sense. The main idea in the analyses was to solve a problem, not to use some particular model or technique. Statistical thinking was the key that made it possible to deal with multidisciplinary methodological ideas at a sufficiently abstract level. After an extraction of the “wireframe model”, the actual search for
a suitable method or analytical approach was in some cases straightforward, but occasionally there was also a need to extend the available methods or even create new ones. In this sense, the methodological problems considered important were not on the (mathematical) details of any single model, but on finding appropriate methodological solutions to various interactively encountered problems. This kind of approach naturally required a lot of preliminary and sensitivity analyses, but it was feasible to report only a small selection of those, and mostly the ones that were required to justify the use of the selected models. The drawback in this kind of underreporting is that the actual process of transforming register-based data into useful information may seem to be easier than it really was [3].

All in all, the methodological approach demonstrated in this thesis represents an alternative style of reasoning for scientific research utilizing register-based data. This style of reasoning could be characterized as methodological constructionism or—in the sense of desire for factual data for data analyses—as hermeneutic positivism. But that is another story.
References


References


References


References


References


208. Christiansen CL, Morris CN: Improving the statistical approach to health care
References


Utilisation of administrative registers using scientific knowledge discovery

Reijo Sund
National Research and Development Centre for Welfare and Health (STAKES), P.O.Box 220, 00531 Helsinki, Finland
Tel.: +358 9 3967 2469; Fax: +358 9 3967 2485; E-mail: reijo.sund@stakes.fi

Received 1 November 2002
Revised 11 December 2002
Accepted 25 February 2003

Abstract. The volume of data being produced for administrative purposes is increasing rapidly. Data must be analysed in order to extract useful information to support decision making. The demand for evidence-based information means that the analysis must be conducted according to the principles of scientific research. Unfortunately, the massive second-hand data sets seem not to fit very well into the traditional methodological paradigm. A secondary data source imposes limitations on the formulation of a problem and concepts, because the measurement can only be based on existing data. The aim of this paper is to present a methodological framework for the utilisation of administrative registers in the creation of scientifically valid information. This is done by discussing fruitful methodological aspects encountered in the practical knowledge-discovery process. The ideas presented originate from many different fields, such as statistics, data mining and sociology. The emphasis lies on understanding connections between problem, data and analysis in the case of massive secondary administrative data sources.

Keywords: Scientific research process, secondary data, administrative registers, event history analysis, statistical data mining, health care provider profiling

1. Introduction

Data have been produced for hundreds of years. The reasons for such production were originally administrative in nature. There was a need for systematically collected numerical facts on a particular subject. Later, a belief in the advantages of quantitative information accompanied by advances in statistical techniques led to the birth of a widely used methodological paradigm for scientific research. Even more recently, advances in information technology have made it possible to more effectively collect and store larger and larger data sets. Administrative information systems have been at the forefront of data gathering, since there have been a growing demand for evidence-based information to support decision making and other administrative purposes. In many cases, however, raw data as such are of little value. Data must be analysed in order to convert them to useful information. Unfortunately data-driven analysis and the massive size of data sets seem not to fit very well into the traditional methodological paradigm.

A traditional way of creating scientifically valid information is to design and carry out a scientific study, in which data production, analysis and reporting are all strictly problem-specific. Data represent tailored measurements of observational reality that are needed in problem solving. Analysis is then used
to reveal ‘how things work’ in empirical data. By interpreting these empirical results in the theoretical context and discussing relevant limitations, it is hopefully possible to create a sophisticated answer to a problem in such a way that it can be evaluated by the scientific community.

The most important challenge for the traditional paradigm is the nature of data. If it is not possible to produce tailored data according to the problem in question, the whole problem-solving process becomes more restricted. The secondary data sources utilised impose limitations on possibilities for technical analysis. Moreover, the formulation of a problem and the associated concepts inevitably becomes opportunistic, because the measurement can only be based on existing data [6]. For some researchers, even the validity of data is suspect if they were not produced for the specific purpose for which they are used. Such criticism is not necessarily based on evidence concerning the quality of the information contained by the data but rather on philosophical beliefs regarding the connection between data origin and data quality [35].

In addition, the growing size of data sets has raised some issues that must be dealt with in a different way: data exploration and description plays a more general role than it does in the case of small data sets; efficient algorithms (and computers) are needed in analyses; the unfeasibility of manual analysis results in separation between the data and the analyst; the manual management of all errors in the data is not possible; and the chance occurrence of patterns is more likely in large data sets. In addition, some databases include the whole population and not just a sample, which changes the nature of the whole statistical inference [13].

In other words, in the analysis of secondary data, the problem must be defined under preconditions arising from the data. Moreover, massive data sets are usually so overwhelming that their processing and analysis is difficult (‘data overload syndrome’). It may also be tempting to screen the data for significant associations without having an adequate prior hypothesis (‘snooping/dredging/fishing’). Rather than tackling too much on what can and can not be done, analysis should achieve an appropriate compromise between the practical problem and the data. This kind of activity has been characterised as ‘greater statistics’, which tends to be inclusive, eclectic with respect to methodology, closely associated with other disciplines and also practised by many non-statisticians [3]. Moreover, the larger the data sets are, the more important the general-science and collaboration aspects of the analysis process seem to become relative to the ‘statistical’ aspects.

Since statistical research has traditionally focused on probabilistic inference based on mathematics, it has not been able to offer very much concrete help in these different circumstances. The practical need for information has led to the development of alternative ways of analysing data, such as data mining [9]. However, these strategies for analysing data have much in common, and formal statistical expertise also provides an excellent basis for the understanding and evaluation of these ‘new’ ideas. Statisticians ought to take advantage of the situation, get involved in interdisciplinary activities, learn from the experience, expand their own minds – and thereby their field – and act as catalysts for the dissemination of insights and methodologies [15].

1.1. The aim of the paper

The aim of this paper is to present a methodological framework for the utilisation of administrative registers in the creation of scientifically valid information. This is done by discussing the fundamental methodological issues encountered in the practical knowledge discovery process. The ideas presented originate from many different fields, such as statistics, data mining and sociology. The paper emphasises a broader understanding of connections between problem, data and analysis in the case of massive secondary data sources.
The development of this paper is heavily influenced by the context of health services research, a field in which the most valuable official data sets are individual-based. In these circumstances, the histories of individuals recorded in administrative data are the most general starting point for any analysis – and a fruitful starting point for this paper, too. A practical real-world example – used to demonstrate the research made by following the guidelines presented in the paper – is also taken from the context of health services research.

2. The scientific knowledge-discovery process

The ultimate goal in all research is – or at least should be – to search for appropriate answers to relevant problems. To find sensible and scientifically valid solutions requires not only a knowledge of the phenomena being investigated but also the capacity to think originally about the nature of inference and the process of conceptualisation. The whole problem-solving process must be conducted in a systematic way, using appropriate methods. Even though it is not in general possible to split research process completely into distinct phases, certain common phases are encountered repeatedly and even simultaneously during the research process. The process does not proceed in such a way that one phase is completed first, after which one moves on to the next. Findings made during the research process inevitably have effects on the final conclusions, since these preliminary findings provide a deeper understanding of the phenomenon and data in question.

One sophisticated characterisation of complex problem-solving, is known as a knowledge-discovery process (see e.g. [2,4,8]). However, despite the practical usefulness of a process description of this kind, some critical issues concerning the scientific nature of the research process are not mentioned at all.

To see what is actually happening in the research process, it is helpful to place it in a wider context. For this purpose, an alternative schema for a research process is presented in Fig. 1. The interactive phases related to this schema are: understanding the phenomenon, understanding the problem, understanding data, data preprocessing, modelling, evaluation and reporting.

Research is conducted in order to obtain an answer to some problem. The researcher has to choose the perspective from which solving the problem is approached, since research requires rigorous communication and careful definitions of concepts related to a phenomenon of interest. However, all researchers have some more or less latent foreknowledge, which provides directions and criteria for understanding the world and also guides understanding gained and interpretations made during the research process. The researcher must be aware that these choices and interpretations are his or her ideologically and historically dependent decisions. The choice of perspective and conceptions may even prevent one from seeing (or enable one to see) something important or interesting.

In order to move from the theoretical level to the empirical level, the researcher must operationalise the research frame in such a way that it is possible to produce useful information from the phenomenon.
and thus create some answers to the problem. This kind of operationalisation is fundamental to the plausibility of research. This is the stage at which the research defines the point of view from which the phenomenon and concepts are examined.

Data represent the empirical world in research. Observed information can not cover all details which are relevant to the phenomenon of interest. The observation and measurement must also be dealt with in a conceptual context. This means that some compromises must be made between the exactness of measurement and the ambiguity of observation. Definitions can not replace the empirical information, but definitions affect the organisation of information and the decisions based on such information. From this point of view, it is easy to argue that data should be produced strictly according to the research problem. That is not the case with secondary data sources, such as administrative registers, since secondary data are originally produced for some other purposes. The available data and the conceptual definitions used in their production are not necessarily compatible with the research problem at hand. As an extreme example, imagine that your observational world is all that you see, hear, smell and feel. Would you trade your observational world for an old black-and-white photo, if you could trade it for a high-quality three-dimensional digital image with full colour scale, including infrared?

All in all, the researcher must decide what to observe, how to conceptualise variables, what kind of operationalisation to make and how to acquire appropriate data. The same freedom also applies to the choice of analysing strategy, interpretation of the results and discussion of alternative interpretations. Therefore it is important to reflect upon decisions made during the research process. Every scientific explanation is valid only in so far as the investigator can provide a rationale for what has been done. At each stage of the research, decisions and arguments must not rest on the uncritical or automatic use of established techniques but on a set of logical decisions which must be taken on theoretical grounds with full consciousness of what is being done. Results must be justified. In the example above, one obvious question to be answered is: Why was that particular black-and-white photo chosen? In this example, ideology may be viewed as the position of and lighting for the camera, and historical dependency as the moment of taking the picture in question.

Since in the case of administrative registers the data typically have certain common properties, it is possible to make some suggestions of which it is useful to be aware if data originating from such a source are to be utilised. The next section briefly describes a few ideas that have been found very worthwhile in practice. In the research process these ideas may be seen as aspects of the following three phases: understanding the problem, understanding data, and data preprocessing.

2.1. Understanding the problem

Time is an essential factor in many problem domains. For example, disease processes evolve in time, and patient records give the history of patients. In research on such dynamic phenomena, interest very often focuses on the sequences of events which occur in time and reflect changes in research objects. A good example is an employment career, which can be regarded as consisting of the sequences and timing of the various jobs and of their association with other events, such as changes in marital status and place of residence. In fact, the importance of the longitudinal event history approach has long been recognised in many areas, especially in the social sciences, econometrics and medical research (see e.g. [5,19,34]).

2.1.1. Operationalisation

For the purposes of problem solving, the problem has to be ‘matched’ against the corresponding domain knowledge and data. This can be done by characterising the important properties of the phenomenon carefully, i.e. by operationalising it as a system. A system is defined as a ‘group of things or parts
Fig. 2. A system shown as a directed graph.

working together or connected in some way so as to form a whole’. This definition of a system requires that the concepts, objects and restrictions related to the system be adequately described.

In the case of event histories, a common approach to characterising the phenomenon is to categorise events in such a way that they describe transitions across a set of discrete states in time. In the simplest case only one qualitative change is possible – for example, the transition from the state ‘alive’ to the state ‘dead’. However, in many cases one transition can not describe the dynamics of a phenomenon in a realistic way. For example, the slow development of a disease which finally results in the state ‘death’, can not be modelled accurately using only the time elapsing between the first diagnosis and death.

If the system is constructed in a reasonable way, the individuals’ event histories can be regarded as ‘paths’ through the system. Pictorially the system can be presented as a directed graph whose nodes and edges have characteristics which describe the properties of possible states and transitions in a system (Fig. 2).

This kind of presentation can be criticised as simplistic or even confusing, since it may seem that nothing happens to an individual as long as he or she remains in the same state. On the other hand, it has been found that this kind of formalisation of a problem is easily understandable even for technical outsiders and that it thus provides a common language for the experts from different fields which is required for effective co-operation between members of a research group.

As a matter of fact, the presentation technique described above does not restrict the number of possible states. If it seems too unrealistic to assume that nothing happens during a stay in a state or that stays in the same state are not comparable, it is not a problem to split a state into a greater number of more specific states. In general there is no limit to this kind of hierarchical splitting. The choice of an appropriate degree of hierarchical structure is a good example of a fundamental decision on how to reach a compromise between exactness of measurement, ambiguity of observation and – most importantly – the practical needs of the problem-solving process in question.

2.2. Understanding data

In register-based analyses, the data themselves impose restrictions on the whole problem-solving task. Such data are known as secondary data, in contrast to the primary data obtained by the investigator through direct observation or interviewing. The use of secondary data raises particular problems associated with the need to take account of the process by which the data were compiled. In general, data must have a form appropriate for storage and analysis purposes, and in addition for intuitive interpretation.
2.2.1. The nature of data

If we apply the traditional Shannon’s communication model in the context of administrative registers, the transmission of information in register utilisation can be regarded as having a structure such as that presented in Fig. 3 (cf. [31]).

The main idea is that some parts of the phenomenon of interest are observed and coded to signal, which is then transmitted using a possibly noisy channel into a database and then decoded in order to obtain proper data concerning the phenomenon. Even though this is a very simple and technical representation of communication, it seems to contain the essential elements needed in the common-sense understanding of secondary data. In other words, it is assumed that there exists some phenomenon which can be observed. Since it is impossible to completely observe all details or make exact measurements, some kind of coding is used to describe things. This coded signal is then stored in a database. The noise can be interpreted as an explanation for measurement compromises, possible inconsistencies and coding errors, and coding practices existing in the stored signal. When this signal is to be utilised, it must be decoded into understandable data, i.e. into a form which tells us something about the phenomenon. This phase is also subject to noise caused by incompatibility of choices and interpretations made by the data producer and the data user.

In other words, there are many problematic preconditions in the case of administrative data. Most importantly, the limitations of data are determined by the choices made by the producer of the data: easily available data are also easier to collect; data are produced using fixed categorisations and may be dependent on producers’ interpretations; there may be many data producers, with consequent variation in production practices and categorisations used; and information systems may not be flexible enough to take changes in phenomena and society into account. In practice, secondary data are interpreted in order to utilise them for purposes other than those for which they were originally intended. The most typical problems encountered in such a situation are the lack of data on important background phenomena and discrepancies between the users and producers of data as regards their beliefs and knowledge concerning the nature of data. It is possible to grasp the real nature of the data only if one understands (a) what factors have led to the production of certain data and a particular information system, (b) what are the sources and consequences of the type of organisation and data structures used, and (c) whether the data have an effect on the understanding of their domain [24,32].

2.2.2. Data structures

Event-history data consist of observations of the form \((\tau_k, D_k)\), where \(\tau_k\) is an ‘occurrence time’ and \(D_k\) is an ‘explanation’ for the event (and \(n\) is the number of observations and \(\tau_1 \leq \tau_2 \leq \ldots \tau_n\) and \(\tau_i < \tau_j\) for at least one observation \(i \neq j\) and \(i,j,k = 1,2,\ldots,n\)).

Usually \(D_k\) consists of a set of attributes (variables). However, not all the attributes necessarily contain important or interesting information. Some attributes can be irrelevant to the solving of particular problem or can be easily derived from other attributes. It is often reasonable to divide up the relevant information into two subsets of attributes. The first subset defines an event type \(E_k\) and the other includes important...
Let \( m \) be the number of distinct occurrence times, \( t_i \) be the \( i \)th distinct occurrence time (\( i = 1, 2, \ldots, m \)) and the event set \( A_i \) be the set of relevant information on observations occurring at the same time, i.e. \( A_i = \{(t_i, E_k, i_k)\} \), where \( i = 1, 2, \ldots, m \) and for every \( i, k \) is over the observations for which \( t_i = f_k(\tau_k) \).

The generalised event sequence \( S \) is defined to be a queue of event sets sorted by the (transformed) occurrence time, i.e. \( S = \langle A_1, A_2, \ldots, A_m \rangle \). In addition, let the conditional event sequence, to be called an event subsequence, be a sequence \( S_\theta = \langle A_i | A_i \in S \text{ and condition } \theta \text{ is true } \rangle \), where \( i = 1, \ldots, m \).

The definition of the generalised event sequence given above is not very rigorous and can be used in the absence of any knowledge about the event-history framework. However, if one uses this framework, generalised event sequences acquire a constructive and intuitively clear interpretation. Moreover, transformations allowed in generalised event sequences make it possible to use the same data structure in the implementation of different statistical and data mining methods.

### 2.2.3. An example of a generalised event sequence

Figure 4 shows an example of event-history data and the corresponding event sequence, which could be produced by system – say system P, presented in Fig. 2. Two graphical representations of this particular event history are also shown. The first is an event-history description of transitions in a system where an individual actually stays in a current state until there is a transition to another state. In the second, only the events occurring are marked on the figure. Since the time between two consecutive events is also the length of stay in a particular state, it is often very useful to include a ‘length of stay’ attribute in the covariate attributes \( i_k \), even though it can be easily calculated from the corresponding occurrence times.

In the definition of a generalised event sequence, parallel occurrences of events are allowed. In principle, the systems approach can be rendered valid in this case by defining each combination of event
sets as a ‘new’ event type. In other words, one can think of the explanation for an event type as a combined description of events occurring at the same time. However, in many cases it is reasonable to assign the parallel events to different event types, since it is possible to find ‘natural’ interpretations for these parallel events.

The data shown in Fig. 4 constitute a description of movements of individual X in a system P. Let Q be a system with two states (state F: ‘married’ and state G: ‘not married’). As a result, data corresponding to the movements of individual X in the system Q have the same form as the data in Fig. 4. Figure 5a shows an example of this kind of situation. Now there are two event histories for individual X, one corresponding to the path through the system P and the other to that through the system Q. In this case it is known that the systems which ‘generate’ the data are parallel, but the two event histories can be combined into one event sequence (Fig. 5b).

In the case of two parallel systems, the information concerning the ‘source system’ of an observation is a very valuable covariate. In general, there can, of course, be more than two data-generating systems. If there is a need to restrict analyses to the observations obtained from some particular system, this can be done easily by using an event subsequence conditional to the corresponding system.

In practice there are always event histories for more than one individual. It is thus trivial to include a covariate which identifies an individual, for example his or her social security number, while the whole data set still has the form of a generalised event sequence. Again it is possible to restrict the analyses to a pretermined set of observations by using event subsequences with suitable conditions.

