UvA-DARE (Digital Academic Repository)

Registers in cardiovascular epidemiology

Reitsma, J.B.

Citation for published version (APA):
Chapter 1

Registers in epidemiological research

The number of registers in health care is growing, and they are increasingly used in public health and clinical studies. Chapter 1 presents examples of studies solely based on data from numbers of cardiovascular hospital admissions and deaths derived from national registers are analysed by age, sex and calendar time.

New developments have contributed to the increased number and increased use of registers. The first development is the progress of information technology. Faster computers, the proliferation of computer networks and the availability of software for analysing large data sets are important factors. "Health care systems as well as the demand for accountability in health care has been increased. By the evidence-based medicine movement and the growing collaboration. Accountability requires data, and from a traditional perspective, this would lead to primary data collection. Using existing data from registers is an appealing alternative from a practical point of view, as the numbers in time and money can be huge compared to primary data collection.

The number and size of health care registers have grown considerably as a result of these two developments. Some of the more recent registers have an administrative background, others have been developed by a more clinical interest. Health care registers have proven to be valuable tools in the surveillance of infectious diseases and the monitoring of some non-infectious diseases, like cancer. The more recent use of administrative registers in clinical research, however, has received mixed appraisal in the medical literature. To quote Laine in an editorial in a volume of the Annals of Internal Medicine, that was completely devoted to this topic: Clinical learning from small sets of meticulously collected primary research and clinical data, not from secondary data collected for billing and other administrative functions. This, however, like much else in health care these days, is changing, like it or not..."
References


1.1 INTRODUCTION

The number of registers in health care is growing, and they are increasingly used in public health and clinical studies. Chapters two through five in this thesis present examples of studies solely based on data from registers. The numbers of cardiovascular hospital admissions and deaths derived from national registers are analysed by age, sex and calendar time.

Two developments have contributed to the increased number and intensified use of registers. The first development is the progress in information technology. Faster computers, the proliferation of computer networks and the reduced costs of computers and storage media facilitate the routine collection of data. A second trend is the requirement that decisions and actions should be founded on factual evidence. Both are universal developments, including the health care system as well. The demand for accountability in health care has been responded by the evidence based medicine movement and the Cochrane collaboration. Accountability requires data, and from a conventional point of view, this would lead to primary data collection. Using existing data from registers is an appealing alternative from a practical point of view, as the savings in time and money can be huge compared to primary data collection.

The number and size of health care registers have grown considerably as a result of these two developments. Some of the more recent registers have an administrative background, others have been developed in a more clinical context. Health care registers have proven to be valuable tools in the surveillance of infectious diseases and the monitoring of some non-infectious diseases, like cancer. The more recent use of administrative registers in clinical research, however, has received mixed appraisal in the medical literature. To quote Laine in an editorial in a volume of the Annals of Internal Medicine that was completely devoted to this topic:

Clinicians learn from small sets of meticulously collected primary research and clinical data, not from secondary data collected for billing and other administrative functions. This, however, like much else in health care these days, is changing, like it or not.

---

1 We distinguish between 'register' and 'registry' conform the English literature, in which a register is the actual listing of records and a registry the whole organisation responsible for one or more registers.
This chapter explores whether this complaint is justified. We examine the potential problems associated with the use of registers in answering various epidemiological research questions, including studies with an etiological, clinical and public health perspective (the latter two encompassing health services research). Our perspective is a general one, but all our examples relate to cardiovascular disease.

In this introductory chapter, we start with some general remarks on the history and objectives of registers in health care (section 1.2) and give a formal characterisation of the registration process (section 1.3). We provide a functional description of epidemiological research in section 1.4, focusing on the timing and presence of information on outcome, determinant and study population, components of the so-called occurrence relation. By examining the interaction between the data requirements of different types of research and the characteristics of data from registers (section 1.5), we identify potential sources of problems. Based on this knowledge, we locate various types of epidemiological research along a line of increasing data complexity. Researchers will face more difficulties if registers are used in research requiring more complex data (section 1.6). We hope that this chapter will contribute to a better understanding of the opportunities and limitations of registers in epidemiological research and illustrate several ways of improving on the use of registers.

1.2 REGISTERS: HISTORY AND OBJECTIVES

Vital records are probably the oldest examples of the use of register data in health care. Foremost they had, and still have, a legal function of providing certification of birth and death. Death records have been used to monitor the health of populations since the early 1700s. Today, they are still used to measure the impact of diseases and injuries, both within and across countries. The first infectious disease registers were established around 1800, as physicians were increasingly required to report patients with specified communicable diseases (concept of 'notifiable' diseases). During the next century the so-called epidemiological transition happened, leading to a shift in the pattern of morbidity and mortality from infectious to non-communicable chronic diseases. This shift prompted the development of monitoring systems for these chronic diseases. The first cancer registers were established around 1950 and many other disease- and treatment-specific registers ensued. Up to then, the main purpose of registers had been surveillance. The primary objective in surveillance is to monitor the incidence or prevalence of specific
health problems in the general population and to characterise those at greatest risk with respect to socio-demographic factors.\textsuperscript{4,25}

During the past two decades, two new trends have become visible. First, a new type of register emerged, the so-called administrative database.\textsuperscript{1,6,7,19} Such a register primarily collects information about certain actions within the health care system for general administrative purposes. Hospital statistics data and claims based registers are typical examples of administrative registers. They usually contain information from a large range of conditions. This is in contrast with disease-specific registers, which focus on a single disease (intervention) or on a related group of diseases (interventions) and have a more clinical background.