2.2.4. Censoring

An additional point to consider is the fact that in practice, data constitute only a narrow window on the dynamics of a phenomenon. In other words, the observations contained in the data set fulfil the condition $a < t_i < b$, where $a$ and $b$ are finite constants and $t_i$ is the occurrence time of an event. The problems caused by a limited observation window are illustrated in Fig. 6, which shows examples of different types of ‘censoring’. Each case corresponds to an individual’s length of stay in some particular state, i.e. each case is combination of two events: transition to the state of interest and away from it. Dotted parts of the lines are unobserved; this corresponds to the case where the ‘true’ transition to or from the state is not observed. Censoring in fact guides the possibilities for analysis, and it must be taken into account in such analyses.

In the cases (a) and (g), there are no observed transitions to or from the state. This kind of censoring can be very problematic if the very first (or last) occurrence of some event type is considered more important than other occurrences (for example the first diagnosis of schizophrenia or the first back-surgery operation). In the case (b), the true transition to the state is not observed, but the individual was
Fig. 6. Examples of censoring.

in that state when the observation began, and the transition time from the state is known. In the case (c) both transitions are observed, but there is a potential problem, because the data outside the observation window are not complete for all individuals (typical situation for hospital discharge data). The case (d) corresponds to the uncensored observation. In the case (e), there has been an unknown or ‘wrong’ transition from the state (drop-outs or ‘competing risk’). It is also possible that the follow-up has ended before the occurrence of the event of interest (case f) or the transitions to and from the state are outside of the observation window (case h).

2.2.5. Statistical interpretations of a generalised event sequence

A generalised event sequence can be interpreted as a sample path of a marked point process, if the occurrence times for all events are distinct. This leads to a very general family of statistical hazard-rate models suitable for censored data (see e.g. [1]). In practice, all methods in the event-history framework are special cases of this general interpretation.

For example, there are many situations in which the ‘calendar time’ of event occurrence is not important, since the ‘real’ information is the time between two consecutive events. In other words, the ‘starting time’ of a follow-up can vary between individuals. A traditional solution, very common in survival analysis, is to transform the time axis from calendar time to ‘failure time’. This can be done using a transformation function \( f_k(\tau_k) = \tau_k - b_k \), where \( b_k \) is the occurrence time of ‘starting event’ of a corresponding individual and \( k = 1, 2, \ldots, n \) (see e.g. [7,18]).

Assuming that the probability of the next event type (state) depends only on time spent in the current state, an appropriate choice for a model is a semi-Markov-model (see e.g. [29]). By restricting the probabilities to change only at discrete time points, the semi-Markov model can be formulated using the Markov chain, where the state space is expanded in a proper way (see e.g. [14]). In the traditional Markov chain, the probability of the next event type (state) depends only on the current state (first-order Markov property), and the probabilities are time-homogeneous. In fact, the Markov chain interpretation corresponds to a situation in which the exact occurrence time is not important and only the order of observations matters. A generalised event sequence can be transformed to an event type sequence of this kind by using the transformation \( f_k(\tau_k) = k(k = 1, 2, \ldots, n) \).

Moreover, by choosing the transformation function \( f_k \) so that it is of the form \( f_k(\tau_k) = c \), where \( c \) is an arbitrary constant and \( k = 1, 2, \ldots, n \), the time dimension of the event sequence can be eliminated and the sequence reduces to the data miner’s classical market basket model (see e.g. [12], Chapter 6).
2.3. Data preprocessing

Usually massive second-hand data sets contain so much information and so many domain-specific features, inaccuracies and problems that raw data as such are not usable. In order to use data in problem solving, there must be understanding of the connections between the problem and the data. In register-based analyses, the problem, domain knowledge and data determine the most suitable model for the final problem solving. Typically it is possible to construct data sets of the event-history type using register data. Preprocessing can also be interpreted as a kind of technical operationalising phase in the research process.

2.3.1. Data abstraction

Often the connections between highly specific raw data and the highly abstract domain knowledge are so complicated that it is not possible to find any direct links between data and knowledge. An intelligent interpretation of raw data must be embedded into analyses, so that the resulting derived data set is at the level of abstraction corresponding to the current problem. Since noise is an unavoidable phenomenon, some kind of data validation and verification which makes use of knowledge should also be performed. This kind of task, performed in order to abstract higher-level concepts from possibly time-stamped data, is called data abstraction (see e.g. [20], Chapter 2; [30]). In the discovery of medical knowledge, data are usually patient-specific, while medical knowledge is patient-independent and consists of generalisations that apply across patients. For example, a complication after surgical operation is a medical concept, but from individual-based data it must be abstracted by using some ‘rules’, such as a list of some particular diagnosis codes recorded in data[RF6].

2.3.2. Data cleaning

Real-world data are very often more or less incomplete, noisy and inconsistent. Data cleaning involves detecting and removing errors and inconsistencies from data in order to improve the quality of data (see e.g. [25]). Compared to data abstraction, cleaning is more data-driven and technically oriented. In other words, corrections of erroneous and inconsistent codes as well as missing values can be usually made to the whole database, but data abstraction always results in problem-specific derived data sets.

2.3.3. Data integration and reduction

Two other common types of preprocessing are known as data integration and data reduction. The idea in data integration is to include data from multiple sources in analyses; the process is also known as record linkage (see e.g. [26,36]).

Data reduction results in a reduced representation of a data set which is much smaller in volume than the original data set, yet produces the same (or almost the same) analytical results (see e.g. [12], Chapter 3). Data reduction can in fact consist of anything from simple database queries to very complicated analyses.

2.3.4. Preprocessing tasks in the event-history framework

Sometimes it can in fact be difficult to distinguish what is operationalising, what is preprocessing and what is modelling. All these phases need interpretation and evaluation of the results. Since these are also highly domain-specific and problem-specific, it is difficult to give any general suggestions concerning tasks and methods for analyses, other than that the problem in question implicitly defines the most suitable technique and that the assumptions underlying techniques applied must be valid.
However, some non-trivial preprocessing tasks in the event-history framework could include: ‘defining’ the state space for some system (What happens to a patient after a surgical operation?); finding interesting and frequent combinations of patterns (What combinations of diagnosis and operation codes are frequent?); assigning patterns to adequate hierarchies (What diagnoses relate to complications of a surgical operation?); and confirming the expert’s ‘hypotheses’ about the phenomenon from data.

Most analysing techniques are feasible only in the case of moderately small data sets, since these typically need access to the whole data set, and the processing time will be directly proportional to the physical file size. In fact, many data mining methods are very well suited in such circumstances and are clearly something that is needed in practical data analysis. It can be stated that a sophisticated preprocessing operation incorporating non-technical domain knowledge in order to scale things down to a size fit for more specific statistical analyses is the most important and time-consuming part of register-based data analysis.

One very straightforward but extremely useful preprocessing technique is ‘pattern remapping’. The idea is to first ‘forget’ the time dimension in the generalised event sequence and to use a levelwise search (see e.g. [21]) to extract the frequent patterns from data, regardless of occurrence times. These patterns (such as a list of medical diagnoses) are then given to a domain expert, who can identify and make hierarchical groupings for the relevant patterns. In the remapping phase, interesting patterns are then assigned to appropriate event types simply by ‘renaming’ interesting patterns according to suggestions made by the domain expert. Finally, the time dimension is restored and records lacking interesting event types are removed. This abstraction usually results in a considerably reduced data set with an interpretation corresponding to the current problem.

3. Practical example: Hip-fracture surgery

This example is a simplified extract from a study, which aimed 1) to develop and implement register-based performance indicators to measure the effectiveness of surgical treatment of hip fracture and 2) to evaluate and compare the effectiveness of health-care providers. The complete results are reported elsewhere [27,33]. The study is a part of a larger project which aims to develop register-based methods for the measurement of effectiveness in specialised health care.

3.1. Defining the problem

The first task of the project was to build up a research group consisting of experts from different fields. The group defined the actual problem more meticulously: the idea was to identify all hip-fracture patients from the Finnish Health Care Register and follow the life events that they encountered after hip-fracture surgery according to register data.

A simplified system related to this particular problem is shown in Fig. 7. As can be seen, the first hip-fracture operation performed on an individual patient has a key role in the characterisation of this phenomenon. Actually, the state preceding the first hip-fracture operation also matters from the clinical point of view, since the patients coming from home are usually in better condition than patients who are already in residential or hospital care. However, the major interest is in the events and pathways of care following the first hip-fracture operation. In this case, these events are classified into four categories. If everything goes well, the patient should return home. Hip fracture is a serious condition for the elderly, and it can also be a starting point or catalyst for other problems which may result in the need for residential or hospital care. A hip-fracture operation may even be followed by fatal complications.
particular categorisation was chosen because the event types comprised the events of interest from the viewpoint of the original problem; and even more importantly, these life events are recorded in various registers.

For the sake of generality, the system’s formulation also allows there to be multiple events after hip-fracture surgery. As a matter of fact, the pathways of care are more interesting than single events in situations such as a cost-effectiveness evaluation. In the formulation of the system, only death is an absorbing state; all other states can be followed by any other state, i.e. only the death state ends a path in the system. Moreover, even though these four states are distinct, they are not necessarily independent. For example, death may be more probable after complications, or it may not be very likely that the patient will return home if long-term residential care is approved. The absorbing death state causes even more problems, because it is not possible to say of any individual that there is an increased risk of complications after death. However, it is possible to speculate what would have happened if death had not occurred. All in all, this kind of quite simple system definition seems to result in an extremely complicated competing risks model, and more simplifications are needed in the actual modelling.

3.2. Understanding and preprocessing data

A cohort of patients with hip fracture in 1998 or 1999 was identified in the Finnish Hospital Care Register using a simple diagnosis-group abstraction (all patients with at least one ICD-10: S72 diagnosis in 1998 or 1999). Using the unique personal identity codes of the patient cohort, data on all inpatient and outpatient hospital care and deaths for this cohort were obtained from the Finnish Health Care Register, the data warehouse of the Finnish Hospital Benchmarking Project and the National Causes of Death Register. The results of these straightforward database queries were integrated into a new data set containing 167,952 records for 17,099 patients.

Each record in this data set corresponds to one care episode in hospital (or death), not any actual event of the system, i.e. each observation includes information, such as patient and hospital ID-numbers, age, sex, area codes and diagnosis and operation codes, as well as dates of admission and discharge (or death). Data cleaning was performed in order to correct impossible simultaneous hospital episodes, systematic
errors in the use of symptom vs. cause diagnoses, and some missing or erroneous attribute values in area codes.

Many types of censoring occur in this data set: the first hip-fracture operation (or other important event) can be outside the observation window; some hospital episodes may have begun before 1998; follow-up finishes at the end of 1999 (there are census data available for the last day of every year in the Finnish Health Care Register); and follow-up may be terminated by the death of a patient.

Operation codes corresponding to hip-fracture surgery were abstracted into two different operation types. Using this and the diagnosis-group abstraction of hip fracture, hip-fracture operations were identified from the data. Since the state preceding a hip-fracture operation was also important, the histories of the patients were traced backwards and the preceding state was classified as home or residential/hospital care by using data abstraction of a more complicated nature.

The forward-direction abstractions were even more complicated, and all event types needed special abstractions and techniques. For example, the acute complication events were identified using the pattern remapping technique, in which all clinically relevant complication diagnoses were remapped to one event type.

3.3. Modelling, evaluation and reporting

For statistical modelling purposes it was assumed that any acute complication event occurring after a hip-fracture operation is an outcome which reflects the effectiveness of the surgical treatment. In addition, deaths and the upper limit of the observation window were assumed to cause censoring of the event of interest. With these assumptions, the modelling reduced to standard survival analysis (see e.g. [7]) where the variables of interest are the time between a hip-fracture operation and a complication or censoring event, and the censoring indicator. These variables were calculated for all of those patients who were (a) aged over 60 years and (b) had been living at home before surgery. The final preprocessed data set had 8824 records, each containing relevant variables for one patient.

In survival analysis, the distribution function of failure time random variable $T$ is $F(t) = P(T \leq t)$, where $t \geq 0$, and $S(t) = P(T > t) = 1 - F(t), t \geq 0$, is the corresponding survival function. Assuming $F$ to be absolute continuous and $f$ to be the corresponding density, the hazard function is defined by $r(t) = f(t)/S(t), t \geq 0$. The differential $r(t)dt = P(T \in dt | T \geq t)$ has the intuitive interpretation of ‘the conditional failure probability at time $t$, given survival to at least $t’$. Moreover, the hazard function determines uniquely the distribution function, and hazard-based models are often a convenient way to handle censored observations.

In spite of the fact that the time dimension includes a lot of information, the actual effectiveness indicators should be as easy as possible to interpret. In this study the continuous time scale was reduced to a simple dichotomous scale which merely indicates whether or not the event of interest has occurred in the case of an individual during some particular ‘limit’ time period. However, it is not obvious how such ‘limit’ times should be chosen. In addition to using clinical knowledge, it is also possible to utilise data empirically in this task. Since the outcome was an acute complication, the hazard function of acute complication occurrences was estimated. According to the hazard function presented in Fig. 8, the probability of acute complications was higher during the first 30-day period after a surgical operation. This finding based on the data corresponded to the domain knowledge and provided some evidence that the data abstraction was done in a proper way.

Since there are censored observations in the data, the proportion of the risk population for whom an event of interest occurred within the ‘limit’ time period does not necessarily correspond to the ‘true’
occurrence rate. In other words, rates must be estimated using more sophisticated methods. In this study the censoring was assumed to be independent (survival experience in the future is not statistically altered by censoring and survival experience in the past). The ‘limit’ time rates were estimated using product-limit (Kaplan-Meier) estimators. In addition, the counts of events were needed, and the corrected ‘observed’ counts were calculated using the relation \( O_i = N_i \hat{F}_i(t) \), where \( i \) indexes the risk population, \( N_i \) is the size of risk population \( i \) and \( \hat{F}_i(t) \) is the estimated cumulative probability of event occurrence at time \( t \) (‘limit’ time rate) in risk population \( i \).

3.4. Profiling the providers

The most useful information is obtained when rates are evaluated for health-care providers, such as hospitals or hospital districts. Profiling analyses of this kind also allow comparisons of effectiveness between providers. However, the profiling can be quite complicated, since there is variation between providers for at least three reasons: 1) differences may be attributable to random variation caused by the size of the provider, 2) the patient case-mix may vary from provider to provider, and 3) providers may differ in the effectiveness of their care. For these reasons, a statistical model for provider profiling in which provider differences are modelled explicitly was constructed in the study.

For an individual patient \( j \), an observed outcome has the form \( (Y_j|x_j, z_j) \), where \( x_j \) is a vector of patient characteristics and \( z_j \) is a provider-specific effectiveness component. The expected outcome for a patient is \( E(Y_j|x_j) \), i.e. a constant degree of effectiveness is assumed. Obviously, if the expected outcome is subtracted from that observed, the remaining residual reflects the effectiveness of care of a provider.

In the case of binary outcomes, a logistic regression is a suitable tool for the calculation of the expected outcomes. The idea is to construct and estimate a model in which the observed outcome is a dependent variable and patient characteristics are independent variables. Using this kind of model, it is possible to calculate predicted values for all individuals using patient characteristics and estimated values of parameters with the inverse logit transformation.
Since the focus in profiling is on providers and not on individuals, the observed and expected outcomes must be aggregated to the provider level, i.e. $O_i = \sum Y_j$ and $E_i = \sum \logit^{-1}(x_j\beta)$, where the sums are over patients treated by provider $i$ and $\beta$ is an estimated parameter vector.

Traditionally the ratio of observed to expected outcomes multiplied by the mean rate is used as a risk-adjusted rate for a provider. This quantity forms the basis for comparisons between providers and can be interpreted as the estimated event rate for a given provider if the population of patients treated were identical to the nationwide case-mix.

Since the observed outcomes $O_i$ are non-negative integers describing frequencies of events, they can be assumed to have a Poisson distribution with unknown mean $\mu_i$. That is, $O_i \sim \text{Poisson}(\mu_i)$, where $\log \mu_i = \log E_i + \theta_i$ and $i$ is the provider index. In other words, it is assumed that the expected outcomes $E_i$ adjust the patient characteristics, and $\theta_i$ describes the variation caused by provider. The use of logarithms guarantees that $\theta_i$ remains positive in the model.

Case-mix adjustment does not necessarily eliminate the variation in performance indicators in such a way that the indicators reflect a provider’s effectiveness of care. For example, in data sets with a hierarchical structure there often exists correlations between observations, and this may result in overestimated differences in profiling analyses. Differences in the sizes of providers may also cause problems. For example, rates can not be estimated accurately for small providers.

Assuming the exchangeability of providers (i.e. that the results for all providers are equal if there is infinite number of (similar) patients), a two-level hierarchical model can be used to solve the problems mentioned above. A simple solution is to assume that variation caused by providers is normally distributed, i.e. $\theta_i \sim N(\alpha, \sigma^2)$, where $\exp(\alpha)$ is the ‘general’ case-mix-adjusted risk ratio and $\sigma^2$ describes the variance between providers (in logarithmic scale). This kind of hierarchical Bayes model needs appropriate prior distributions for the hyperparameters $\alpha$ and $\sigma^2$, such as $\alpha \sim N(0, 10^6)$ and $\sigma^{-2} \sim \Gamma(0.001, 0.001)$. The estimation of posterior distributions for parameters of interest (in this case $\theta_i$s for all providers $i$) can be done using Markov chain Monte Carlo (MCMC) simulation [22].

Multiplying the obtained adjusted risk ratios by the mean rate results in easily interpretable risk-adjusted rates for providers. The use of a hierarchical multilevel model eliminates many drawbacks of traditional profiling analyses. Multilevel models are well suited to the simultaneous calculation of many confidence intervals (multiple comparisons problem), and they give more conservative estimates for differences between providers than do traditional methods [11].

3.4.1. Example drawn from the results

Figure 9 shows the hospital-district-specific risk-adjusted rates for 30-day complication rates. The mean 30-day acute complication rate was 13.0%, and there were no significant differences between districts according to the 95% confidence intervals of rates. However, according to the 50% confidence intervals, hospital districts 9 and 15 seem to have slightly higher rates than others. For more results, see [27] and [33].

4. Conclusions

This paper has presented a methodological framework for the utilisation of administrative registers in the creation of scientifically valid information. This has been done by discussing essential methodological criteria encountered in the practical research process and by combining fruitful methodological ideas from different fields, such as statistics, data mining and sociology. The emphasis has been on the understanding of connections between problem, data and analysis in the case of secondary data sources.
In the introduction, the problem caused by a changing paradigm of data analysis was discussed. Different perspectives are needed in analyses. Bringing carefully defined ideas originating from some field into the body of knowledge concerning other fields may create new possibilities or solve some problems – if the scholars working in the ‘object field’ are open-minded enough. Moreover, some pieces of knowledge, such as the criteria for scientific information and the phases of research processes, may be ‘self-evident’ issues for many researchers and data analysts, but the listing of these principles is useful in any case, since it helps to anticipate and avoid the most crucial pitfalls.

The paper contains some key points concerning the research process when it is based on massive secondary data sources. The secondary nature and massive size of data sets highlight the importance of scientific issues as compared to the technical questions. Effective research utilising massive secondary data sources requires broad expertise and active collaboration, since it is waste of time to reinvent things which are trivial to some other scholars. This can easily lead to the problem that there is no common language shared by experts in different fields. The ‘slaves of ideologies’ think differently and have different latent assumptions, so it can be difficult to find suitable compromises needed in order to achieve reasonable results. An even bigger mistake is to forget the possibly major impact that decisions made during the execution of a study can have on its outcomes. In addition, beliefs concerning the nature of secondary data are not the same for producers and users, and this generates unknown amounts of ‘noise’ for the results if such issues are not considered carefully. Sophisticated preprocessing – incorporating knowledge from the non-technical domain in order to scale things down to a size fit for more detailed

Fig. 9. Hospital district-specific risk-adjusted 30-day complication rates in 1998 and 1999. Rates as complications per 100 patients. Thin parts of lines correspond to 95% confidence intervals and thick parts correspond to 50% confidence intervals. Mean rate is 13 complications per 100 patients (dotted line).
statistical analyses – is the most important and time-consuming element in the register-based data analysis. All in all, it can be stated that ‘data are not collected, but produced; research results are not findings, but creations’ [17].

There is a great deal of literature concerning the utilisation of register-based data in research (see e.g. [6,10,16,23,28,35,37]). However, the perspective is usually very closely connected to the domain field or to some particular problem, which makes it difficult to extract more generally applicable ideas. Moreover, a purely problem-oriented approach can easily lead to a situation in which the contents of the data are taken too literally. For example, one pitfall is to think that medical diagnoses are recorded in administrative registers in just as detailed a manner as the diagnosis classification allows.

There are also two very common erroneous beliefs concerning the utilisation of register data in research. It has been stated that the formation of research data is often technically easy, and that register data are ‘hard’ data representing the truth and providing evidence-based quantitative information. In spite of the fact that time-consuming and expensive production of new data is avoided, according to the principles presented above, the data preprocessing, which is full of ideologically dependent qualitative choices, is the most fundamental and the most difficult part of register-based research. Data analysis is never ‘easy record linking’. It is at least as important a part of the research as the domain knowledge. All data analysts should remember that they are not just assistants to ‘real researchers’ but full and equal members of the research group. Collaboration is the key to obtaining the results. Nobody can be an expert on everything.

The principles of the research process that were presented above are the basis for the actual utilisation framework. Since the ideas were somewhat abstract in nature, some suggestions for concrete and practically useful interpretations were considered more carefully. The event-history framework was suggested as a well-developed option in outlining the characteristics of dynamic phenomena. A systems approach to the operationalisation of dynamic phenomenon was described as a helpful tool in generating a common language for experts from different fields. The structure of event-history data was shown to have the form of a generalised event sequence and to be a suitable data structure for the handling of second-hand register data. We drew attention to aspects of censoring which, in connection with event-history data, implicitly define possible model types for problem solving and provide a glimpse of the data-generating process, which has an intuitive interpretation in the event-history framework. Moreover, a wide variety of suitable data structures for traditional models can be produced from the generalised event sequence as special cases using the time-transformation property presented. In addition, common preprocessing types were reviewed, some examples of possible preprocessing tasks were given and a pattern remapping technique was formulated. Finally the ideas were illustrated by means of a practical example. It was seen that register-based data analysis becomes very complicated and challenging even in seemingly simple situations.

All in all, this paper is simply a monologue presented by a statistician who is specialised in computational statistics and has a background in the social sciences, research interests in problems related to health services research, and experience in the production of official statistics, the maintenance of information systems, the development of statistical software, the teaching of statistics, and consultation with other researchers. In other words, the issues presented in this article are just an ‘insider’s’ collection of pieces of preliminary knowledge which are essential to the conduct of scientific research based on administrative registers.

Acknowledgements

This study is supported in part by research grant from the Academy of Finland.
References


Comparing Properties of Audit Data and Routinely Collected Register Data in Case of Performance Assessment of Hip Fracture Treatment in Finland

R. Sund, I. Nurmi-Lüthje, P. Lüthje, S. Tanninen, A. Narinen, I. Keskimäki
1National Research and Development Centre for Welfare and Health, Helsinki, Finland
2Health Centre of Kuopio Region, Kuopio, Finland
3Kuusankoski Regional Hospital, Kuusankoski, Finland

Summary
Objectives: To compare prospective audit data and secondary administrative register data in the production of performance assessment information in the case of hip fracture treatment, and to cross-validate the quality of information.

Methods: First, a conceptual model for the performance assessment of hip fracture treatment was defined. This model was then utilized in comparisons between the prospective audit data concerning 106 consecutive hip fracture patients from the Kuusankoski Regional Hospital and corresponding register data from the Finnish Health Care Register and the Causes of Death Register. We examined the completeness of registration of patients and also the accuracy and degree of completeness of the registered data. Observed differences were checked against the medical records.

Results: Register data lack clinical detail, but outperform prospective data in the recording of inpatient care history. Completeness of the register data is very good. The accuracy of easily measurable variables in the register is at least 95%. The agreement between register and audit data was 86.3% for detailed hip fracture diagnoses. Polynomial correlation between the functional dependency variables was 0.68.

Conclusions: Register and audit data have certain limitations and problems, but both seem to be suitable for the performance assessment of hip fracture treatment. To improve the feasibility of the administrative register data, the voluntary input of additional hip fracture event data to the register should be made possible. Standardized instructions that guide the use of available register classifications in a sensible way would improve the quality of data.

Keywords
Validity, completeness, accuracy, information systems, conceptual model

Introduction
The theoretical principles of the performance assessment of health systems have been widely studied [1, 2]. Performance assessment requires a multidisciplinary approach and is closely connected to issues such as quality of health care [3, 4], systems analysis [5], knowledge management [6], and decision support systems [7]. Continous improvement techniques are required in the design and management of information systems suitable for performance assessment [8, 9]. There exist several methodological challenges: complicated interactions between the physiological onset of disease, health system, treatment decisions, human behavior, society and expenditures need to be integrated under the same theoretical model [10, 11]. In practice, the most useful approach to performance assessment seems to be a disease-based comparison of episodes of care [12, 13]. Relevant performance criteria can be extracted from the evidence-based treatment guidelines [14]. Typically comparisons are made between the providers of care [15], though it requires the adjustment of known risk factors that may disturb the comparisons [16-18].

Practical and reliable information on evidence-based performance assessment in the case of hip fracture treatment is becoming increasingy necessary [19]. Good experiences of the Swedish Rikshöft-registry have led to the formalization of data production for the standardized audit of hip fractures in Europe (SAHFE) [20]. Data complying with the SAHFE standards have also been produced in some hospitals in Finland [21-23]. This kind of prospective collection of clinical data obviously represents the practical “state-of-the-art” consensus for data requirements for assessing the treatment of hip fractures. Unfortunately, the separate data production requires extra work and resources, and therefore it is unlikely that the extensive voluntary data collection required by SAHFE type audit data would be feasible in all hospitals. It would be practical and cost-effective if routinely collected administrative data could be used for performance assessment purposes [24]. In fact, in Finland administrative registers have been utilized in performance assessment in the case of hip fractures [25-27].

In principle, the Finnish health registers offer a very attractive and flexible environment for research purposes, because the universal personal identification numbers are used in all registers [28]. Deterministic record linkage can be used and the potential difficulties with complex probabilistic linkages are thus avoided [29-31]. Also the accuracy of the most important variables in the Finnish registers is known to be good [32-35]. Unfortunately, the available evidence concerning the validity of register data can not be directly generalized to the case of hip fracture. One study of acute pelvic fractures in Finland during 1988 has investigated the correspondence between the register data and the medical records of 114 patients (10% of 1212 patients) [36]. The accuracy of the data of the register was found to be at least 95% for most important variables excluding secondary diagnosis (80%) and place of injury (75%). In another study investigating hip fracture incidence in
one health care district from 1982-3 and from 1992-3 in Finland, the data on hip fracture patients (n = 668) identified from hospital files was corrected using information from medical records [37]. The hospital discharge register missed 6.6% of patients in 1982-3 and 2.6% in 1992-3, and a high percentage of false or multiple data were found (41.8% in 1982-3 and 21.7% in 1992-3). Besides these, we do not know of any other Finnish studies that analyze the accuracy or completeness of the national hospital discharge register in regard to fracture patients.

From a general perspective, accuracy and completeness can not be defined uniquely without an assumption of a golden standard of measurement, and data validity must also be judged against the intended utilization purposes [38, 39]. In studying incidence, the definition of a golden standard is straightforward, but this is not the case in performance assessment. Fortunately, the data requirements for incidence calculation and for determining the starting points of care episodes in performance assessment are very similar [40]. This resemblance between two approaches implicitly shows that assumptions of a golden standard are very simplistic in incidence studies: certain observable facts are measured at a fixed time point. In performance assessment the main interest is on the dynamic processes of care, which essentially means dealing with different perspectives of measurement changing in time [41]. For instance, it is obvious that the detailed diagnosis of hip fracture is more important in the operating room than during the final stages of rehabilitation, and an assumption of constant accuracy is practically unrealistic. Moreover, the actual diagnosis of hip fracture is only an interpretation even after seeing x-rays, which suggests that the assumption of a golden standard may be erroneous [42]. This potential ambiguity is clearer in other diagnoses such as dementia or schizophrenia, and inevitable in the cases of more complex concepts such as post-operative complication or health status [43]. These problems become even more concrete and critical, if secondary data – in other words data collected originally for some other purpose – are to be used [24, 26, 44].

To make any statements concerning the quality of (secondary) data, it is essential to first outline the properties of required measures carefully under a fixed conceptual model (determine how data result from the theory), and then evaluate the observed data against these requirements (examine how well the theory can be reconstructed using the properties of actual observed data). In other words, to analyze the quality of available data for performance assessment a pragmatically useful compromise between problem-oriented and data-driven theories is required. In this sense, the definition for a golden standard of measurement becomes hermeneutic: the key issue is to understand why the data are like they are without fixing the reality using any single data source. However, it can be assumed that primary data collected for hip fracture audit purposes correspond more closely to any reasonable golden standard than secondary register data that were originally collected for other purposes.

Objectives

The aims of this study were to define a conceptual model for producing data for performance assessment in the case of hip fracture treatment, to compare two different sources of data – prospective clinical audit data (designed for monitoring treatment quality) and secondary administrative register data (used mainly in compiling statistics) – in the production of performance assessment information, and also to cross-validate the quality of these data sources.

Methods

For the definition of the conceptual model, the list of SAHFE-variables was chosen to represent an adequate base for data requirements (http://www.sahfe.ort.lu.se/guide.html). These are actual operationalizations of those patient level concepts considered important in the case of hip fracture. In this study, the dimensions of the conceptual model were abstracted from the actual operationalized measures, and then complemented using theoretical models of performance assessment [1, 18, 45, 46]. Each phenomenon represented by a concept was assumed to be attributable to an individual on a continuous time scale. It was also assumed that each concept can be described using a systems approach with a (limited or unlimited) number of states so that the system is always in some of these states at each time point. The concepts were then classified into more general groups in terms of their contextual interpretation, temporal stability and theoretical measurement properties. This qualitative classification procedure was repeated until the resulting conceptual model was considered satisfactory in the sense of Occam’s razor principle.

Audit data were collected prospectively for 106 consecutive hip fracture patients in the Kusankoski Regional Hospital between January 1, 1999 and January 31, 2000 [22, 23]. Patients treated in the surgical ward at the Kusankoski Regional Hospital during the same period were identified from the Finnish Health Care Register, and all records of these patients from 1987-2002 were extracted from the Finnish Hospital Discharge Register, the Finnish Health Care Register, and the Causes of Death Register using the unique personal identification numbers of the patient population. Detected differences between data sources were further checked manually from the medical records of these patients (PL). The ethical committee of the Kymenlaakso Hospital District approved the study. Permission to use the data was obtained from the Kusankoski Regional Hospital, the National Research and Development Centre for Welfare and Health, and also from Statistics Finland.

The selected measures of different data sources were first matched to the defined conceptual model. The completeness of registration of hip fracture patients as well as the accuracy and degree of completeness of the registered data were examined in both the audit and register data using standard methodology [38]. For some variables it was more reasonable to measure agreement than accuracy. This required approaches based on modeling the decision-making process, such as a polychoric correlation model (a form of latent trait model) [47].
tion of agreements that were not due to chance was measured using a delta coefficient, which is based on the model of multiple-choice tests and avoids certain known deficiencies of the kappa coefficient [48]. Kappa may be considered as an approximation of delta, as both yield values that are very similar when at least one marginal distribution is balanced or when both marginal distributions are moderately unbalanced in the same direction [49].

Results

The dimensions of the conceptual model and selected measures attributable to these dimensions in various data sources are reported in Table 1. Of the dimensions, the biological constants do not change in time and therefore one measurement is generalizable to all times. Biological events, accident/fall history, and hip fracture event represent dimensions for which actual values of measures are recorded in the proximity of some observable event. All other dimensions relate to phenomena that potentially change in time and should be continuously monitored. In practice, time must be fixed for actual measurement. In the audit data, the measurement is done in connection with a hip fracture event and for certain measures also two weeks, four months, and one year after the fracture. In the register data, the recording takes place at each discharge. As expected, the content of audit data is richer than in the register data, even after using indirect measures of certain concepts. However, the measurement frequency in register data is superior to audit data, allowing complete observations of event histories for inpatient care.

Completeness of Registration

There were 104 patients in the prospective data (two had a new fracture at the other side of the hip during the follow-up). In the register there were 111 patients who were admitted to the surgical inpatient ward at the Kuusankoski Regional Hospital during the study period with a hip fracture diagnosis. It turned out that there were in total 105 patients with a confirmed hip fracture diagnosis (Table 2). The audit data missed one case and the register data two cases indicating very good completeness. The patient missing from the audit data had an impacted hip fracture (femoral neck) and was erroneously excluded from the prospective study. Both cases missing from the register data had a diagnosis of distal femur fracture recorded in the register. Three extra hip fracture patients with clearly false hip fracture diagnoses were found in the register. For the first patient, there was a coding error in the medical record diagnosis (S42.0 → S72.0) and the second patient had a distal femur fracture, but for the third patient any explanation for an obvious error was not found. One patient with hip fracture diagnosis in the register had a periprosthetic hip fracture and therefore this patient was not a hip fracture patient. Another patient was admitted because of suspected hip fracture, but no fracture was found. Three patients excluded from the audit data were having on-going treatment for a recent earlier hip fracture – one operated on at the Kuusankoski Regional Hospital and two at some other hospital. It also turned out that four patients included in the audit data had had – according to the register – an earlier hip fracture during the preceding ten years. The side of the fracture is recorded very seldom in the register, but the medical records revealed that only one of these patients had the new fracture on the same side as the earlier fracture. The positive agreement between register and audit data was very good (94.9%), and would be even higher if appropriate data abstraction rules for identification of false positives are defined. If data abstraction rules are not used, the register data will overestimate the number of new hip fractures.

Completeness of Registration for Re-operations and New Fractures

Two of the 105 hip fracture patients needed an acute re-operation during the first admission. Neither of these re-operations was detectable from the register data. In fact, the data structure of the Finnish Health Care Register allows only one operation day to be recorded during one period, and recording of two similar operation codes for the same period may seem to be an error. In both cases, the data in the register corresponded to the latter operation.

Three patients with a new hip fracture during the first year of follow-up were detected from the register. Two were also found from the audit data and the third was confirmed using the information in the medical record. The (confirmed) register data also revealed that one of these three patients had a third operation because of problems with the hip prosthesis during the first year of follow-up.