Second, the interest of measuring the nation's health beyond the output of mortality statistics and infectious disease registers is growing.\textsuperscript{22} Monitoring the population’s health by means of comprehensive summary measures of health (combining mortality and morbidity information) has been proposed, posing new challenges to nation-wide data collection among others through registers, health interviews and health examinations.\textsuperscript{26}

This chapter discusses the role of registers in general, despite the large variation in original purpose, content and size of registers (table 1). The only registers we exclude are local registers (single physician or single institution) or registers with a short-term perspective, because of their different logistics. The justification for reviewing registers at large is the fact that all registers have some important points in common (table 1):\textsuperscript{9,11,27,28}

1. records are created in response to a registration event
2. the aim is to register consecutive registration events
3. for each event a comprehensive and similar set of data is recorded (fixed format)

The first characteristic distinguishes registers from populations that arise from specific research activities, such as health examinations and interviews. If functional, we will distinguish between administrative and disease-specific registers. The way in which registries collect and record information is the subject of our attention in the next section.
Table 1. Some examples of administrative and disease-specific registers.

<table>
<thead>
<tr>
<th>Type of registers</th>
<th>Registration event and other features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital statistics data</td>
<td>A single hospitalisation; registers are either based on a stratified sample of hospitals or have complete coverage</td>
</tr>
<tr>
<td>National register of</td>
<td></td>
</tr>
<tr>
<td>hospital admissions</td>
<td></td>
</tr>
<tr>
<td>Insurance and billing</td>
<td>Each billing transaction or prescription order</td>
</tr>
<tr>
<td>data</td>
<td></td>
</tr>
<tr>
<td>Health maintenance</td>
<td></td>
</tr>
<tr>
<td>organisations</td>
<td></td>
</tr>
<tr>
<td>Mortality data</td>
<td>The death of a person; in most countries the register has complete national coverage and a long history; legal responsibility for notification</td>
</tr>
<tr>
<td>National registers of</td>
<td></td>
</tr>
<tr>
<td>causes of death</td>
<td></td>
</tr>
<tr>
<td>Perinatal registers</td>
<td>The birth of one or more babies; records often contain information about both mother and offspring</td>
</tr>
<tr>
<td>Obstetric and neonatal</td>
<td></td>
</tr>
<tr>
<td>databases</td>
<td></td>
</tr>
<tr>
<td>Disease registers</td>
<td>Being diagnosed with a particular disease; for some infectious diseases a legal responsibility for notification</td>
</tr>
<tr>
<td>Infectious disease</td>
<td></td>
</tr>
<tr>
<td>Congenital malformations</td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td></td>
</tr>
<tr>
<td>Trauma registers</td>
<td>An accident severe enough to warrant medical attention or hospitalisation</td>
</tr>
<tr>
<td>Treatment registers</td>
<td>Undergoing a specific treatment or operation</td>
</tr>
<tr>
<td>Coronary artery bypass</td>
<td></td>
</tr>
<tr>
<td>surgery</td>
<td></td>
</tr>
<tr>
<td>PTCA</td>
<td></td>
</tr>
<tr>
<td>Heart (renal) transplant</td>
<td></td>
</tr>
<tr>
<td>Renal replacement therapy</td>
<td></td>
</tr>
</tbody>
</table>
1.3 REGISTRATION PROCESS: CASE CAPTURING AND DATA RECORDING

Figure 1 provides a formal representation of the registration process. A distinction is made between the capturing process and the recording process. The capturing process consists of three mechanisms that determine whether a record will be created or not in response to a certain health event. These mechanisms are depicted below the horizontal line that represents calendar time. The recording process determines what information will be recorded and how. The recording process is presented above the time-axis. We examine the two processes distinguishing, if appropriate, between administrative and disease-specific registers.

Capturing process

Whether a record is created, depends on three mechanisms. First, a health event must trigger a registration event. Second, the registration event must happen within the target population of the register. And finally, after detection, cases must be notified to the appropriate personnel within the registry. These three mechanisms are elaborated on below.

Registration event in relation to health event

We distinguish between registration event and health event (figure 1). A registration event is the intended event to be captured in the register (case definition). It has a clear link to the primary objective of the register. Examples of registration events are the death of a person, a hospitalisation, a billing action, and the assignment of a particular diagnosis, like cancer after pathological confirmation (see also table 1).

The health event is the event (disease or intervention) within a patient that triggers the registration event. The health event is frequently the actual subject of interest in epidemiological research. The relation between registration event and health event differs between disease-specific and administrative registers.