Accuracy and Completeness of Easily Measurable Variables

The completeness and accuracy of the register data were examined using 106 hip fracture events with complete audit data. Completeness indicates the percentage of cases with a recorded value for the variable in question. Accuracy tells the proportion of correctly recorded values for cases having at least some recorded values. The correctness – telling the overall utility (measured as the product of completeness and accuracy) of the variable – was very good for most of the variables (Table 3). The accuracy of admission source and discharge destination was improved by using record linkage instead of variables in the index admission period. Discharge day also required record linkage, because each transfer from one ward to another even in the same hospital results in a new discharge in the register. Missing operation codes are a nuisance in the case of hip fractures, because they may also indicate conservative treatment. The data structure of the register does not suit well to recording acute re-operations during the same hospital period, resulting in the recording of wrong codes and days for primary operation. The extra operation code indicating the side of fracture was used very rarely. Accuracy for the place of accident was poor. This was mainly because of incapability of the ICD-10 classification to
### Table 1 Conceptual model dimensions and their realizations in different data sources

<table>
<thead>
<tr>
<th>Category</th>
<th>SAHFE measure</th>
<th>Audit data measure</th>
<th>Register data measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological facts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biological events</td>
<td>date of birth, death, year of menopause, menarche</td>
<td>date of birth and death</td>
<td>date of birth and death</td>
</tr>
<tr>
<td>Biological constants (demographic)</td>
<td>sex</td>
<td>see</td>
<td>see</td>
</tr>
<tr>
<td>Biological measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical examinations</td>
<td>height, weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory tests</td>
<td>hemoglobin, creatine, albumen, bone density</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographic history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household composition</td>
<td>living alone</td>
<td>marital status, living alone</td>
<td></td>
</tr>
<tr>
<td>Place of living</td>
<td>residential status</td>
<td>residential status, area of living</td>
<td>area of living, level of care</td>
</tr>
<tr>
<td>Socioeconomic history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education, occupation, economic resources</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health related behavior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk behavior</td>
<td>smoking, alcohol intake</td>
<td>alcoholic usage</td>
<td>alcohol related inpatient care</td>
</tr>
<tr>
<td>Diet, activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective quality of life</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall well-being</td>
<td>pain, psychological state, fear of fall</td>
<td>Pain</td>
<td></td>
</tr>
<tr>
<td>Objective need for care</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases/symptoms</td>
<td>comorbidity, complications</td>
<td>Complications</td>
<td>recorded diagnoses, use of care</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>walking</td>
<td>walking, activities of daily living</td>
<td>need for care</td>
</tr>
<tr>
<td>Cognitive functioning</td>
<td>Abbreviated mental test score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of care</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical aid</td>
<td>walking aids</td>
<td>use of assistive technology, changes in living environment</td>
<td></td>
</tr>
<tr>
<td>Medication</td>
<td></td>
<td>use of painkillers</td>
<td></td>
</tr>
<tr>
<td>Health care utilization</td>
<td>provider, type of stay, length of stay</td>
<td>type of care, rehabilitation, dates for hospital visits</td>
<td>provider, type of care, admission and discharge dates, diagnoses, operations</td>
</tr>
<tr>
<td>Accident/hospitalization history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Properties of fall</td>
<td>date of fall, place of fall</td>
<td>date of fall, place of fall</td>
<td>place of and reason for hospitalized injury</td>
</tr>
<tr>
<td>Hip fracture event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Properties of fracture</td>
<td>time of fracture, occurrence place, fracture type, fracture side</td>
<td>day of fracture, reason for fracture, occurrence place, fracture type, fracture side</td>
<td>place of and reason for hospitalized injury</td>
</tr>
<tr>
<td>Initial treatment process</td>
<td>hospital id, admission time, admission source, start of operation, reason for operative delay, type of surgeon and anesthetist, type of anesthetic, operative method, length of surgery, type of prophylaxis, complications in operations, reoperations, time to mobilization, discharge time, discharge destination</td>
<td>admission day, admission source, day of operation, reason for operative delay, operative method, complications in operation, reoperations, discharge day, discharge destination</td>
<td>hospital id, admission day, admission source, diagnoses, day of operation, operation codes, discharge day, discharge destination</td>
</tr>
<tr>
<td>Clinical stability</td>
<td>ASA grade, comorbidities, existing fractures, reason for operative delay in medically unfit patients</td>
<td>ASA grade, reason for operative delay in medically unfit patients</td>
<td>recorded diagnoses, use of care</td>
</tr>
</tbody>
</table>

1) Indirect measurement requiring record linkage
In both data sets:  
98 new hip fracture patients  
4 patients with earlier fracture  
Missing from audit data:  
1 verified new patient  
2 operated in other hospitals  
1 recent earlier fracture  
1 earlier total prophylaxis  
1 suspected hip fracture  
3 false-positive register diagnoses  
Missing from register:  
2 false-negative register diagnoses  
Positive agreement:  
94.9% (95% CI: 90.3-96.8%)  
Overestimation in the register:  
7.2% (95% CI: 3.2-13.7%)  
Underestimation in the register:  
1.9% (95% CI: 0.2-6.7%)

Table 2  
Completeness of data sets  

Table 3  
Completeness and accuracy of easily measurable variables in the register

Table 4  
Accuracy of hip fracture diagnosis in the register

Accuracy and Reliability of Hip Fracture Diagnoses

Register diagnosis was recorded as hip fracture (ICD-10: S72.0, S72.1, S72.2) in 98.1% (95% CI: 93.4-99.8%) of cases (Table 3). However, the evaluation of the exact accuracy of the main diagnosis was not straightforward, as different classifications of hip fractures were used in the register and in the audit data. Following the methodological suggestions, the accuracy of diagnoses was addressed by considering the bone anatomy as a biological entity identifying true fracture status [50].

Hip fractures are commonly classified into intra- and extra-capsular according to their relationship to the capsular attachment of the hip. In terms of current data, fractures of the neck of the femur are intra-capsular, and other hip fractures are extra-capsular. The bone anatomy may be used to subdivide extra-capsular fractures further into basocervical, trochanteric and subtrochanteric fractures. Unfortunately, there is no separate diagnosis code in ICD-10 for basocervical fracture. Using a detailed fracture classification of femoral neck fracture, trochanteric fracture or subtrochanteric fracture, the agreement between audit data and register was 86.3% (95% CI: 79.4-92.2%). Pairwise agreement analyses revealed that femoral neck and subtrochanteric fractures were correctly identified in the register (high sensitivity), but identification of trochanteric fractures was not as accurate (Table 4). This is most probably because the extremities are easier to identify correctly. This interpretation is supported by the fact that the register diagnoses of trochanteric fracture were practically always true trochanteric fractures (high positive predictive value). This was not the case for femoral neck fractures or subtrochanteric register-diagnoses which were occasionally used for true trochanteric fractures (low positive predictive value). Missclassification at the borders of the trochanteric region of the femur may be one reason. Another cause is the ICD-10 classification utilized in Finland, which suggests the use of femoral neck diagnosis for unspecified hip fractures. The proportion of agreements that were not due to chance was very good as measured with the delta coefficient (81.3%; 95% CI: 71.3-91.3%) or the kappa coefficient (73.5%; 95% CI: 61.2-85.7%). In addition, since it was reasonable to assume that the underlying trait is continuous, the poly-

separate home accidents from the accidents occurring in residential care and falls occurring outdoors from the ones occurring indoors.
choric correlation coefficient was calculated indicating a very high consistency between the audit data and register (0.92; 95% CI: 0.86-0.98).

**Agreement between Functional Dependency Measures**

In the audit data, functional dependency was measured using an activities of daily living (ADL) type of scale [23]. In the Finnish Health Care Register, overall dependency (and need for care) is recorded using a scale of five categories [51]. The completeness of dependency variable in the register was 100%. The boxplots describing dependency measures (at discharge) are shown in Figure 1. The upper categories of the register-based measure are combined due to the small number of observations. The polyserial correlation between the variables is 0.68 (95% CI: 0.55-0.81), showing rather good consistency. Means and medians behave reasonably, but Figure 1 indicates that the most commonly used category in the register data has a very wide range. It seems that high register categorizations truly reflect bad functionality, but absolute interpretation of the register scale at an individual level may be erroneous.

**Follow-up Information**

State diagrams for the register and audit data summarizing the follow-up information are shown in Figure 2. The overall shapes are pretty similar, but the register data have more accuracy. The dates of deaths, lengths of initial hospitalizations, and residential statuses at four months and at one year after the fracture in the audit data are accurate, but the follow-up information concerning acute care subsequent to the initial hospitalization in the audit data is partly incomplete and therefore technically difficult to handle. This can be most clearly seen from the proportion of patients in nursing homes, which is first overestimated and later on underestimated (Fig. 2).

**Discussion**

In this study two forms of data production for performance assessment purposes were compared in the case of hip fracture treatment. A justifiable comparison required a definition of the conceptual model, which allowed for a systematic structuring of links between observable measures and theoretical concepts. Such models are only rarely reported [46], even though the explicit introduction of a conceptual model obviously significantly improves the understanding of the problem and makes the most important assumptions affecting the actual results of analyses visible.

Systematic approaches can be used for the development of a conceptual model [52]. Quite often conceptual models had to be represented so that it becomes possible to consider several different perspectives simultaneously [53-56]. This is particularly challenging if data are used for other purposes than originally intended [57]. In this study, a pragmatic mapping between concepts from a performance assessment theory, SAHFE metadata, and structured data entry in prospective audit data and secondary register data was constructed. In spite of the generality of the current model, it is limited in the sense that it deals with individual level concepts only. For benchmarking purposes, provider level phenomena that give explanations of observed differences between providers of care should also be considered. One very important individual level measure – namely costs – is also missing from the model. It would have special measurement properties, because it is cumulative in time.

As expected, the register data lacked clinical detail. However, most of the more detailed measures in the audit data were ac-
usually related to phenomena that continuously change in time. Such data require tacit knowledge to be useful in a more general context [58]. On the other hand, register data have a data structure that allows the complete observation of inpatient care history, and therefore outperforms prospective data in this sense if some delay in data production is tolerable. Making use of demographic, socioeconomic and medication histories data – which are, in principle, available in other Finnish registers – could also improve certain details of register data, but it requires extensive work to obtain and preprocess such – possibly expensive – data.

In this study, it was assumed that all hip fracture patients were identifiable from the register or audit data. This assumption is valid so far as patients were treated at the orthopedic inpatient ward at the Kuusankoski Regional Hospital, because it is very unlikely that two complete, but mutually different registrations would miss a patient. In fact, even the patients treated conservatively or operated on at another hospital are routinely referred to an orthopedic inpatient ward and would be registered. In theory, some patients admitted to the outpatient emergency department who were not resident in the operation area of the Kuusankoski Regional Hospital and thus directly transferred to their own local hospital, could have been missed from the data. However, in our study capture-recapture analyses estimating the total number of cases were not applicable, because there was structural dependency between data sources (the occurrence of a patient in audit data also indicated expected occurrence in register data) [59, 60]. Therefore the completeness was evaluated by identifying all hip fracture cases from audit data and register data, and then all incompatibilities were checked against the information in medical records. In this sense, the differences in completeness between data sources indicate the different definitions of hip fracture or errors in register data.

The completeness of audit data was excellent, and also the completeness of register data seemed to be very good. In fact, there is a tendency to overestimation if no appropriate data abstraction rules are used. For example, it is more reasonable to count new hip fracture cases than every patient with an ongoing hip-fracture-related care episode [61, 62]. However, unless the clinical judgment of a hip fracture case is accurately and completely recorded, register-based estimates for the numbers of hip fractures are prone to bias. The appropriateness of such data abstraction rules can be partially verified by comparing the hospital-specific numbers of consecutive hip fractures during particular time periods – available from prospective studies – to register-data-based estimates for the same hospitals and the same time periods. It must be noted that even this kind of a definition has some drawbacks in incidence calculations. The number of patients is hospital-specific, but the available risk-population data refers to geographical areas and one can therefore argue that patients living in the same area but operated in some other hospital should also be counted as cases, and patients from other areas should be excluded.

Our definition for a golden standard of measurement in this study was hermeneutic. In practice, three types of comparisons between data sources were utilized (even though the boundaries between types are not clear cut). The first type was applied in accuracy comparisons: the same rather easily measurable variable was available in both data sources, and detected differences could be further checked from the medical records (e.g. Table 3). The second type was for the comparison of theoretically similar information that has substantially different operationalizations in both data sources (e.g. Figs. 1 and 2). The third type of comparison was a combination of the first two types: both approaches can be applied (e.g. analyses of the detailed fracture classification).
The key idea was to make versatile comparisons that are methodologically justified and practically relevant.

The accuracy of the most easily measurable variables in the register is very good. However, the data structure for the recording of performed operations is not optimal in the register. For example, only one operation day is allowed for one hospitalization even though the patient may be re-operated on during the same hospitalization. The content of these fields could also be improved by giving standardized instructions that guide practitioners to also record the relevant additional operation codes available in the classification, such as which side the fracture is on. Such additional metadata should also be feasible to use to improve the quality of register data in general in cases where the use of structured data entries is known to vary between coders [39, 63, 64].

For the purposes of performance assessment, follow-up information is particularly essential. Mortality and length of initial hospitalization periods were recorded equally in both data sets, but otherwise it was not easy to derive information that was exactly compatible. Information on the use of inpatient health services on a daily basis was available directly from the register while deriving similar information from the audit data required certain assumptions and interpolation. The collection of follow-up information from the audit data requires extra work and is prone to different biases as is the routine collection of register data. Even though the use of health services is recorded better in the register, other measurements of useful concepts such as use of technical aid, functional disability and pain offer information that is not directly available from the register. However, because the register includes a disability measure, it actually offers complementary information (at admission, discharge and census points). This information would be more valuable if some commonly used functional disability or quality of life measure had been available in the register instead of a non-validated ad hoc measure [65].

On the other hand, the register includes information concerning the provider-specific use of health services with diagnostic-related group information for the total patient population. This allows the calculation of costs by summing up the suitably (price-) weighted information on the use of health services. This same idea for calculating costs can be further utilized, though the weights this time correspond not to prices but to the mean health or disability status of persons receiving specific forms of care, and this results in health status measure which is unbiased (in the same sense as costs can be unbiased) at the population level. Examination of such ideas is out of the scope of this study, but may offer fruitful directions for further research.

As a conclusion, register and audit data both have certain limitations and problems, but seem to be suitable for the performance assessment of hip fracture treatment. Voluntary input of additional hip fracture event data to the register should be made possible, because the best compromise for practical performance assessment purposes would be reached by a prospective recording of hip fracture event data combined with extraction of care histories of these patients from the registers.

Acknowledgments

This work was financially supported by a research grant from the Academy of Finland. The funding agreement ensured the authors’ independence in designing the study, interpreting the data, and writing and publishing the report.

References

13. Grady ML, Schwartz HA, editors. Medical effectiveness research data methods. Rockville: Agen-
20. Parker MJ, Currie CT, Mountain JA, Thongre
Analytic Perspective

Utilization of routinely collected administrative data in monitoring the incidence of aging dependent hip fracture

Reijo Sund*

Address: National Research and Development Centre for Welfare and Health (STAKES), Health Services and Policy Research, P.O. Box 220, FIN-00531 Helsinki, Finland
Email: Reijo Sund* - reijo.sund@stakes.fi

* Corresponding author

Abstract

Societies are facing challenges as the public health burden increases in tandem with population aging. Local information systems are needed that would allow a continuous monitoring of the incidence and effectiveness of treatments. This study investigates the possibilities of routinely collected administrative data as a data source for hip fracture incidence monitoring in Finland.

The study demonstrates that a straightforward use of register data results in biased estimates for the numbers of hip fractures. An interpretation of hip fractures from the population aging point of view offers an alternative perspective for hip fracture incidence calculation. This enables development of a generalizable method for probabilistic detection of starting points of hip fracture care episodes. Several risk factor and risk population extraction techniques required in register-based data analyses are also demonstrated. Finally, it is shown that empirical evidence suggests that hip fracture incidence is proportional to population level disability prevalence.

In conclusion, Finnish administrative data makes it possible to derive data for rather detailed population level risk factor stratification. Certain limitations of register-based data can be partly avoided by synthesizing data-sensitive methodological solutions during the analysis process.

Background

Societies are facing challenges as the public health burden increases in tandem with population aging. In order to fulfill the growing need for information for prevention and performance assessment purposes, local information systems are needed that would allow a continuous monitoring of the incidence and effectiveness of treatments for important health problems. However, additional data production requires funding and resources, and it would be beneficial if the required information could be produced using existing administrative data such as hospital discharge registers.

In fact, there is a great deal literature on doing descriptive epidemiology with administrative data [1,2]. The main problem with secondary administrative data is that the straightforward application of standard epidemiological practices may not be feasible, because the data collection cannot be tailored to meet the needs of the exact research problem, as with separate primary data collection. Therefore, it is obvious that the validity of secondary data depends on the research question in mind [3]. In other words, an extra interpretation and preprocessing phase that aims to find an adequate compromise between problem-driven and data-driven approaches becomes an important part of the research process [4]. Even though
this phase is commonly encountered while using secondary data in epidemiological studies, it is seldom reported explicitly. In fact, a more coherent report of the results can be achieved by stating that the required compromises were known a priori even if those choices were actually “findings” of the preprocessing phase. The drawback in underreporting is that the data utilization may seem to be easier than it really is. In the worst case, this may lead to uncrirical and mechanical repetition of “good examples” in circumstances where the assumptions of those examples are not valid.

This study examines the possibilities of routinely collected Finnish administrative data as a data source for an epidemiologic surveillance system. The basic idea is to demonstrate what kind of issues may require some “rethinking” during a concrete empirical research process that utilizes secondary data. Hip fracture incidence monitoring is used as an example application. First, an example is given of how conventional use of data turns out to produce improper estimates. Because data are not fully compatible with the traditional conceptualization, some alternative theoretical ideas originating from health services and aging research are used as additional links between the problem of hip fracture incidence monitoring and secondary data. After that several pragmatic issues confronted during the empirical research process on hip fracture incidence are examined from a methodological point of view. This study aims to offer practically useful methodological perspectives for the utilization of data by emphasizing the importance of continuous data-sensitive problem solving.

Analysis

Conventional approaches for determining the annual numbers of hip fractures by using hospital discharge data

Hip fractures represent a worldwide major public health burden whose impact is expanding as the population ages, with hip fracture incidence rates increasing exponentially with age [5,6]. For an individual, hip fracture is a serious and painful condition that requires and invariably results in acute hospital treatment. Hip fracture is also relatively easy to diagnose (compared with many other health problems), and practically all recorded hip fracture diagnoses (regardless whether principal or secondary) in hospital discharge abstracts reflect hip fracture treatment. Therefore, it can be expected that hospital discharge data are a good source for the identification of hip fracture patients.

Pitfalls in the use of hospital discharge data

There are also several known potential pitfalls in using hospital discharge data for calculating the numbers of injuries [7-10]. One general nuisance is that only hospitalized injuries are observed. In the case of hip fracture that is not a problem, because virtually all patients with hip fracture require hospital treatment, which results in a discharge record with a hip fracture diagnosis. Another potential pitfall is related to multiple hospitalizations of a single patient. In principle, record linkage allows detection of multiple hospitalizations of the same patient, and the real problem is to define which hospitalizations should be considered as readmissions. There are also other (more or less data-source specific) problems — such as the use of diagnosis codes in the selection of cases — which are typically special cases of common challenges for the use of secondary register data in (epidemiological) research, and require careful data preprocessing that incorporates tacit knowledge formalized in terms of appropriate data abstraction rules [2-4].

Hospital discharge data in Finland

In Finland, hospital discharge data are available in the Finnish Health Care Register which records data for all inpatient care discharges in institutions with 24-hour personnel and for outpatient surgical operations. Census data are also collected at the final day of each year in order to capture the ongoing care periods. The register is nationwide, which in international terms is exceptional in that it has such extensive legislative coverage of all public and private service providers. Each record in the register includes data on patient and provider ID-numbers, age, sex, area codes, and diagnosis and operation codes, as well as dates of admission, operation and discharge. Patient ID is a unique identification number given to all Finnish citizens and permanent residents. A system of unique identification numbers has been operating since 1968 and is used in all Finnish registers. Importantly, this system allows complete deterministic record linkage within and between Finnish registers [11]. Data quality is also shown to be good [12-18]. In the case of hip fracture, the completeness is very good and the accuracy of easily measurable variables is at least 95% [18].

Register based numbers of hip fractures in Finland

Finnish register data have been previously used as a data source for the calculation of the number of hip fractures [19,20]. For the purposes of this study, all discharge abstracts with a primary or secondary hip fracture diagnosis (10th revision of the International Classification of Diseases diagnoses S72.0, S72.1 and S72.2) from 1998–2002 were identified in the Finnish Health Care Register. The mean number of hospital discharge records (including census records) with hip fracture diagnosis was 14430 per year during 1998–2002 in Finland (Table 1).

However, the annual numbers of hip fracture discharge records do not tell the actual number of hip fracture events. A typical care chain for a hip fracture patient consists of acute hospital treatment on a surgical ward and
follow-up care on a general ward or in a specialist rehabilitation unit. In addition, a patient may be readmitted for the same fracture after initial discharge to home or to a long-term facility. In order to prevent multiple counting of cases, an individual-based record linkage is recommended for the detection of multiple hospitalizations for the same fracture [10]. This is not a problem with Finnish data, and the previous Finnish studies have identified all hospital discharge records with hip fracture diagnosis and then used calendar year boundaries to exclude multiple hospitalizations of each patient. The corresponding annual number of patients with hip fracture discharge was on average 7605 per year between 1998 and 2002 (Table 1).

Unfortunately, the use of calendar year boundaries makes such an exclusion approach artificial and has no epidemiological justification, because it is clear that calendar year boundaries result in the fact that the related hip fracture free clearance periods (time from the beginning of the year to a first fracture hospitalization of the year) vary per patient. This has two serious consequences. First, it is likely that many patients having their fracture during the final months of each year are erroneously counted as separate cases for two years, because it takes at least four months until the maximum restoration in terms of residential status (hospitalizations) is reached at the population level [21]. Second, one patient may have more than one hip fracture per year resulting in some undercounting. These drawbacks are an indication of need for a more appropriate estimate of the annual number of hip fracture events.

It is reasonable to hypothesize that virtually all recorded hospitalizations with a hip fracture diagnosis combined with hip fracture operation represent a new hip fracture, and that the date of an acute admission for surgery can be used as an estimator for the actual occurrence date of the hip fracture. The mean number of such operations was 6084 per year in Finland during 1998–2002 (Table 1). Unfortunately, this simple definition underestimates the true number of hip fractures. Even though a small number of (incorrectly recorded) re-operations may be erroneously included, the definition excludes patients treated conservatively (non-operatively) as well as patients who die before the operation. It is also known that the recording of hip fracture operations in The Finnish Health Care Register has not been totally complete [18,22]. This means that even some hip fracture related admissions without recorded operations should be considered indications of fresh hip fracture. The problem is to define which admissions reflect fresh hip fractures, and an easy solution is to use a constant hip fracture free clearance period to sort out new admissions from readmissions. If the new admission is defined as a hip fracture related record with no other hip fracture admissions for that individual in the previous two years, the mean number of hip fractures is 6787 per year (Table 1). As the use of long clearance periods may exclude some true subsequent fractures, the numbers of hip fractures were calculated also with a two-month hip fracture free clearance period. With this criterion, the mean number of hip fractures was 7258 per year (Table 1).

In principle, further diagnosis-based exclusion rules could be utilized in order to exclude certain nonstandard cases, such as pathological hip fractures, arthrosis related fractures, and cases with multiple fractures or orthopedic aftercare. However, variations in coding practices partially invalidate the use of such data abstraction rules, which assume recording of secondary diagnoses. In practice, even the clinical criterion of hip fracture may have some

Table 1: Number of hip fractures in Finland 1998–2002

<table>
<thead>
<tr>
<th>Year</th>
<th>Records with hip fracture diagnosis in the Finnish Health Care Register</th>
<th>The number of patients with hip fracture diagnosis</th>
<th>The number of patients with hip fracture discharge</th>
<th>Admissions with hip fracture diagnosis</th>
<th>Of which:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14089</td>
<td>7817</td>
<td>7575</td>
<td>13219</td>
<td>Of which:</td>
</tr>
<tr>
<td>1998</td>
<td>13818</td>
<td>7661</td>
<td>7465</td>
<td>12965</td>
<td>- recorded hip fracture operation</td>
</tr>
<tr>
<td></td>
<td>14192</td>
<td>7435</td>
<td>7425</td>
<td>13381</td>
<td>5934</td>
</tr>
<tr>
<td>2000</td>
<td>14978</td>
<td>7924</td>
<td>7706</td>
<td>13818</td>
<td>6742</td>
</tr>
<tr>
<td>2001</td>
<td>15071</td>
<td>8030</td>
<td>7854</td>
<td>13381</td>
<td>7246</td>
</tr>
<tr>
<td>2002</td>
<td>14430</td>
<td>7815</td>
<td>7605</td>
<td>14430</td>
<td>7136</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7411</td>
</tr>
</tbody>
</table>

Admissions with first record with hip fracture diagnosis in ten years

<table>
<thead>
<tr>
<th>Year</th>
<th>Of which:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>Of which:</td>
</tr>
<tr>
<td>1999</td>
<td>- persons aged 50 or more</td>
</tr>
<tr>
<td>2000</td>
<td>5990</td>
</tr>
<tr>
<td>2001</td>
<td>5853</td>
</tr>
<tr>
<td>2002</td>
<td>5932</td>
</tr>
<tr>
<td>Mean</td>
<td>6083</td>
</tr>
<tr>
<td>2003</td>
<td>6141</td>
</tr>
<tr>
<td>2004</td>
<td>6000</td>
</tr>
<tr>
<td>Mean</td>
<td>5551</td>
</tr>
<tr>
<td>2005</td>
<td>5413</td>
</tr>
<tr>
<td>2006</td>
<td>5543</td>
</tr>
<tr>
<td>2007</td>
<td>5644</td>
</tr>
<tr>
<td>2008</td>
<td>5667</td>
</tr>
<tr>
<td>Mean</td>
<td>5564</td>
</tr>
</tbody>
</table>
In any case, following the reasoning described above, it is likely that the mean number of hip fractures is somewhere between 6800–7200 per year in Finland between 1998 and 2002. The estimate for the number of hip fractures in 1998 is 6742–7246. Even the upper limit is significantly smaller than in the earlier Finnish study [20], where the annual number of hip fractures in 1998 was reported to be 7698. The reasons for the small difference in relation to the number of hip fracture discharges in 1998 (Table 1) are not known, but are most probably due to difficulties in reconstruction of the exact operationalization of some inclusion and exclusion criteria used in the earlier study. Also some overestimation is likely to occur because of the artificial definition of fresh hip fracture, as was described above. In conclusion, the earlier study reports a number that seems to be about 10% too large (corresponding to about 700 hip fractures).

Alternative theoretical approaches

As was demonstrated above, it is difficult to derive adequate estimates for the number of hip fractures even with the careful use of register data. The main problem is that the secondary data are not completely compatible with the theoretical objective of calculating the number of fresh hip fracture events. In other words, it would be beneficial if the theoretical objectives could be modified so that they are in agreement with the limits of the available register data. Fortunately, certain methodological ideas originating from health services and aging research are also useful in epidemiological applications.

Care episode approach

The definition for an incident case is a common problem in epidemiologic research. At the conceptual level this means that all events related to the same underlying disorder should be recognized. This is the fundamental idea in the care episode approach [23]. The care episode approach is widely applied in health economics and in health services research. The key point is that the care episode approach offers a sound methodological framework for dealing with multiple hospitalizations in terms of hospital discharge data, and it also provides a sound basis for the measurement of incidence, because an appropriate episode measure combining related hospitalizations is less subject to over- or undercounting of cases [24].

Unfortunately, a sound methodological framework does not automatically solve empirical difficulties [25]. One of the biggest practical problems with linked register data is to determine the starting point of the care episode. In principle, the starting point of a care episode for hip fracture should be rather easy to determine, because the actual event of fracture (or hospitalization following that event) is an obvious index point. However, in practice it is challenging to make a distinction between true subsequent hip fractures and hospitalizations due to ongoing treatment episodes or reoperations in terms of data. In standard practice, it is common to use the first health problem-related event available in the data as an index point. Another widely used approach is to use certain clearance periods to determine the appropriate index points. A third technique is to use external complementary data in determining (the number of) incident cases [26]. Other approaches seem to be rare [27].

Intuitively, it seems to be a good idea to determine an individual’s first hip fracture occurrence and consider it as a starting point for a (chronic) care episode. In terms of hospital discharge data this corresponds to the detection of an individual’s first hospitalization with recorded hip fracture diagnosis. The problem is that the hip fracture itself is not a chronic condition, but a remediable health deficiency. Therefore, an alternative theoretical interpretation for the first hip fracture is needed.

Gompertzian interpretation of aging-related hip fractures

One particularly interesting fact is that hip fracture incidence increases exponentially with age [5]. This kind of functional dependency (Gompertz law) has very strong interpretations from the point of view of population aging [28-30]. In this sense, perspectives from aging research may strengthen the traditional epidemiological interpretations (and vice versa) [31,32].

Biologically, aging can be seen as a complex process occurring stochastically in organs and tissues after reaching adulthood (and mainly after reproductive maturity), which results in irreversible damage accumulation and vulnerability to the failures in maintaining the integrity of tissues and organs [33,34]. A particularly fruitful approach is to consider aging-related events of interest as failures in components of a biological system [35]. An exponential increase of incidence with age means that the intensity of failures is constant (as the probability of failure is constant in time, the cumulative probability of failure is exponential). Therefore, it is not surprising that most aging-related (Gompertzian) conditions are closely connected to (cumulative) alterations in certain tissues or organs.

However, hip fracture is not just an aging-related failure in some single tissue or organ, but is typically the result of an accidental event (such as a fall). In other words, cumulative damage in some tissues (such as bone and muscle) increases the probability of (accidental or pathological) fracture, but it is also likely that some aging-related conditions increase the risk of accident. In this sense, hip fracture seems to capture effects of non-fatal failures in several
components of the human body. Among the Gomperzian conditions, only death seems to have a similar multidimensional interpretation as it captures fatal failure(s) in any vital components of human body. In other words, an interesting (methodological) analogy can be seen between death and hip fracture.

It is also obvious that serious but non-fatal failures in components of the human body result in some kind of disabilities. Such disabilities — typically measured in terms of reduced functional capacity and coping with physical activities of daily living — are important risk factors for hip fracture [36,37]. In other words, it is reasonable to hypothesize that the prevalence of such risk factors in the underlying population is related to the incidence of (aging dependent) hip fractures. Following this interpretation, it becomes clear that the occurrence of first aging-dependent hip fracture gives an approximate upper limit for the time of development of critical hip fracture risk factors, because the risk factors must have exceeded the critical level before the event of (low-energetic aging dependent) hip fracture. From this it follows that the first aging-related hip fracture can be seen as an indication of a chronic (disability) condition, and any subsequent hip fractures represent just a nuisance for related population level interpretations. In conclusion, there are theoretical justifications for the determination of the first aging-related hip fracture occurrence, which are in concordance with the restrictions of available hospital discharge data. This interpretation is also of particular importance in a theoretical sense, because it suggests that a commonly available time series of hip fracture incidence may also reflect the more general disability trends of the population, which is a testable hypothesis.

Calculating the number of hip fractures

So far, the methodological problem of the definition of the hip fracture episode has been reduced to a detection of first aging-related hip fractures in the population (starting points of the first hip fracture care episodes). This is not necessarily a straightforward task in practice, because the available data allow only limited backward follow-up time and it remains unknown whether the first hip fracture found in the data really is the first hip fracture of the person. Probabilistic methods can be used to correct the number of observed persons to the number of persons with the first appearance of a chronic condition [27].

However, if there is a need to identify the persons with a first hip fracture in terms of care episode (instead of just calculating the number of persons with a first hip fracture), it is inconvenient that the actual backward follow-up times (definitions of first hip fractures) vary between persons. In addition, the subsequent hip fracture of a person is biasing interpretations only as far as the risk of subsequent hip fracture is significantly higher than the risk of first aging-related hip fracture. In other words, there may be a cut-off point after which the probability of having a new hip fracture is reduced to the level it would be even without preceding hip fracture, i.e. preceding hip fracture is an “uninformative” predictor of the new hip fracture. By examining the observed and expected probabilities of preceding hip fractures it becomes possible to give justification for a suitable clearance period to be used. Importantly, this probabilistic argument for the definition of the starting points of the care episodes also generalizes easily to other disorders.

**Probabilistic determination of the starting point of care episode**

In order to determine the cut-off point for a clearance period, the first admissions (index points) with the (principal or secondary) diagnosis of hip fracture in 1998–2002 were identified for each patient. To ensure sufficient (backward) follow-up time for each patient, all available discharge records of the hip fracture population from 1987–2002 were obtained from the registers using the unique personal identification numbers as linkage keys. For each index point, backward time to the previous hip fracture admission of the same patient was calculated. Time was measured in months. There were data from the years 1987–2002, so the minimum (backward) follow-up time was 11 years.

The first task was to calculate the expected probabilities of earlier hip fractures in the hip fracture population of interest. The problem here is that reasonable estimates for the incidence of first hip fractures are needed for the calculation. In this study, it was hypothesized that clearance periods between one and ten years may result in a reasonable definition of first hip fracture, and the mean age- and sex-group specific incidences between 1998 and 2002 were calculated by using one year and ten year clearance periods. In order to calculate the actual expected probabilities, the logarithms of the mean hip fracture incidence rates for different age-groups were used in estimating the exponential trend in age-group specific rates by using (log-)linear regression analysis for both sexes separately [38]. This functional relationship allows the interpolation of an incidence rate for persons who are over 40 years old [29]. The individuals with an observed hip fracture were then followed (backwards) in time, with age correspondingly corrected at each time point for all persons in the risk population. With a known age and sex distribution of the risk population, it was possible to predict the expected probability of hip fracture by using the estimated log-linear relationship between age and hip fracture incidence. For ages below 40 years, a constant incidence rate was used in the prediction. These probabilities were then summed across the risk population resulting in an expected number of hip fractures that was finally divided by the size of the risk population.
population, giving an expected (conditional) probability of hip fracture (for each time point). The expected probabilities are shown in Figure 1 using dotted lines. As can be seen, the difference between expected probabilities based on one- and ten-year clearance periods is quite small.

In order to calculate the observed probabilities, a hazard function giving the (conditional) probabilities of having a preceding hip fracture admission as a function of backward time (i.e. time was measured from index admission to a preceding hip fracture admission) in months was estimated nonparametrically using a product-limit estimator, and smoothed using a polynomial moving average. As the estimated hazard function in Figure 1 shows there was an increased and non-stabilized (non-constant) risk for an admission involving hip fracture occurring about 120 months after the previous admission related to hip fracture and the observed risk remains higher than expected in the ten-year period. If a more formal way is needed to determine where the lines meet, it is straightforward to calculate confidence intervals for the hazard function. In conclusion, it seems that a clearance period between seven and ten years is needed until the risk is reduced to the same level as that without a preceding hip fracture. In this study, a conservative ten-year clearance period was selected. The ten-year criterion has also been used in other studies [39].

Figure 1
Probability of having a preceding hip fracture as a function of backward time in months from the first fracture in 1998–2002. The smaller picture is a tenfold magnification of the final months. Dotted curves represent the expected probabilities of having a hip fracture (upper curve is calculated with a one-year clearance period and the lower curve with a ten-year clearance period, see text for more information).

Estimating the number of new hip fractures with limited data

It is unfortunate that no shorter than a ten-year criterion seems to be suitable for hip fractures, because in most countries there may not be the required data available for such a long backward follow-up period. However, if an estimate for the overall number of first-ever hip fractures is enough, and there is no need to identify which patients actually had their first hip fracture, data with limited backward follow-up can also be used.

The numbers of patients without a preceding hip fracture as a function of backward time in the logarithmic scale for 1998 are drawn in Figure 2. As can be seen, there is almost a linear relationship between logarithmic time and the number of patients without a preceding hip fracture. In other words, even data with quite a short backward follow-up time allow the estimation of such a linear trend resulting in a reasonable prediction for the number of hip fracture patients at the ten year cut-off point (or at any other desirable cut-off point) for a clearance period. For example, if only one year data for backward follow-up are available, the idea is to calculate time from index admission to previous fracture or to the beginning of (backward) follow-up. Then for each day (preceding the index admission) the number of patients with a longer time distance for the previous fracture is calculated. Finally, a logarithmic transformation is applied to the day variable, and a linear regression model (where the number of fractures is a response and constant and logarithmic day are explanatory variables) is estimated. Predictions from this linear model can be easily extrapolated to any cut-off point of interest.

Risk factor extraction

Using Finnish register data it is feasible to identify hip fracture patients who have their first aging-related hip fracture in the sense described above. This makes it possible to determine the status of certain hip fracture risk factors available in the data — such as age, sex, institutionalization, urbanity, season and year [40-45] — in relation to these patients. Methodologically, three different risk factor extraction techniques can be separated: internal, external, and empirical.

Internal extraction

Internal extraction corresponds to the use of data abstraction rules within the register data. For example, using the Finnish data it is not enough to determine the index hip fracture admission, but the actual day of hip fracture must be inferred using appropriate algorithms. It is obvious that the care episode related to hip fracture starts from the first contact with the health care system after the actual event of fracture. However, the index admission could be for a long stay in residential care, while the diagnosis referred to an accident that happened near the discharge.
day, or the index admission could be for a surgical period, while the actual accident had happened before the admission. After the detection of initial surgical admission, more abstraction rules can be developed. For example, the most accurate diagnoses can be extracted from the data corresponding to the surgical treatments following the fractures, since those are the first ones based on x-ray and physical examinations. For the purposes of this study, a person was classified as an institutionalized long-term care patient if he or she was admitted to the surgical ward from some institution providing inpatient care, and if he or she also had a recorded administrative decision for long-term care or had received inpatient care for more than six months during the year preceding the fracture.

**Empirical extraction**
In empirical extraction, preliminary analyses of the available data are used to justify the definitions of variables and data abstraction rules required in the internal extraction. For example, it is not obvious what kind of definition for seasons should be used. This problem can be solved when using the Finnish register data, which allow accurate calculation of daily numbers of hip fractures. After smoothing out the random variation in absolute numbers by a moving-average technique, the mean of daily numbers of the new hip fractures was 15.2 for persons aged 50 or older (Figure 3). There was small but clear seasonal variation so that 53.5 per cent of fractures occurred during the winter/spring season (from November to April) compared to 46.5% during the summer/fall season (from May to October). Data also revealed that there had been some very “hazardous” days during the winter season, but seasonal variation was almost completely due to non-institutionalized persons (Figure 3). Using these results (of preliminary analyses) in the definition of seasons is an example of empirical extraction. In fact, the method developed above for the definition of first aging-related hip fracture is another example of empirical extraction.