Disease-specific registers intend to capture one specific disease (condition, intervention), which means that there is a close relation between the health event and the registration event. Diseases for which registers have been set up have one or more of the following characteristics: they lead to health care activities, they have a high case fatality, they generate high costs, or they induce significant danger to the patient's family and beyond (epidemic threat). In technical terms, diseases prevail that represent acute episodes rather than chronic states, anatomical rather than physiological/functional diagnosis.
The registration process

Administrative registers capture particular actions in the health care system, like admissions, billing actions or drug prescriptions. Here, the relation between the registration event and the health event is not straightforward. Coverage with respect to the registration event itself, however, is usually complete.

Target population
A second mechanism that determines whether a registration event will be recorded in a register is the fact whether the event happens within the target population. Some registers are based on a sample of events; for instance the hospital register in the United States is based on a large stratified sample of hospitals. Other registers are defined to a geographical region or to persons insured by a particular company. Knowing the size and structure of the target population is essential to determine the appropriate denominator (population
at risk) in incidence and prevalence measures. Using registers with complete national coverage avoids problems of unrepresentative coverage and not knowing the number of persons at risk.

Notification to the registry

The third mechanism that determines capturing is the notification of cases to the registry. Disease-specific registers rely on the co-operation of individual doctors or other health care workers for notification. Cases will not be recorded if doctors are unaware of the existence of a register or if they are unwilling to participate. In some instances, doctors have a legal responsibility to notify cases, such as for some infectious diseases and death. However, this does not guarantee full coverage. In administrative registers several actions ensure that events are rarely missed. The quality of the information provided, however, is of greater concern. This will be the subject of interest in the next paragraph on the recording process.

Recording process

After discussing the three mechanisms that determine whether a record will be created or not, we will now elaborate on the recording process itself. This process determines which information will be recorded, how, and how well. The recording process yields two separate data clusters. The first cluster consists of information related to the health event that triggered the creation of a record. The second cluster involves information that is recorded additionally. We will make a few remarks about both clusters.

Event information

A critical feature of any register is the classification and codification of the health event. A classification is an integrated scheme of a defined number of mutually exclusive codes that can be assigned to each case. The purpose and setting of a register are important factors in deciding which classification to use. Administrative registers have to use broad classification schemes that cover the full range of diseases and conditions, such as the International Classification of Diseases (ICD). Disease-specific registers permit the use of more detailed classification schemes, often specifically developed for a particular disease. A point of general concern is the lack of opportunities to record information on severity or stage of the primary event, especially in broad classification schemes such as the ICD. Disease-specific classifications offer more room to record information on severity or stage, like the TNM classification in cancer registers.

Codification is the process of assigning the right code to an individual case. Codification usually requires specific and ongoing education and
The existence of a 'semantic gap' between the language used in the classification system and the language being used by physicians to describe the patient's condition can result in erroneous codes and in too many a-specific codes or missings.

**Additional information**

The amount and nature of the additional information varies widely between registers. The following general observations can be made.

First, social-demographic information like age, gender and place of residence are present in almost any register. Information on ethnicity, education and social economic status is recorded in only a few registers. Registers containing full name and address or a unique identification number are rare in the Netherlands. Or if such information is present, it is inaccessible to researchers due to current privacy regulations.

Second, the type of information that is additionally recorded has a direct link to the primary objective of a register. Consequently, a limited amount of clinical information is recorded in registers with an administrative background. An example is the limited amount of information about the existing health status of patients.

Third, feasibility of data collection is a major issue given the ongoing nature of registers. This limits the recording of information requiring additional efforts to obtain. Consequently, information on events occurring before or after the registration event are infrequently present or recorded with lower validity. For instance, follow-up information in registers is either absent or short-term (for instance vital status at discharge in hospital registers). The same applies to information on past exposures. Registers often lack this information because of the variety and complexity of measuring past exposures. Even in disease-specific registers, like cancer registers, only crude information on smoking and on occupational histories is recorded.

**In summary**

Registers are characterised by a capturing and a recording process. The capturing process determines which cases will be present in the register. The recording process determines the type of information that will be recorded for each case. Registers naturally focus on the event itself rather than on what happened before and thereafter. Disease-specific registers compared to administrative registers perform better in the recording of clinically relevant data. However, disease-specific registers are usually targeted at health events with an acute episode leading to health care contact rather than at chronic health states.
1.4 **EPIDEMIOLOGICAL RESEARCH: FUNCTIONAL CLASSIFICATION**

A variety of problems can be the subject of interest in epidemiological research. A widely used classification of epidemiological research is based on the purpose of a study, leading to categories like etiological studies, diagnostic studies, prognostic studies, intervention studies, quality of care research, and surveillance. In general, epidemiological studies intend to count or measure the occurrence of a health related phenomenon (the outcome) in a population of interest, and to study the factors (the determinants) on which the occurrence depend. This formal description of an epidemiological study, also known as the occurrence relation, was given by Miettinen. A typical example of epidemiological research would be a study investigating whether the risk of stroke (outcome of interest) in adult patients with atrial fibrillation (population of interest) varies in relation to the size of the left atrium (determinant of interest). The word 'determinant' is used in a broad and neutral way, and includes all possible factors, both causal and non-causal, on which the frequency of occurrence depend.