**Risk population data**
For incidence calculations data on risk population are also needed. Typically, the official population figures are available in administrative databases with stratification according to age, sex and area of living. In this study, population figures (taken on the last day of the years 1997–2002) in 5-year groups were obtained for each municipality (local administrative unit in Finland) from the Social and Health Service Statistical Database (SOTKA). Municipality works as an aggregate unit in external extraction, and also allows easy determination of population figures for any combination of these basic units.

**External extraction**
In external extraction, variables that link individuals to aggregate levels are first internally extracted and then external data that describe aggregate units are linked to each individual. For example, for the purposes of this study, the area code (municipality) at the index admission was used to classify each patient as rural or urban (including semi-urban areas) using the official grouping defined by Statistics Finland (rural municipalities are those municipalities in which less than 60 per cent of the population lives in urban settlements). In general, any other patient characteristic attributable to area-specific phenomena could also be used here instead of urbanity.
Risk population for internally extracted risk factors

It is more difficult to determine risk population for internally extracted risk factors. In this study, long-term institutional care was one risk factor of interest. Fortunately, data from the Finnish Health Care Register can be used to calculate the total numbers of clients in long-term institutional care on the last day of the years 1997–2002, since the register includes all inpatient hospital and nursing home care in Finland. This individual level data can be easily aggregated to appropriate groups (such as stratification by age, sex and municipality), and further subdivision of population figures according to long-term institutional care is possible. However, during the research process it turned out that the calculation technique used in official statistics concerning institutional care was inappropriate for the purposes of epidemiological studies. Therefore, the significantly downwards biased numbers were corrected using a technique reported elsewhere [46].

This procedure finally resulted in simultaneously observed risk populations on the final day of the years 1997–2002 with stratification by sex, age, urbanity, and institutionalization. The derivation of such exceptionally detailed nationwide population figures was possible because of common aggregate units (municipalities) in the databases and registers.

Trends for risk populations

Observed risk populations on the final day of the years of interest are not good approximations for the size of the risk population (or the follow-up time for risk population). A common method is to interpolate the mean population by using observed census data on two consecutive years (approximating also the follow-up time of the risk population, if measured in person years). This method can be generalized by modeling the trends extractable from observed risk populations. In this study, regression models with the constant, year and square(year) as regressors were fitted for each group with a stratification by sex, age, institutionalization, and urbanity (with age groups 50 to 64, 65 to 74, 75 to 84 and 85+). These models were then used in the approximation of the required risk populations varying by year and — importantly — by (empirically extracted) seasons, too. Finally, there were the observed numbers of hip fracture events and the follow-up times (in person years) for 320 groups.

Limitations of risk population data

It is well known that an accurate number of the population at risk is an essential requirement for incidence calculations [47]. In this study it was possible to derive population-based denominators for each group of interest. These denominators are desirable in terms of accuracy, but not perfect, because an assumption of a stable population (no short-term fluctuations in migration or mortality) is needed for interpolating census day population figures to appropriate mean follow-up times.

In addition, as a ten-year clearance period was used in this study, a person is not at risk of a new hip fracture for the 10 years following a hip fracture and should be — in principle — left out of the risk population. However, this correction was not done in this study, because adequate (group-specific) hip fracture prevalence data were not available. The bias resulting from keeping the prevalent pool in the risk population was considered to be insignificant, as the number of hip fractures is very small in relation to the population in younger age-groups and mortality following the hip fracture is very high for the older age-groups. In fact, relative bias [1-(uncorrected incidence/corrected incidence)] of mean incidence seemed to be less than 2%. Moreover, even if some group-specific bias exists, the direction is towards conservative estimates (underestimation rather than overestimation of incidences).

In conclusion, there is still room for improvement in deriving risk population data. However, the problems are insignificant in terms of the “iceberg phenomenon” involved in the denominator problem [47], and there is no reason to believe that the data used in this study would not reflect the true epidemiology of hip fracture.

Hip fracture incidence

After data preparation, it is simple to examine the univariate effects of population-level risk factors using standard methodology. It is quite straightforward to extend such analyses for other stratifications of interest, such as area specific estimates (possibly requiring empirical-Bayes estimation) revealing regional variation in hip fracture incidence. Results of such basic analyses are reported elsewhere [48]. Only exemplary results on the simultaneous effects of risk factors with an interpretation offering a new epidemiological perspective for hip fracture are given here.

In standard practice, only a few risk factors are typically considered simultaneously in population level incidence studies. Risk factors of interest are included into the model, and estimates are interpreted as independent effects (effects adjusted for other risk factors in the model). The problem is that there may be complicated interactions between the risk factors, which distort straightforward interpretations.

In this study, the Poisson-regression model was used in the simultaneous examination of several risk factors. Other analyses have indicated that the (long-lasting) increase in age-adjusted incidence had recently stabilized
in Finland, with the age-adjusted incidence almost constant between 1999 and 2002 [48]. As the interest in this study was not in secular trends of incidence but in the effects of risk factors, the final Poisson regression model was estimated using the combined data from 1999 to 2002. All main effects (sex, age, institutionalization, urbanity, and season) and statistically significant interactions up to the third degree were included in the model. The goodness of fit was acceptable in terms of Pearson chi-square statistic. Because there were a lot of significant interactions between risk factors, it was rather difficult to report all relevant estimates in easily interpretable form. Therefore, the group-specific rates adjusted for all other factors were calculated by anti-logging the corresponding linear predictors (without the offset) based on the estimated model. These adjusted rates represent the systematic effects extracted from the data by the model. Overlapping confidence intervals of two groups indicate that there is no statistically significant difference between rates in these groups.

Alternative epidemiological perspective for hip fracture incidence

The results in Table 2 show that there is a higher hip fracture incidence among older persons and women than among younger persons and men managing at home. However, there is no significant effect of sex on incidence among institutionalized persons. The incidence is higher during the winter time for all persons with a functional status potentially good enough to allow for walking outdoors (non-institutionalized under age 85). Urbanization is not associated with significantly higher hip fracture incidence in Finland. As it is known that institutionalization is a sign of reduced coping with daily activities [49] and that women tend to have more functional limitations and physical disability than men as age increases [50], the results are in concordance with the hypothesis suggesting that hip fracture incidence is proportional to the prevalence of disability related hip fracture risk factors in the underlying population.

In order to test this hypothesis, a correlation was calculated between the nationwide data on the prevalence of outdoor walking ability — being a strong disability-related risk factor for hip fracture [51] — of persons aged 65 to 84 and the hip fracture incidences for the corresponding age and sex groups for each year between 1998–2002. Data on walking ability status for the population was extracted from consecutive nationwide surveys on health behavior among older people conducted by the National Public Health Institute [52]. Linear regression analysis showed that the prevalence of individuals unable to walk outdoors explained 97.5% of the variation in the hip fracture incidence. In conclusion, the hypothesized association between hip fracture incidence and prevalence of (disability-related) hip fracture risk factors was not falsified, and deserves further examination.

Conclusion

This study evaluated the feasibility of producing information for the purposes of monitoring hip fracture incidence based on Finnish administrative register data. This type of information production is not a simple task, but requires creative use of rather complicated methodology. It was shown that straightforward use of register data results in biased estimates of the numbers of hip fractures, and that even the sophisticated use of data is dependent on the more or less ambiguous definitions of actual hip fractures.

In this study, the definition of hip fracture was linked to population aging. This reasoning resulted in the suggestion that hip fracture-related disability has certain interesting methodological similarities to the event of death. Moreover, an example from the results empirically supported this interpretation by demonstrating that hip fracture incidence is associated with a certain type of aging-related disability. Actually it seems that hip fracture incidence is directly proportional to the prevalence of population level disability. In this sense, hip fracture incidence trends also reveal the more general disability trends of the population.

Methodologically, while using register-based data, the determination of first aging-related hip fractures is easier than the determination of all hip fractures. Moreover, the detection of the first aging-related hip fracture event closely resembles the determination of the starting point of the hip fracture care episode. A method was developed based on an observed (and expected) hazard function, that allows probabilistic justification for the starting point of the care episode. This method is applicable in the general care episode approach for a wide variety of health problems regardless of the intention to use care episodes in incidence calculations.

Furthermore, three different techniques for risk factor (and corresponding risk population) extraction were demonstrated. Finnish administrative data makes it possible to derive data for a rather detailed population level risk factor stratification, but such a feature has certain more or less data-specific limitations which can be partly avoided by developing methodological solutions to the encountered problems. One area requiring further work that is beyond the scope of this study is the more careful estimation of hip fracture prevalence in Finland.

Finally, this study tries to demonstrate that the traditional methodological paradigm with an assumption of theory-driven data collection and fixed methods has limited suitability for data analyses that utilize secondary data. Two
main ideas should be noted. First is the empirical justification of theory-driven operationalized definitions, and second is the theoretical model that is compatible with theory-driven as well as data-driven approaches. For example, in this study register data gave only uncertain estimates for the number of all hip fractures, so the definition and theory underlying the hip fractures had to be revised. After theory revision, empirical justification for the extraction of risk factors was utilized, which finally gave results that supported the revised theory. In fact, it may be difficult to make a clear distinction between definitions, preliminary analyses justifying these definitions, and actual results of the study. In conclusion, it seems that more alternative perspectives are needed in the analyses of the more detailed (secondary) the available data is, the more alternative perspectives are needed in the analyses of such data and in the reporting of the results of the analyses.

Competing interests
The author(s) declare that they have no competing interests.

Authors’ contributions
RS carried out the whole research process.

Acknowledgements
This work was financially supported by the research grant from the Academy of Finland. The funding agreement ensured the author’s independence in designing the study, interpreting the data, writing, and publishing the report.

References


IV
**Kuinka laskea ikääntyneiden pitkääikaisasiakkaiden määrä rekisteritietojen perusteella?**


**REIJO SUND, SARI KAUPPINEN**

**JOHDANTO**


Virallisii tilastoihin (ks. mm. Ikääntyneiden sosiaali- ja terveyspalvelut 2002) pitkääikaisasiakkaat lasketaan vain kyseessä olevan tilastovuoden asiakaslaskeutuimen avulla. Tällöin pitkääikaisasiakkaiksi määritellään ne henkilöt, joilla on asiakaslaskeutuimen perusteella voimassa oleva pitkään ja joka asiakaslaskeutuimen mukaan ovat siihen mennessä ollut aikaisemmin asiakaslaskeutuimen kyseessä olevassa asiakaslaskeutuimen yli 90 päivää.

Kun selvitetään riskiväestötä epidemiologista tutkimusta varten, on tärkeää saada tavoitteet kyseessä olevan väestöryhmän mahdollisimman kattavasti. Näin ollen saattaa olla tarpeen käyttää hyväksi myös myöhempien asiakaslaskeutujen ja päätyneiden hoitojaksojen tiedot.

Tämän tutkimuksen tarkoituksena on kuvata ikääntyneiden pitkääikaisasiakkaiden määrä, ikä- ja sukupuoliryhmittäin vuosille 1997–2002 sekä selvitää, miten arvioitaan yhden vuoden asiakaslaskeutumisesta laskettu pitkääikaisasiakkaiden lukumäärä eroaa siitä, joka saadaan, kun otetaan huomioon myös myöhempien asiakaslaskeutujen ja päätyneiden hoitojaksojen tiedot.

**AINEISTO JA MENETELMÄ**

Tutkimusta varten sosiaali- ja terveydenhuollon hoitoilmoitusrekisteristä poimittiin kaikki yli 50-vuotiaiden asiakkaiden pitkääikaisishoidon krite-


Aineisto jaettiin lisäksi sukupuolen, iän ja laitostyyppiin mukaan ryhmiin. Sukupuoliryhmä indeksikoittaa iillä (i = 1, 2; 1 = mies ja 2 = nainen), ikäryhmä iillä (j = 1, 2, 3, 4; 1 = 50–64, 2 = 65–74, 3 = 75–84- ja 4 = yli 85-vuotiaat) ja laitostyyppiä indeksillä k (k = 1, 2; 1 = terveyskeskus tai sairaala ja 2 = vanhainkoti tai vastaava). Näillä merkinnöillä vuoden x asiakaslaskentatietojen perustuvaa lukumäärä ryhmässä ik:k j on O_{ik} (x) ja koko aineistoa hyödyntäen saatavasta lukumäärä E_{ik} (x). Edelleen p_{ik} (x) = O_{ik} (x) / E_{ik} (x) on laskentatietojen osuus koko aineiston perusteella saadusta lukumäärästä ryhmässä ik:nuonnax.

Tavoitteena oli muodostaa malli, jota käytäten olisi mahdollista arvioidaa ryhmien ik:iä välissä eroja osuuudessa p ja toisaalta ennustaa E_{ik} = O_{ik} / p_{ik} halutulle vuodelle x, kun O_{ik} tunnetaan. Olettamalla, että p_{ik} = (1 / E_{ik})Poission(\lambda_{ik},E_{ik}) - jossa \lambda_{ik} on teoreettinen osuus ryhmässä ik:j ja \lambda_{ik}E_{ik} on O_{ik}:n odotusarvo, päädyttiin yleistettyyn log-lineaariselle Poission-regressionmalliin. Koska selektiivänä olivat osuudet, jotka mallissa painottaa jotakin havaintoja p_{ik} lukumäärällä E_{ik}. Ryhmättyymuuttujien lisäksi mallissa käytettiin selittävänä muuttujana myös vuotta. Mallia estimoitaisiin käytettiin vuosien 1997–2001 havaintoja ja pyrittiin ennustamaan vuoden 2002 osuudet p_{ik} (2002) kaikille ryhmille ik:k.

**Tulokset**


Taulukossa 3 on raportoitu aineistoa tarkoituksenmukaisimpana vuonna mallin tulokset. Osoitettua on siis selitetty ikäryhmää, laitostyyppiä, sukupuolta sekä ikä- ja laitostyypistyttävä välieffektiä vuosavaiheen muuttujien avulla. Mallin yhteensopivuus aineiston kanssa mittattuna \chi^2-
estillä on erinomainen (p = 0,947), eikä minkään interaktioteoriain lisäarviointia enää parantanut malla merkkittävästi. Perusryhmänä on käytetty vanhainkodissa tai vastaavassa osuus yli 84-vuotiaita naisia ja vuosimuuttujan perusuovudeksi on valittu vuosi 2000. Ikäryhmän, laitostyyppiin ja sukupuolen päävaikutukset ovat tilastollisesti merkittyinä, vaikka ne kontrolloiditaan toistensa ja ryhmävastaisen vuosimuuttujan suhteen. Kattavuus siis parannee selvästi iän myötä (p < 0,0001). Osoitut ovat isomia sairaaloiden ja terveyskeskusten vuodeosastoilla kuin vanhainkodissa tai vastaavissa (p < 0,0001) ja kattavuus on parempi naisten osalta (p = 0,02). Lineaarinen vuosieffekti on kasvava vanhainkotien tai vastaavien osalta kaikissa ikäryhmissä. Sairaaloiden ja terveyskeskusten vuodeosastoilla on laskeva trendi vuorimassa ikäryhmässä, mutta muissa ikäryhmissä tilastollisesti merkittävää lineaarista trendiä ei ole havaittavissa.
## Taulukko 1.

<table>
<thead>
<tr>
<th>50–64 Miehet</th>
<th>Vanhainkoti tai vastaava</th>
<th>Molemmat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terveyskeskus tai sairaala</td>
<td>Laskenta</td>
<td>Osuus</td>
</tr>
<tr>
<td>Laskenta</td>
<td>Osuus</td>
<td>Kaikki</td>
</tr>
<tr>
<td>1997</td>
<td>626</td>
<td>80,4 %</td>
</tr>
<tr>
<td>1998</td>
<td>577</td>
<td>79,8 %</td>
</tr>
<tr>
<td>1999</td>
<td>610</td>
<td>81,2 %</td>
</tr>
<tr>
<td>2000</td>
<td>608</td>
<td>74,7 %</td>
</tr>
<tr>
<td>2001</td>
<td>595</td>
<td>74,8 %</td>
</tr>
<tr>
<td>2002</td>
<td>572</td>
<td>73,4 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>65–74 Miehet</th>
<th>Vanhainkoti tai vastaava</th>
<th>Molemmat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terveyskeskus tai sairaala</td>
<td>Laskenta</td>
<td>Osuus</td>
</tr>
<tr>
<td>Laskenta</td>
<td>Osuus</td>
<td>Kaikki</td>
</tr>
<tr>
<td>1997</td>
<td>972</td>
<td>80,5 %</td>
</tr>
<tr>
<td>1998</td>
<td>965</td>
<td>83,1 %</td>
</tr>
<tr>
<td>1999</td>
<td>947</td>
<td>83,3 %</td>
</tr>
<tr>
<td>2000</td>
<td>894</td>
<td>84,3 %</td>
</tr>
<tr>
<td>2001</td>
<td>929</td>
<td>82,7 %</td>
</tr>
<tr>
<td>2002</td>
<td>844</td>
<td>83,7 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>75–84 Miehet</th>
<th>Vanhainkoti tai vastaava</th>
<th>Molemmat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terveyskeskus tai sairaala</td>
<td>Laskenta</td>
<td>Osuus</td>
</tr>
<tr>
<td>Laskenta</td>
<td>Osuus</td>
<td>Kaikki</td>
</tr>
<tr>
<td>1997</td>
<td>1398</td>
<td>82,4 %</td>
</tr>
<tr>
<td>1998</td>
<td>1374</td>
<td>83,0 %</td>
</tr>
<tr>
<td>1999</td>
<td>1421</td>
<td>84,6 %</td>
</tr>
<tr>
<td>2000</td>
<td>1442</td>
<td>83,2 %</td>
</tr>
<tr>
<td>2001</td>
<td>1437</td>
<td>84,4 %</td>
</tr>
<tr>
<td>2002</td>
<td>1481</td>
<td>85,3 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>85+ Miehet</th>
<th>Vanhainkoti tai vastaava</th>
<th>Molemmat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terveyskeskus tai sairaala</td>
<td>Laskenta</td>
<td>Osuus</td>
</tr>
<tr>
<td>Laskenta</td>
<td>Osuus</td>
<td>Kaikki</td>
</tr>
<tr>
<td>1997</td>
<td>944</td>
<td>84,7 %</td>
</tr>
<tr>
<td>1998</td>
<td>941</td>
<td>88,3 %</td>
</tr>
<tr>
<td>1999</td>
<td>907</td>
<td>84,6 %</td>
</tr>
<tr>
<td>2000</td>
<td>899</td>
<td>84,5 %</td>
</tr>
<tr>
<td>2001</td>
<td>898</td>
<td>87,0 %</td>
</tr>
<tr>
<td>2002</td>
<td>874</td>
<td>88,3 %</td>
</tr>
</tbody>
</table>

1) Kyseessä olevan vuoden asiakaslaskentatietojen mukainen pitkääikaisaikakkaiden määrä  
2) Asiakaslaskentatietojen mukainen määrä suhteutettuna muilla pitkääikaisaikkkien kriteerit täyttävillä hoitoimoi-
tuksilla täydennetyn aineiston mukaisen määrään  
3) Kaikkien pitkääikaisaikoiden kriteerit täyttävien hoitoimoiituksien perusteella laskettu pitkääikaisaikakkaiden
määrä  
4) Pitkääikaisaikakkaiden määrä yhteensä (palvelutyypin mukaan erottelematta)  
5) Niiden asiakkaiden osuus, jotka aineiston mukaan olivat samaan aikaan sekä terveydenhuollon vuodeosastolla
että vanhainkodissa tai vastaavassa  
6) Lihavoitetut luvut ovat vuoden 2002 ennustettuja arvoja (katso selitys tekstistä)

**POHDINTA**
Tässä tutkimuksessa havaittiin, että pelkästään yhden vuoden asiakaslaskentatiedoista laskemall
ka pitkääikaisaikakkaiden lukumäärä on selvästi pienempi kuin silloin, kun otetaan huomioon
myös myöhäisten asiakaslaskentojen ja päättä
neiden hoitojaksojen tiedot. Tähän on monta
mahdollista sytä.