To discuss the opportunities and limitations associated with the use of data from registers in epidemiological research, we will divide epidemiological research into 'single moment' and 'serial' studies. Our broad classification is not intended to replace existing typologies, but merely serves as a starting point to arrive at a better understanding why some research questions are more difficult to address through registers than others. The critical distinction between single moment and serial studies is found in the timing of information on the components of the occurrence relation (population of interest, determinants and outcome). If the information on all components is present and readily measurable at a single moment in time, we refer to them as single moment studies. On the other hand, if any of the information is separated in time, we will call them serial studies. We will elaborate on the differences between single moment and serial studies by examining the characteristics of some studies belonging to each of the two classes.

**Single moment studies**

The earmark of single moment studies is that any information on the population and/or determinant of interest is present at the moment of the outcome measurement. There is no intention to obtain historical information or to collect prospective follow-up information.

The first group of single moment studies are cross-sectional studies. In cross-sectional studies, the general population or a sample thereof is examined to establish the prevalence of a condition or to determine the distribution of a
measurement. A graphical representation is given at the left of figure 2.

Examples include the estimation of the number of persons with diabetes type 2 in region X by asking about present medication use and the determination of the distribution of cholesterol levels in a stratified sample of the Dutch population aged between 18 and 60 years. The reason that these examples are single moment studies is that all necessary information, like population defining information (living in town X; aged between 18-60 years) and the outcome information (diabetes yes/no; level of cholesterol), is present at a single moment in time. These studies would remain single moment studies if the number of diabetic patients or the level of cholesterol is analysed by age, sex and socio-economic status, as the information on these types of determinants is also present at that same moment. They are no longer single moment studies, if information about past exposures or prospective follow-up information would be collected.

The second group of single moment studies determines the frequency of outcome events emerging from a dynamic population during a specified calendar time period (right side of figure 2). Again, the critical feature is that any additional information collected from the persons having the event is limited to factors present at the time of the event. A typical example is a study examining the number of cardiovascular deaths by age and sex in the Netherlands during the last decade. The additional information necessary to execute this study, such as age and sex, is present at the time of establishing the cause of death. One of the reasons why we can classify this type of research as a single moment study is because of the particular nature of dynamic populations. Examples of dynamic populations are the citizens of a specific town or all persons insured by a particular health insurance plan. Persons can enter or leave the dynamic population during the period under study. The fact that dynamic populations are defined by a qualifying state (living in town X) rather than a qualifying event (having had a myocardial infarction) means that the population defining information is present at the moment of the outcome event.
In general, the main focus of single moment studies is on measuring the frequency of a condition or the distribution of a measurement in the general population. Studying relationships between determinants and the outcome is often less important, partly because the determinants in single moment studies are limited to more or less stable factors present at the moment of outcome, such as age, sex, marital status and socio-economic class.

Serial studies
The distinctive feature of serial studies is that the different pieces of information on the components of the occurrence relation within a person are separated in time. The timing of the outcome information does not coincide with the timing of the population defining event and/or the timing of the determinants of interest.

The first group of serial studies are (prospective) cohort studies. A defined set of persons, the cohort, is followed over time on an individual basis in these studies and the outcome is established in all members of the cohort. A pictorial representation is given at the left of figure 3.
Serial studies

Cohort type studies

A1

Defined population (closed)

Person time (follow-up information)

Cases

A2

Defined population (closed)

Person time (follow-up information)

Controls

Figure 3. Pictorial representation of serial studies. Persons with \( \bullet \) and without \( \bullet \) the outcome of interest; persons with + and without – the determinant of interest.

The reason why the components of the occurrence relation are separated in time results from the definition of a cohort. A cohort is a closed set of individuals in which membership is gained through some qualifying event rather than a qualifying state as in dynamic populations. This qualifying event and the moment of outcome occurrence are separated in time. An example would be a study measuring quality of life in patients one year after their heart transplantation. This example illustrates that the moment of the qualifying event (heart transplantation) is separated in time from the occurrence of the outcome (the quality of life measurement). Schematic drawing A1 in figure 3 represents such a study. The period of separation can be very short, for example in a study determining the percentage of patients receiving aspirin in the acute phase of their myocardial infarction. Another example of cohort study is a clinical trial investigating whether ACE inhibition improves survival compared to placebo in patients with mild to moderate heart failure. Here the population qualifying event (patients with mild to moderate heart failure) and the determinant of interest (receiving ACE inhibitors or placebo) are separated in time from the outcome event (being total mortality). This type is depicted as A2 in figure 3. Again, individual follow-up is required to investigate the occurrence relation.
The second group of serial studies are case-control like studies. Case-control studies start with a group of patients having the outcome of interest (cases) and a group of persons without the condition of interest (controls). Both groups are thoroughly questioned about past exposures of interest. A pictorial representation of case-control studies involving twice as many controls as cases is given at the right side of figure 3. In these studies, there is a separation in time between the presence of the determinant (past exposure) and the timing of the outcome of interest (having a particular condition). Although the recording of information can be carried out at a single moment of time, we still refer to these studies as serial, as the occurrence relation involves a temporal relationship within an individual.

**In summary**

We distinguish between single moment studies and serial studies based on the functional difference whether the occurrence relation involves a temporal relationship within a person. In single moment studies, the population defining information, the information on potential determinants and the outcome information are all present at the same moment. The absence of a temporal direction means that single moment studies are mainly descriptive in nature. In serial studies, there is a temporal relation between the timing of the (cohort) qualifying event or the determinant status and the subsequent outcome. The presence of a temporal direction in serial studies may provide the opportunity to study the occurrence relation, under conditions, in causal terms.

### 1.5 Interaction between Data Requirements of Research and Characteristics of Register Data

Whether a research question can successfully be addressed using data from one or more registers depends on many factors. In this section we will discuss some potential areas of tension between the data requirements of research questions and the characteristics of data from registers. In this discussion we fall back on the concepts introduced in the previous paragraphs. We start with our simplified division of epidemiological research into single moment and serial studies to review in an analytic fashion the use of registers in both types of research. A more pragmatic overview of the opportunities and limitations of registers in various epidemiological research is given in section 1.6.
Single moment studies

Single moment studies and registers have corresponding purposes. Single moment studies aim to determine the distribution of health related phenomena in a general population, whereas registers aim to record consecutive registration events within a target population. Two potential problems can arise when using register data in single moment studies. The first problem is the preference of registers to record specific types of health events, the second problem relates to the sensitivity and specificity of the registration process.

Health event preference in registers

In section 1.3 we postulated some common characteristics of diseases that are captured in registers. We identified the following three characteristics:
1. diseases with major medical consequences or high public appeal;
2. diseases with a direct link to specific activities in the health care system;
3. diseases with an acute episode (event-like) are easier to define and to capture than chronic states or functional diagnosis.

We will illustrate these points by looking at two examples and determine which characteristics apply. For coronary bypass surgery all characteristics apply. Consequently, the number of coronary bypass operations can be studied in many countries using administrative and specific registers. Atrial fibrillation is at the other end of the spectrum: (1) the medical consequences of atrial fibrillation have long been underrated; (2) atrial fibrillation does not directly lead to health care contact due to a lack of symptoms or a-specific symptoms; (3) the chronic and variable course of atrial fibrillation hampers case finding. This profile of atrial fibrillation means that specific registers for atrial fibrillation are rare and that administrative registers are unfit to measure the frequency of atrial fibrillation.

Sensitivity and specificity of the registration process

The sensitivity and specificity of the registration process come into play once a potential register is available to study the condition of interest. Sensitivity is the ability of the register to capture and correctly classify all cases of interest from the researcher’s point of view, whereas specificity refers to the capability of avoiding non-cases to be counted. Sensitivity is a major issue in studies aiming to determine the frequency of conditions, as in single moment studies. Cases can be missed during the capturing process or the recording process, as illustrated in figure 4.
Interaction between single moment studies and register data

<table>
<thead>
<tr>
<th>Research</th>
<th>Using code E' to study outcome E</th>
<th></th>
<th></th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>observed = 1+3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>truth = 1+2+4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Recording process
- correct coding of E as E'
- incorrect coding of E as E''
- (incorrect) coding of non-E as E'

Capturing process
- captured
- not captured

Calendar time
- Record
- Missing
- Record
- Record

Truth
- outcome E
- non-E outcome

General population (dynamic)

Figure 4. Capturing and recording of cases in single moment studies.

Case 2 in figure 4 represents a case being missed through failure of one of the three mechanisms of the capturing process (see section 1.3). The main source of missed cases in disease-specific registers is related to the quality of the notification step. In contrast, the relation between health event and registration event is the main problem in administrative registers, as notification is usually complete with respect to the registration event. The relationship between health event and registration event in administrative registers can be complex. Studying heart failure using hospital discharge data is an illustrative example (see also chapter 4 of this thesis). First, hospital data are insufficient to determine either the incidence or prevalence of heart failure in the Netherlands, as many heart failure patients are managed outside the hospital. Trends in discharges for heart failure can still provide useful information, if the sensitivity is more or less stable during the study period.
In the case of heart failure a change in admission policy, however, could seriously affect the interpretation of time trends in heart failure admissions.

The recording process can also be responsible for 'lost' cases. Captured cases that receive a wrong code (false-negative codings) will not be counted. Case 4 in figure 4 exemplifies such a case. Incorrect coding has received much attention in the medical literature with most examples coming from the codification of causes of death.

The opposite problem also occurs; erroneously counting unrelated cases which is illustrated by case 3 in figure 4. There are two reasons for this lack of specificity: errors during codification (false-positive codings) or ambiguity of the classification scheme itself. This can lead to specific health events being recorded under several distinct codes or to health events being concealed within a larger group of related diseases, all having the same code.

**Serial studies**

Many of the problems mentioned under single moments studies are equally present in serial studies. In addition, there is the complexity of information that is separated in time. We discuss three areas of problems: health event preference of registers, the bringing together of information separated in time, and the nature of the occurrence relation: descriptive or causal.