Research Report 174  
STAKES 2008  
Methodological Perspectives for Register-Based Health System Performance Assessment
Taulukko 2.  

<table>
<thead>
<tr>
<th>50–64 Naiset</th>
<th>65–74 Naiset</th>
<th>75–84 Naiset</th>
<th>85– Naiset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laskenta</td>
<td>Osuus</td>
<td>Molemmat</td>
<td>Vanhanlaki tai vastavaa</td>
</tr>
<tr>
<td>1997</td>
<td>455</td>
<td>78,7 %</td>
<td>578</td>
</tr>
<tr>
<td>1998</td>
<td>434</td>
<td>73,9 %</td>
<td>587</td>
</tr>
<tr>
<td>1999</td>
<td>427</td>
<td>78,8 %</td>
<td>542</td>
</tr>
<tr>
<td>2000</td>
<td>406</td>
<td>73,7 %</td>
<td>551</td>
</tr>
<tr>
<td>2001</td>
<td>394</td>
<td>71,4 %</td>
<td>552</td>
</tr>
<tr>
<td>2002</td>
<td>453</td>
<td>74,6 %</td>
<td>607</td>
</tr>
</tbody>
</table>

1) Kyseessä olevan vuoden asiakaslaskentatietojen mukaisen pitkäaikaisiasiakkaiden määrä
2) Asiakaslaskentatietojen mukaan määrä suhteutettuna muilla pitkäaikaisihoitotennyt ja hoitoimoi-
tuksilla täydennettynä asiakaisen mukaan määrän
3) Kaikkien pitkäaikaisihoitojen mukaan pitkäaikaisihoitotennynä ja määrän
4) Pitkäaikaisiasiakkaiden määrä yhteensä (pelvettyyniin mukaan erottelu)
5) Niiden asiakaisihoitojen osuus jokaisen mukaan olivat samaan aikaan sekä terveydenhuollon vuodeosastolla
6) Lihaoidut luvut ovat vuoden 2002 ennestettuja arvoja (katso selitys tekstistä)

PITKÄAIKAISIAISIAKKUUDEN KRITEERIT EIVÄT VÄLTÄMÄTTÄ OLE
VOIMASSA VIELÄ LASKENTAPÄIVÄNÄ

Pitkäaikaisiakkauksu ei välttämättä ole ollut voimassa asiakaslaskentapäivänä, vaikka myöhempi
iliomietusten mukaan hoitojakso osoittautui
yli 90 vuorokautta kestäväksi. Kun aineistosta
poimitaan pitkäaikaisihoitojen tunnusmerkit täy-
tävät hoitojakso, pitkäaikaisiakkauksen kohdis-
tu näissä hoitojaksoissa takautovin kiikoko
hoitojakson.

Jos varsinaisen pääosuus pitkäaikaisihoitosta on tehtävä vasta laskentapäivän jälkeen, asiakkaan
 ei kuulukaan olla mukaan laskentapäivän
pitkä-
aikaisiakkaiden lukumäärässä. Jos ellekin
asiakasryhmille pitkäaikaisihoitosta päätöstä ei edes
tehdä, jolloin asiakas on pitkäaikainen vasta,
Taulukko 3.
Pitkäaikaisiasiakkaiden asiakaslaskennoissa raportoitavaan osuuteen vaikuttavia tekijöitä kuvaavaan Poisson-regressiomallin tulokset 1.

<table>
<thead>
<tr>
<th>Perusosuus</th>
<th>95-% luottamusväli</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0,8620</td>
</tr>
</tbody>
</table>

Osuus referenssiryhmän osuudesta

<table>
<thead>
<tr>
<th>Ikäryhmä</th>
<th>95-% luottamusväli</th>
</tr>
</thead>
<tbody>
<tr>
<td>50–64</td>
<td>0,8755</td>
</tr>
<tr>
<td>65–74</td>
<td>0,9458</td>
</tr>
<tr>
<td>75–84</td>
<td>0,9655</td>
</tr>
<tr>
<td>85+</td>
<td>1,0000 (referenssiryhmä)</td>
</tr>
</tbody>
</table>

Palvelutyypit

<table>
<thead>
<tr>
<th>Vuodeosasto</th>
<th>95-% luottamusväli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vanhainkoti</td>
<td>1,0288</td>
</tr>
</tbody>
</table>

Sukupuoli

| Mies | 9,9834 | 0,973 | 0,994 |
| Nainen | 1,000 (referenssiryhmä) |

Osuus edellisen vuoden osuudesta

<table>
<thead>
<tr>
<th>Vuosiefekti (perusvuosi 2000)</th>
<th>95-% luottamusväli</th>
</tr>
</thead>
<tbody>
<tr>
<td>50–64 vuodeosasto</td>
<td>0,9803</td>
</tr>
<tr>
<td>50–64 vanhainkoti</td>
<td>1,0244</td>
</tr>
<tr>
<td>65–74 vuodeosasto</td>
<td>1,0073</td>
</tr>
<tr>
<td>65–74 vanhainkoti</td>
<td>1,0229</td>
</tr>
<tr>
<td>75–84 vuodeosasto</td>
<td>1,0066</td>
</tr>
<tr>
<td>75–84 vanhainkoti</td>
<td>1,0208</td>
</tr>
<tr>
<td>85+ vuodeosasto</td>
<td>1,0062</td>
</tr>
<tr>
<td>85+ vanhainkoti</td>
<td>1,0099</td>
</tr>
</tbody>
</table>

1) Taulukossa raportoidulla mallilla voidaan laskea taulukoiden 1 ja 2 ennustetut osuudet. Esimerkiksi 75–84-vuotiaat vuodeosastolla olevat miehet: 0,8620 (perusosuus) × 0,9655 (ikäryhmä 75–84) × 1,0288 (vuodeosasto) × 0,9834 (mies) × 1,0066 (vuosiefekti 2000 → 2001) × 1,0066 (vuosiefekti 2001 → 2002) = 0,853

kun hoitojakso on kestänyt yli 90 vuorokautta. Esimerkiksi tehostettu palveluasuminen ei ole laitoshoitoa, joten sen asiakkaille ei tehdä pitkäaikaisen hoitoon perustuvaa päätöstä. Myöskään sairaalat eivät pääasiasa ole tarkoittettu pitkäaikaisen hoitokseen voimakasta, joten niissä olisi perustauduttava harvoin pitkäaikaisen hoitoon ja syy pitkään hoitojakson on varsinaisen sairaanhoitoon pitkittyminen esimerkiksi kompleksia ja vakavaa.

Edellä mainittujen seikkojen vaikutusten arviointi rakentaa aineistosta laskettuun korkeintaan 90 vuorokautta laskentatapavänä hoidossa olleiden asiakkaiden määrä on myös rähmäs. Vaikka näin saadut luvut sisältävät myös en alle 90 vuorokautta hoidossa olleet, joille ei onko sen jälkeen pitkäaikaistettu hoidossa, se näyttää sitä, että pitkäaikaisista asiakkaiden "takauksesta" kirjautuminen myös hoitojakson pitkäaikaishoidossa kriteerin täyttymistä edeltäneelle ajalle selittää valtaosan terveyskeskusten ja sairaaloiden asiakaslas-

centojen pitkäaikaisiasiakkaiden puuttumisesta

laskentatiedoissa. Sen jälkeen vanhalämpöissä ja

vastaavissa on asiakkaita, joiden puuttuminen

laskentatietojen pitkäaikaisassiakkaita ei selity

takauksesta kirjautumisessa. Näiden "selittämät-

äidin" osuus on huomattava ikäryhmä (50–64-v.)
lukuun ottamatta suurten prosenttien luokkaa.

HOITOILMOITSAINEISTON KATTAUVUUSONGELMAT

Vaikka määritelty suurin osa laskentatiedostoa

puuttuusta pitkäaikaisiasiakkaita näyttääkin selit

lytymään siitä, että ne eivät todellisuudessa vielä

ole olleetkaan pitkäaikaisiaasiakkaita, laskentatapavänä, osa havaittuista eroista voi myös puutteista hoitoilmoitusaineistojen kattavuudessa. Tämä voisi selittää eroja terveydenhuollon hoitoilmoitustensa on, sillä kokemusperäisesti arvioidaan, että terveydenhuollon

asiakaslaskentamoitukset saatuutteet jättää use

ammin ilmoitattavaan kuin päätyneet hoitojak-

Sosiaalihuollon haitoiloimotuksaineistoa on kuitenkin kertaamme vasta vajaan kymmenen vuotta, joten varsinkin tarkastelujakson alkupuolella kattavuus on ollut nykyistä heikompa. Sitä tarkastelujakson alkuvuosina näkyy suuremmi ero laskentatietojen ja kaikkien (myös myöhempien) ilmoitusten perusteella pitkäaikaisiaakkaiden määrissä selitty nimenomaan kattavuuden vuosittaisella parannemisella.

**Virheet yksittäisissä hoitoloimotuksissa**

Myös virheet yksittäisissä hoitoloimotuksissa voivat olla syynä osaan luvuissa näkyvistä eroista. Joskus pitkäaikaisihoitot ongelmia on ollut tehty, mutta sitä ei huomata kirjata hoitoloimotustietoihin. Erityisesti sosiaalihuollon, kun tiedot kerta- tään pääasiassa käsityksellä, tällaisista tiedoista on välttämät, lähdetä tarkastamaan, vaan kohta jätetään tyhjäksi.


**Päälkekkäiset hoitojaksot**


**Rajoittuminen pitkäaikaispäättökellusiin**

Hoitoilmoitustiedoihin Vaikka edellä on pohdittu eräitä hoitoloimotusrekisteriin liittyviä potentiaalisia ongelmia, on tärkeää pitää mielellä, että tässä tutkimuksessa käytetty aineisto ei varsinaisesti todista mitään hoitoloimotusrekisterin kattavuudesta, sillä pelkkien pitkäaikaisuuden kriteerit täyttävien hoitokapkojen käyttö aiheuttaa jo sinänsä ongelmia. Jos pitkäaikaisiakkaan hoitojaksot esimerkiksi sairaalakäynnin takia keskeytetvät alue 90 vuorokaudessa, asiakas ilman merkintää pitkäaikaisihoitoinen päätöksestä eivät tule lainkaan mukaan pitkäaikaisiakkaiden poimintaan. Tämä ongelma on kuitenkin rikkinäinen, mittaetaan pitkäaikaisiakkan tulevat kriteerit täyttävien hoitokapkojen tavoin. Tämä ongelma on olemassa rikkinäisesti ja pitkäaikaisiakan poimintaan tulevat yhden vuoden asiakaslaskentatietoja tai myös myöhempiä hoitoloimotuksia. Pitkäaikaisuuden kriteereillä rajatulla aineistolla ei pystytä myöskään kuvaamaan kattavasti asiakkaan kokoa hoitoketjua, sillä asiakkaan hoitojaksot saattaa todellisuudessa olla rekisterisissä, mutta ei tule mukaan aineistoon, jos se ei täytä pitkäaikaisuuuden kriteerejä.

Käytettäessä sekä poistoilmoitus että laskentatietoa olisi asiakkaan lukumäärät periaatteessa mahdollista päätellä jokaiselle päivälle erikseen. Tämä kuitenkin johtaa hankaluuksiin, sillä rajattessa asiakkaat rekisteristä pitkäaikaispäätöksen perusteella, päivitytävien palauttioiden kuvaajasta tulee ”sahamainen” siten, että kunakin vuoden viimeisen päivän ja seuraavan vuoden ensimmäisen päivän populaatioi eroavat toisistaan jopa kymmenen prosenttia. Näin suuri päivämuutos ei ole todellisuudessa realistinen. Ilmiö johtuu siitä, että kaikille asiakkaille ei
löyty veden asiakasläskentäteidosta. Tällaisen aineiston avulla saattaisi lisätietoa myös siitä, ilmoitetaanko lähtö- ja tulopäät ovat oikeansuuntaisesti. Edelleen tällaiseen kokonaisaineiston avulla olisi mahdollista laskea asiakkaan "koko laitoshistorian kesto, joka on hoitoilmoituseen rikottu" (=tiedot ovat jaksokohdaisia, joten esimerkiksi sairaalakäynti katkaisee vanhinkotihoidon) takia muuten ongelmallisista.
This study reports the age and sex specific counts of older people in long-term care in Finland from 1997–2002, based on two different calculation techniques. It turns out that when using only the single year census data, the number of long-term clients appear to be fewer than when using additional information from the discharge abstracts. Neither of these techniques is better than the other, and the intended use of data determines which is more suitable.

Reijo Sund
VTM, tilastotieteilijä
Vaikutusvuuden ja oikeudenmukaisuuden tutkimusryhmä
Stakes

Sari Kauppinen
FM, erikoissuunnittelija
Tilastot-ryhmä
Stakes

Kirjallisuus


Lonkkamurtumien ilmaantuvuus Suomessa 1998–2002


Aineisto ja menetelmät

Ensimmäisten lonkkamurtumien määrä oli vuosien 1998–2002 aikana keskimäärin noin 6 000 vuodessa (taulukko 1). Näistä runsaat 5 500 tapahtui vuodessa 50 vuotta täyttäneidenpotilaista. Kyseisistä lonkkamurtumapotilaista valtaosassa on runsaat 6 000 tapahtui vuodessa (taulukko 1). Näistä runsaat 5 500 tapahtui vuodessa 50 vuotta täyttäneidenpotilaista. Kyseisistä lonkkamurtumapotilaista valtaosassa on runsaat...
osa (72 %) oli naisia. Noin 29 % murtumista tapahtui laitoshoidossa ja 19 % pitkäaikaisessa laitoshoidossa. Liukuvalla keskiarvolla tasoitetut lonkkamurtumien päivittäiset määrät on esitetty kuvassa 1. Lonkkamurtumia tapahtui keskimäärin 15,2 päivässä, ja havaittavissa oli pienehköä mutta selkeää kausivaihtelua: noin 54 % murtumista tapahtui marras–huhtikuussa ja noin 46 % touko–lokakuussa. Aineiston mukaan talvikauden aikana murtumia tapahtui muutamina päivänä poikkeuksellisen paljon. On myös huomion arvoista, että pitkäaikaispotilaiden murtumissa ei ollut havaittavissa kausivaihtelua (kuva 1).

Ikä- ja sukupuolivakiituja ilmaantuvuus oli vuonna 1998 suurempi kuin muina vuosina, joiden aikana ilmaantuvuus pysyi varsin vakaa (taulukko 2). Tämä on merkki lonkkamurtumien ilmaantuvuuden tasautumisesta myös Suomessa. Vakioitu ilmaantuvuus oli naisilla suurempi kuin miehillä ja erityisen suuri pitkäaikaisessa laitoshoidossa olleilla (taulukko 2). Pitkäaikaisessa laitoshoidossa olleiden miesten ja naisten ilmaantuvuusluvuissa ei kuitenkaan ollut merkitsevää eroa.


**Pohdinta**

Suomessa käytännössä kaikki lonkkamurtumat hoidetaan sairaalassa, joten on odotettavaa, että lonkkamurtumataan liittyvät hoitojakson koot ovat valtakunnalliseen hoitoilmoitusrekisteriin. Lonkkamurtumapotilaiden kattava identifiointi rekisteristä on siis periaatteessa mahdollista sillä oletuksella, että tiedot on kirjattu kattavasti ja luotettavasti rekisteriin. Tuoreen validointitutkimuksen mukaan kattavuus on lonkkamurtuman osalta erittäin hyvä ja vain noin 2 %:iin kirurgian erikoisalan hoitojaksoista on liitetty virheellisesti muu kuin lonkkamurtumadiagnosti
Suurempia hankaluuksia kuin tavanomaiset kattavuus- ja luotettavuusongelmat aiheuttaa rekisteritietoihin perustuvissa ilmaantuvuuslaskeluissa. Suurin osa kaikkien lonkkamurtumien järjestelyistä on saa vastuukehityksen tukikohdasta. Myös pitkään tapahtuneita tapauksia voidaan selvittää aikaisemmin katsottavissa rekisterissä poistamatta niiden merkitystä. Keskitymällä vain kaikkien lonkkamurtumien järjestelyihin voidaan selvittää, mitä tällaisissa tapauksissa on mahdollista virheellisesti tulkita.

Vertailuun helpottaa, että lähes kaikki muutkimot, jotka ovat ilmainneet lonkkamurtumiaan, olivat saman toimijan liitteenä. Tämä vastaa suurempaa kohdeaineistoa, joka edelleen on mahdollista selventää virheellisistä tulkinnista.