**Health event preference of registers**

The preference of registers to capture event-like phenomena means that some populations and outcomes of interest are less likely to be recorded in registers. Similar observations that have been made under single moment studies apply here. Populations of interest are limited to those that can be defined through registers. Prime candidates, therefore, are diseases for which a specific register has been set up, diseases or procedures with high hospitalisation rates and patients receiving medications. Capturing all cases is not a prerequisite. In studying patients with acute myocardial infarction it is not necessary to examine all patients. Information on stage and severity is more significant, either to make meaningful prognostic subgroups (descriptive relations) or to achieve control of confounding in comparative studies. This subject is discussed in section 1.6.

**Pieces of information separated in time**

Serial studies mean pieces of information that are separated in time. Historical and/or prospective information has to be available to carry out serial studies. In essence, four different situations can occur, which are depicted as I to IV in figure 5. We briefly discuss these four situations.
In situation I the information that defines the population of interest and the outcome information is recorded within the same record. This means that follow-up information is recorded in addition to the information on the event that defined the population of interest. A textbook example of situation I is the recording of the vital status at discharge for each hospitalisation.

Situation II resembles the case-control design situation. Historical information on the determinants of interest is recorded in addition to the outcome event. This provides the opportunity to study a temporal relationship between past exposures and subsequent risk of developing a particular disease. For instance, the prevalence of diabetes type 2 among patients with acute myocardial infarction is compared to the prevalence of diabetes among an age-matched control group.

**Interaction between serial studies and register data**

![Diagram showing interaction between serial studies and register data](image)

**Figure 5.** Information from multiple time points in serial studies that must be connected.
In situation III the information about the population defining event and the outcome information is present in the same register, but in different records. Bringing this information together is simple if a personal identification number is present and correctly applied. Record linkage techniques can provide a solution in the absence of an identification number. An illustration of situation III can be found in chapter 7 of thesis. Patients who were hospitalised for acute myocardial infarction and discharged alive constitute the population of interest in that chapter. We examined the number and causes of cardiovascular readmissions (outcome event) within this cohort. Although the information on the population defining event (discharged alive after acute myocardial infarction) and the outcome events (cardiovascular readmissions) were present within a single register, the national register of hospital admissions, we had to use probabilistic linkage techniques to recognise readmissions due to the absence of a unique personal identifier.

The difference between situation III and IV is the fact that the population defining information and the outcome information are recorded in two distinct registers. In these situations, a unique and shared identification number is even more exceptional. Therefore, record linkage will often be necessary. A noticeable exception are Scandinavian countries, their registers contain a national health care information number.\textsuperscript{45,46}

Descriptive or causal relation
Serial studies may provide the opportunity to study the occurrence relation in causal terms. In causal studies there is a strong need to prevent or to control for the effects of differences in prognostic factors (control of confounding).\textsuperscript{47-49} There are two general, non-exclusive approaches to achieve this goal, either through study design (in particular through random assignment) or through some adjustment procedure in the analysis, like stratification or some multivariate modelling technique. Using observational data, like registers, in causal or comparative studies requires detailed information on stage and severity of the health event and on the presence of co-morbidity to achieve adequate control of confounding.

In summary
The potential problems associated with the use of registers are bigger in serial studies than in single moment studies. All in all, we can identify at least four sources of problems associated with the use of registers. First, the preference of registers to record event-like phenomena, thereby limiting the number and type of outcomes and populations that can be studied. Second, the lack of sensitivity or specificity of the registration process, leading to missed cases
and false-positive cases. Third, the difficulty of obtaining historical or prospective information in order to study temporal relations. And finally, the lack of detailed and accurate clinical information to make meaningful subgroups in descriptive studies or to adjust for differences in case mix in comparative studies. In the next section, we will examine the extent to which these problems affect a variety of practical study situations.

1.6 OPPORTUNITIES AND LIMITATIONS OF REGISTERS IN EPIDEMIOLOGICAL RESEARCH: A PRACTICAL GUIDELINE

In this final section we will provide a pragmatic perspective on the role of registers in epidemiological research using the concepts presented in the previous sections. Some of the advantages and disadvantages of using register data in different categories of epidemiological studies will be given. We will review the following categories of studies:

- population health studies
- descriptive quality of care studies
- prognostic studies
- etiological studies
- comparative quality of care studies
- efficacy studies

We will place these study categories on an axis of increasing data complexity and, therefore, increasing difficulty to address them with registers (figure 6). In addition, we will briefly discuss the potential role of registers in studies with primary data collection.

Population health studies

Population health studies describe the health of a general population by measuring the frequency or distribution of diseases and risk factors. Chapters two through five of this thesis belong to this category. In these chapters the number of deaths and hospital admissions caused by different cardiovascular conditions are analysed by age, sex, and calendar year. Additional examples can be found in the reports on the health status and forecasts of the Dutch population.50,51

The majority of the population health studies can be classified as single moment studies, as the frequency of occurrence in itself is the main object of interest in public health studies. Registers can provide valuable information for single moment studies, if the condition of interest is likely to be captured in
a register (see also section 1.5). In addition, the structure and size of the population should be known in order to interpret the numbers correctly (appropriate denominator). Given the natural link between registers and single moment studies it comes as no surprise that registers have a long tradition in studies describing population health, starting with tabulations of causes of death and the surveillance of infectious diseases.

The use of registers is, however, limited to 'catchable' events, excluding phenomena like lifestyles, risk factors and functional health states. The growing interest in composite descriptive measures, like DALY's, requires information not present in today's registers. Other approaches, like health interviews and health examinations, are needed to study health phenomena that do not have an acute episode or do not lead to health care contact. Both health interviews and health examinations can, however, suffer from selective participation.

Descriptive quality of care studies

Quality of care research addresses two questions. First, are we doing the right things to the right patient (appropriateness), and secondly, do we perform well in providing the appropriate care (skill). Examples of descriptive quality of care studies are studies determining the percentage of patients with acute myocardial infarction that receive aspirin or the percentage of CT-scanned patients in stroke. Other types of quality of care studies describe the variation in health care use or practice between individual physicians, institutions or countries. Examples include the comparison of the number of hysterectomies or bypass operations per 100,000 between countries. Descriptive figures are particular valuable if a normative figure is present or if a striking variation is found, unlikely to be explained by differences in population characteristics. The importance of differences in case mix is discussed in more detail under comparative quality of care studies.

The majority of the quality of care studies are serial studies, as typically a clinical population of interest is defined in which the outcome is measured. Administrative registers can provide valuable descriptive measures as they contain large numbers of unselected cases, reflecting day-to-day practice.

The main drawback of (administrative) registers in quality of care studies is the limited number of relevant outcomes that are recorded in these registers. The presence of information on vital status at discharge, on the number of procedures performed, and sometimes on subscribed medicines means that these outcomes are most frequently used. Extending the role of registers in quality of care studies requires additional recording of data on the process of
care, which is complex, and the recording of other follow-up information besides mortality. One limitation which could be solved relatively simple is the inability to distinguish between co-existing conditions present at admission from complications occurring during hospital stay. If a patient is coded at discharge as having an acute myocardial infarction, it is impossible to determine whether this happened some time before the admission or during the stay in hospital. This difference is vital in quality of care studies.

Prognostic studies

Prognostic studies aim to describe the course of a disease in relation to prognostic factors. An example would be a study describing 30-day mortality in patients with acute myocardial infarction in relation to age, sex, size of the infarction, presence of diabetes, etc. Results from prognostic studies are used to predict the future course in new patients.

Prognostic studies typically have a cohort design. They start with a group of persons with a defined disease, following these patients forward in time and measuring clinical outcomes. Prognostic studies in our classification are serial studies because of the time difference between the onset of the disease and the subsequent outcome. Registers provide a fitting starting point to make valuable risk predictions, as they contain information on a large number of consecutive cases.

The use of registers in prognostic studies is, however, hampered by two problems. The first problem is related to the recording of outcome information. Establishing complete follow-up information is laboursome and, hence, difficult to achieve from an organisational and financial point of view. Outcome information in registers is therefore often limited to short-term consequences and/or mortality. An example is the recording of vital status at discharge in hospital registers. For many diseases, the interest would be in longer periods of follow-up and in outcomes other than death. So, prognostic studies using registers have been carried out in stroke patients to analyse mortality, but in order to study quality of life after stroke, primary data collection was needed. A longer period of follow-up can be created by linking records from administrative or disease-specific registers with mortality data. The second problem that can be encountered is the limited amount of clinical information recorded in registers. This reduces the possibilities to study the impact of various prognostic factors.
Registers ‘ideal’

Describing population health

Counting alone in population at large

Descriptive quality of care

Long term prognosis

Comparative quality of care

Comparative or causal serial studies

Short term prognosis

Etiological studies

Comparative effectiveness

Single moment studies

Descriptive serial studies

As we move along the line, there is a growing complexity of data requirements and subsequently more problems when registers are used as sole providers of information. Data requirements and complexity increase:

- if pieces of information are separated in time
- if longer periods of follow-up are needed
- if the population of interest, the determinants and the outcome or not based on event-like concepts
- if the importance of control of confounding grows (etiologial or comparative studies)

Figure 6. Types of epidemiological studies placed along a line of increasing data complexity.

Etiological studies

Etiological studies examine the relationship between the exposure to putative risk factor(s) and the subsequent incidence of disease. Exposure can take place at a single point of time or, more often, takes place over a period of time, although some risk factors are inherited. Most etiological studies are serial studies requiring detailed information on past exposures and subsequent outcome within an individual. Furthermore, the relationship is studied in causal terms and adequate control of confounding is paramount. The case-control design is often used for efficiency reasons, as many diseases have long latency periods between exposure and first manifestations.

The role of registers in etiological studies is limited. Registers have been used in case-control like studies, in which registers supply the cases and controls. This requires that registers record information on past exposures as additional information. Measuring exposure is difficult, in particular cumulative measures, and therefore hardly recorded in registers. A notable exception is the use of registers in studying side-effects, mainly because
accurate listings of subscribed medicines exist. Pharmaco-epidemiologists now extensively use administrative registers to investigate patterns of drug related side-effects.  