(Sund ym., julkaisematon käsikirjoitus).

Suurempia hankaluksia kuin tavanomaiset kattavuus- ja luotettavuusongelmat aiheuttaa rekisteritietoihin perustuvissa ilmaantuvuuslaskelmissa lonkkamurtumien järjestelyissä. Myös pitkään tapahtuneita tapauksia voidaan selvittää aikaisemmin katsottavissa rekisterissä poistamatta niiden merkitystä. Keskitymällä vain kaikkien lonkkamurtumien järjestelyihin voidaan selvittää, mitä tällaisissa tapauksissa on mahdollista virheellisesti tulkita.

Vertailuun helpottaa, että lähes kaikki muutkimot, jotka ovat ilmainneet lonkkamurtumiaan, olivat saman toimijan liitteenä. Tämä vastaa suurempaa kohdeaineistoa, joka edelleen on mahdollista selventää virheellisistä tulkinnista.
Lonkkamurtumien ilmaantuvuus Suomessa 1998–2002

<table>
<thead>
<tr>
<th>Naiset</th>
<th>Ikäväkioitut ilmaantuvuus</th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miehet</td>
<td>Ikkäväkioitut ilmaantuvuus</td>
<td>57</td>
<td>59</td>
<td>55</td>
<td>52</td>
<td>54</td>
</tr>
<tr>
<td>Ikkäväkioitut ilmaantuvuus</td>
<td>216</td>
<td>193</td>
<td>193</td>
<td>194</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>Ikkäväkioitut ilmaantuvuus</td>
<td>208</td>
<td>167</td>
<td>187</td>
<td>192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ikkäväkioitut ilmaantuvuus</td>
<td>228</td>
<td>200</td>
<td>178</td>
<td>187</td>
<td>192</td>
<td></td>
</tr>
<tr>
<td>Ikkäväkioitut ilmaantuvuus</td>
<td>367</td>
<td>385</td>
<td>363</td>
<td>362</td>
<td>368</td>
<td></td>
</tr>
<tr>
<td>Ikkäväkioitut ilmaantuvuus</td>
<td>311</td>
<td>315</td>
<td>316</td>
<td>315</td>
<td>311</td>
<td></td>
</tr>
<tr>
<td>Ikkäväkioitut ilmaantuvuus</td>
<td>2114</td>
<td>1796</td>
<td>1957</td>
<td>1910</td>
<td>1707</td>
<td></td>
</tr>
<tr>
<td>Pitkäaikaisessa laitoshoidossa olleet</td>
<td>1650</td>
<td>1582</td>
<td>1485</td>
<td>1371</td>
<td>1747</td>
<td></td>
</tr>
<tr>
<td>Muut kuin pitkäaikaisessa laitoshoidossa olleet</td>
<td>307</td>
<td>294</td>
<td>293</td>
<td>293</td>
<td>286</td>
<td></td>
</tr>
</tbody>
</table>

Näennäisesti samanlaisista määritelmistä huolimatta tämän tutkimuksen arviot lonkkamurtumien lukumääristä ovat selvästi pienempiä kuin aiemmin tutkimuksissa raportoidut (Kannus et al. 1999, Lüthje et al. 2001). Esimerkiksi vuonna 1998 lonkkamurtumien todellen määrä oli välillä 6,800–7,200, eli aiemmat arviot ovat noin 6–14 % liian suuria. Aiemmin tutkimuksissa jokaisen henkilön ensimmäinen rekisteristä löytyvää lonkkamurtumadiagnosti-

Tässä tutkimuksessa todettiin, että naisilla oli suurempi lonkkamurtuman riski kuin miehillä paitsi pitkäaikaisessa laitoshoidossa, joka oli sukupuolten eri maailmat ja henkilöittäin yhteydessä muuttunut lonkkamurattuman riskiin. Pitkäaikaisessa laitoshoidossa olevien osalta eillä olleet havaittavissa vuoden aiheuttaneet liukaisut ja heikentyminen liittyvät kauasvaihtelua. Vaikka muiden osalta talvikausi oli yhteydessä lisääntynneen suuruuden murtumavuonna, vaikka kyse olisikin kiihtäätämaista samasta hoitoepisodista.

Lopuksi


Toimintakyky heikentee iän myötä ja naisilla nopeammin kuin miehillä (Murtagh ja Hubert 2004). Laitoshoitoon joutuneen heijastelee kuin kykyta kauasvaihtelua. Vaikka muiden osalta talvikausi oli yhteydessä lisääntynneen suuruuden murtumavuonna, vaikka kyse olisikin kiihtäätämästä samasta hoitoepisodista.

Methodological Perspectives for Register-Based Health System Performance Assessment
Research Report 174
STAKES 2008

R. Sund

Kirjallisuutta


REUO SUND, VTM, tilastotieteilijä reijo.sund@stakes.fi
Stakes, vaikuttavuuden ja oikeudenmukaisuuden tutkimusryhmä PL 220, 00531 Helsinki

Lonkkamurtumien ilmaantuvuus Suomessa 1998–2002

Lonkkamurtumien ilmaantuvuus Suomessa 1998–2002
Quality effects of operative delay on mortality in hip fracture treatment

R Sund, A Liski

Background: Most hip fracture patients undergo surgery, but there is conflicting evidence on the relation between the timing of surgery and the outcome of treatment. There is considerable variation in the length of surgical delays between hospitals, possibly reflecting the quality of care.

Methods: The effects of operative delay on mortality were estimated using various statistical methods applied to observational data from 16,881 first-time hip fracture patients aged 65 or older from 47 hospitals (providers) in Finland in 1998–2001.

Results: A prolonged in-hospital operative delay was associated with a higher mortality of hip fracture patients in individual level analyses, but the instrumental variable approach indicated that the individual level effect was not caused by the operative delay but by inappropriate methodological assumptions. There was extensive variation between providers in the proportion of late surgery patients. Provider level analyses showed that the effects of the provider of operative delay on mortality are quite small, but there is a clear association between the proportion of late surgery patients and non-optimal treatment.

Conclusions: If provider level heterogeneity is not explicitly taken into account, studies of the effects of surgical delay on outcomes are prone to serious bias. The proportion of patients with prolonged waiting time for surgery at the provider level seems to work as an effective evidence-based quality indicator. Providers should reduce unnecessary delays to surgery and identify more carefully patients not suitable for early surgery.
operation, and discharge (or death). The completeness and accuracy of the registers is known to be good.6–11

Data preprocessing
Data were preprocessed so that information concerning patients aged 65 or older with a first hip fracture could be accurately identified.12 For the purposes of this study, the data were restricted to patients operated on using an internal fixation, a prosthesis, or a total hip replacement. Patients with incomplete data were excluded from the analyses. A few providers (hospitals) had not reported operation dates for the years 1998–1999, but otherwise no systematic bias was found in the sensitivity analyses. The final study population consisted of 16 881 patients, which represents 83.3% of all first time hip fracture patients aged 65 or older. Forty six providers had at least 50 cases, while the remaining 154 cases from other providers were combined into a single figure. The existence of possible co-morbidities was extracted for each patient from his or her medical history using the diagnosis codes recorded in the data.13–15 The extraction method was adapted and updated from the original Charlson co-morbidity categories16 and applied to the current data set.

Statistical analyses
Cumulative probabilities of unadjusted mortality were calculated using the product limit estimators. Excess risk of death was defined as difference in cumulative probabilities of mortality between two comparable groups. Hazard ratios were obtained using the proportional hazards model allowing the adjustment of observed confounders. Since only a limited number of covariates were observed, the biasing influence of unobserved heterogeneity was statistically controlled by incorporating the gamma frailty term into the model.20

The significance of the differences between the unadjusted patient characteristics was determined using the z test (binary variables) or $\chi^2$ test (categorical variables). The logistic regression model was used to calculate odds ratios and significances of differences between adjusted patient characteristics. The same logistic model was also used for predicting the late surgery probability for each patient. These probabilities were then aggregated to the provider level, resulting in the expected number of late surgery patients for each provider.21 The provider specific ratios between observed and expected numbers of late surgery patients were further modelled using the hierarchical gamma-Poisson model to obtain the comparable risk adjusted proportions of late surgery patients for providers.22 The weekday adjusted proportion was calculated using the predictive margin technique.23 In the instrumental variables analysis the 1 year adjusted mortality was calculated for each admission day of the week and the null hypothesis of equal mortalities was tested using the $\chi^2$ goodness of fit test.24

RESULTS
Effect of length of operative delay on mortality
The cumulative probabilities of unadjusted mortality rates attributable to different lengths of operative delay are shown in fig 1. Waiting times of 0–2 nights had the same effect on mortality, but there was a significant increase in mortality with waiting times of 3–4 nights (p<0.001), and waiting times of 5 or more nights resulted in an even higher mortality rate. Early surgery was defined as a waiting time of 0–2 nights (n = 14 426, 85.5%) and late surgery as a waiting time of at least 3 nights (n = 2455, 14.5%). The unadjusted hazard ratio between late and early surgery was 1.24 (95% confidence interval (CI) 1.13 to 1.34, p<0.0001).

Excess risk of death
The unadjusted excess risk of death for late surgery compared with early surgery quickly rose to 3%, then slowly increased during follow up to about 5% at 1 year (fig 2). Complementary analyses revealed that the excess risk was quite stable and near to its maximum at 1–3 years and then began to decrease slowly. As the lower confidence limits in fig 2 indicate, the unadjusted excess risk of death became statistically significant 2 weeks after the operation.

Characteristics of early and late surgery groups
Characteristics of the patients with hip fracture classified into early and late surgery groups are shown in table 1. Together with the corresponding odds ratios for late surgery adjusted for the other characteristics in the table. Early surgery was associated with being 85 years or older, of female sex, having a pertrochanteric fracture, and receiving long term inpatient care during the year preceding the admission. Admission on a Wednesday and a medical history of dementia or cataract were also associated with early surgery. The risks for late surgery were likely to be increased by admission from a nursing home, health centre or readmission shortly after

![Figure 1](https://example.com/1.png) Figure 1. Cumulative probabilities of unadjusted mortality following hip fracture for operative delays of various lengths.

![Figure 2](https://example.com/2.png) Figure 2 Unadjusted excess risk of death (with 95% confidence interval) for late hip fracture surgery.
discharge, admission between Thursday and Saturday, and co-morbidities including cardiovascular disease and peripheral vascular disease (and, prior to adjustment, diabetes, cerebrovascular disease, chronic pulmonary disease, hypertension, and anaemia). The proportion of patients having late surgery substantially decreased during the period of the study.

**Adjusted effect of operative delay on mortality**

The hazard ratio for the late surgery reduced to 1.18 (95% CI 1.09 to 1.28, p<0.0001) after adjusting for the provider and the characteristics in table 1. Even after controlling for the unobserved covariates, the hazard ratio remained significant, indicating an increased mortality for late surgery. However, after the pseudo randomised assignment of patients into early and late surgery groups using admission day of the week as an instrumental variable, the difference in the adjusted 1 year mortality rate was found not to be attributable to operative delay (p = 0.069).

**Proportions of late surgery patients**

There was extensive variation in the proportions of late surgery patients between providers (fig 3). The conservative 95% confidence interval crosses the potentially achievable
Trend analysis at provider level

The simultaneous examination of the provider level proportions of late surgery patients and 1 year mortality rates showed that there was a statistically significant positive trend for an association between a larger share of late surgery patients and 1 year overall mortality (p<0.05), and that a smaller proportion of late surgery patients was (non-linearly) associated with a higher mortality rate for these patients (fig 4).

Trend analyses for severity groups

The associations were further analysed according to severity (measured as predicted 1 year mortality based on the same observed patient characteristics as in table 1) in five severity groups of patients (table 2). The proportion of patients undergoing late surgery was only slightly higher in the more severe patient groups but the 1 year mortality rate increased significantly with severity, as expected. The slope of the trend of overall mortality increased with the severity but was statistically significant only for the most severe patients. The positive slope of early surgery mortality was statistically significant in severity groups 3 and 4, but for the most severe patients this association disappeared. A smaller proportion of late surgery patients was associated with a higher late surgery mortality rate in all groups except the most severe patients.

DISCUSSION

In this study the definition of late surgery turned out to be 3 or more nights in hospital after admission, since shorter waiting times were not associated with higher unadjusted mortality. A similar “long” delay is commonly used as the definition of late surgery in other studies. The unadjusted excess mortality for late surgery patients increased during the year following the operation. It is interesting that the short term mortality differed only slightly between early and late surgery groups, but the difference became much clearer after the perioperative period. This may be partly due to the fact that almost half the perioperative deaths are unavoidable in an unselected population. If patients are going to die shortly after the fracture—regardless of the quality of the hip fracture treatment provided—it is obvious that the treatment effect turns out to be small. The treatment effect becomes observable for long term mortality but is potentially biased because of heterogeneity between the early and late surgery groups in patient characteristics affecting the outcome.

Several patient characteristics differed between the early and late surgery groups, but the prolonged operative delay seemed to increase the (long term) mortality significantly even after adjustment for these observed patient characteristics. There were no detailed clinical data available which obviously makes the adjustments only partial. The bias attributable to the unobserved covariates was controlled by allowing extra variation in the model, but the treatment effect remained significant. However, this kind of model is still prone to bias if the relationships between the risk factors and mortality are not correctly specified in the model. The pseudo randomizing instrumental variable approach also require strong and questionable assumptions, the sizeable difference in the estimates of the effect of operative delay on mortality between methods means that there is a need for hypotheses to explain such differences.

For instance, one could try to separate the acceptable delays from the unacceptable ones or one could record the reason for the delay in the data. In the strictest sense, this kind of approach needs the assumption that the acceptable and unacceptable delays can be defined and measured uniquely. In practice, an acceptable delay corresponds to a clinical decision to postpone the operation and the assumption is approximately fulfilled if the clinical practice remains constant. Because the data used in this study did not include the reason for the delay, the hypothesis of constant clinical practice was examined indirectly using provider level analyses.
The provider specific proportion of late surgery patients, adjusted for observed patient characteristics, showed that there was extensive variation between providers, which is a common finding. A simple performance assessment interpretation is that the percentage of late surgery patients can be reduced to a potentially achievable level that can also be interpreted as the upper limit for the proportion of acceptable delayed patients. Consequently, the expected proportion of unacceptable delayed patients is the proportion of late surgery patients exceeding this upper limit. This leads to the hypothesis that the overall mortality of hip fracture patients should increase with an increasing proportion of late surgery patients given that the longer operative delay would have an adverse effect on mortality. In fact, a statistically significant trend between the provider specific proportions of late surgery patients and overall 1 year mortality gives empirical evidence for the hypothesis, but the actual volume of the effect was small. Another provider level hypothesis is that the long term mortality of late surgery patients is higher if only the patients unfit for surgery are delayed, since the unfit condition for surgery is also a risk factor for 1 year mortality. This hypothesis was also supported by empirical evidence which revealed a non-linear association between the provider specific proportion of late surgery patients and the 1 year mortality rate for these patients.

A more careful examination of the groups of patients with different predicted mortality (severity) gave even more insight into the provider level association between the proportion of late surgery patients and mortality. Intuitively speaking, the least severe patients are young, come from home, and are without severe medical conditions. Correspondingly, the most severe patients are older with much co-morbidity and come from residential care. For patients in the least severe group, mortality was higher in the late surgery patients for all providers. This probably indicates that, in this group, hip fracture related mortality is caused by the medical problems which require late surgery and the prolonged surgical delay does not itself increase the mortality. For patients in the most severe group mortality was also higher for the late surgery patients for all providers. However, in this group the mortality rate in the late surgery patients was lower in the providers with a small proportion of late surgery patients and there were no differences in mortality rates among early surgery patients between providers. This means that, in this group, it is essential to perform early surgery in all patients who can withstand it because the significantly prolonged surgical delay makes the patient’s condition worse and increases mortality. For groups 2–4 the interpretation is more difficult than for patients in the extreme groups. Since the mortality rate in early surgery patients increases significantly and is higher even than the mortality rate of late surgery patients for providers with a large proportion of late surgery patients, it seems that in these groups an operation performed too early may cause more harm than the prolonged waiting time.

Other differences between providers in the proportion of patients with late surgery result mainly from lack of resources such as temporary unavailability of the operating rooms or surgeons and problems with obtaining medical clearance from other specialties in off hours. The importance of these factors can be illustrated using the effect of the admission day on the in-hospital operation delay; the mean proportion of patients undergoing late surgery would be reduced from 14.5% to 9.0% (p<0.0001) if all patients had been admitted on “the best day”.

In spite of the large nationwide database and the advanced methods, the observational study design and administrative data require a special orientation to the analyses and caution should be exercised when interpreting the results. Data were available for 1998–2001 and the practices may have disproportionately changed for the different providers during
that time. The results must therefore be interpreted as the mean situation between the years 1998–2001 for each provider. Non-operated patients were excluded from the analyses since the operative delay was not defined uniquely for these patients. Inclusion of these patients would certainly increase the mortality rate since the typical non-operated patient is in a worse condition and dies before any operation can even take place. The trend analyses are intended to summarise the systematic parts of the variation between the providers. However, there seems to be some unexplained variation, especially in the proportions of late surgery patients, because of the small number of patients for some providers. In the absence of the trend analyses need to be confirmed and validated in further studies.

Study designs which explicitly take into account the heterogeneity at the provider level are a prerequisite for unbiased results since the selection bias may vary between providers. One could also include the provider level characteristics in the analyses, but easily observable characteristics such as hospital type or number of surgeons typically allow only indirect interpretations. A more direct approach would be to observe the performance of providers continuously and to ask the providers themselves to clarify for their performance.

In Finland the mean proportion of late surgery patients has decreased during the years of the study, but extensive variation between providers exists. In addition to possible system problems, there also seems to be variation in the clinical practices for judging which patients are suitable for early surgery. Identification of good practices and the release of a national clinical guideline for hip fracture treatment will probably improve the situation. In this sense, the proportion of patients with a prolonged waiting time for hip fracture surgery operation seems to be an effective evidence based quality indicator, since it clearly indicates an area for improvement. The effects of operative delay on mortality at the patient level are quite small, but at the provider level the association between the proportion of late surgery patients and non-optimal treatment is clear.

Authors’ affiliations
R Sund, A Liski, National Research and Development Centre for Welfare and Health (STAKES), Outcomes and Equity Research, Helsinki, Finland
This research was supported by a research grant from the Academy of Finland.

REFERENCES