Health interviews and health examinations lead to large cohorts in which the distribution of one or more risk factors is known. These cohorts could be used to study subsequent outcomes. Obtaining long-term follow-up on a large number of persons is a practical difficulty. One promising way to achieve this goal is linking these data sources with mortality or discharge data. Ecological studies have been used as an alternative to circumvent the fact that individual follow-up cannot be established. Comparing national cholesterol levels with cardiovascular mortality across countries is a typical example of such a study. However, many analytical problems exist when either the exposure or the outcome or both are not measured at an individual level.

Comparative quality of care studies

Comparative quality of care studies compare or even rank hospitals or individual physicians with respect to certain outcomes. An illustrative example is the study examining in-hospital mortality rates after carotid surgery among hospitals in the United States. The results of these studies are interpreted in a comparative way: how much of the observed variation is 'explained' by differences in patient populations (case mix) and how much by differences in hospitals, whatever these factors may be. The crude comparison of in-hospital mortality may be confounded through differences in case mix among hospitals. Some hospitals may operate upon older and sicker patients than others.

Administrative registers have regularly been used in comparative quality of care studies. The main advantage of using register data is the presence of a large number of cases reflecting daily practice. The use of registers remains controversial because of three reasons. First, out of necessity, outcomes can only be measured in terms of event-like concepts that are recorded in registers. Therefore, (in-hospital) mortality is the main outcome in present studies, despite inherent limitations due to the varying length of stay and short duration of follow-up. Second, measures of quality of care are hardly recorded, making differences in outcome between institutions or physicians hard to interpret. This is one of the reasons why surgical procedures are so often studied. Third, because results are interpreted in a comparative way, differences in case mix between institutions and physicians must be recorded in sufficient detail to adjust for differences in prognostic factors (case mix).
Several scales and indexes have been developed to measure co-morbidity. Because of the lack of information on co-morbidity in administrative registers, researchers have used historical diagnoses to gain insight into the pre-existing health status of patients. However, historical diagnoses can be a questionable source of information to measure co-morbidity (incomplete information and severity not recorded). The discussion of residual confounding after adjustment always remains in studies using observational data.

Efficacy studies

Registers have been used to compare groups treated with different medications or with different treatment strategies, like medical versus surgical treatment. The use of observational data (registers) to compare efficacy has raised fierce debates in the literature. Control of confounding, or the lack of, it is the central issue in these debates. In daily practice, clinicians treat patients in a specific way, because they think the patient needs that particular treatment, thereby mixing prognosis with treatment decisions. Correction for important prognostic factors may remove part of this bias, but subtle differences are likely to remain. This is known as confounding by (contra)indication.

The following example of the use of β-blockers after acute myocardial may illustrate this problem. Treating myocardial infarct patients with β-blockers reduces mortality, but the presence of heart failure and chronic obstructive pulmonary disease are relative contraindications. Using observational data to compare mortality in myocardial infarct patients with and without β-blockers is difficult, as patients with heart failure (and hence a shorter life-expectancy) are less likely to be treated with β-blockers. The effect of β-blockers will therefore be overestimated in the unadjusted analysis. However, even if the presence of heart failure had been recorded as a yes/no variable, a stratified analysis will not help in deciding whether treatment of β-blockers is also beneficial to patients with heart failure. As heart failure can range from mild to life threatening, it is very likely that heart failure is less severe in patients treated with β-blockers than in patients in which physicians did not subscribe β-blockers. Consequently, differences in the severity of heart failure may still account for the observed benefit of β-blockers in heart failure patients despite adjustment. Recent reports have shown, however, that treatment effects estimated from observational data sources can be made quite similar to results obtained from clinical trials. More studies of this type are needed.
Studies of unintended effects (like unexpected side-effects) are less affected by this so-called confounding by indication. In these types of studies there is a dissociation between the reason for exposure and the outcome. This opens the door to study these effects in a non-experimental setting, including registers (see also under etiological studies).

Clinical trials provide and will continue to provide the best evidence for the efficacy of therapies. But, like any other tool, clinical trials have their strengths and limitations. Logistic problems are present because of the large number of questions to be studied, in trials with rare outcomes or with outcomes that lie far in the future. In addition, the homogenising entry restrictions that give the clinical trials their statistical power, also preclude the generalisation of the results to the everyday patient with more symptoms, higher age and more co-morbidity. In several cases, registers can provide additional information or can even be a reasonable alternative.

**Additional use of registers in studies with primary data collection**

Up to this point, we have focused on research in which one or more registers were the sole providers of information. Registers, however, can also be used for specific purposes in studies collecting primary data. We briefly discuss two functions.

The first group of research comprises studies in which registers are scanned to identify study subjects (sampling frame). After identification, data is then collected from these subjects in a traditional manner using medical charts and/or patient interviews. This approach can be used in case-control studies for the identification of cases or for drawing controls. Many examples of using registers in case-control studies can be found among cancer studies. Registers can also serve as the starting point for prospective cohort studies or trials.

A second group of research consists of studies in which the population and determinants of interest are assembled in the traditional way, but registers are employed to identify outcome events during follow-up. Complete and valid outcome information on all patients is a prerequisite in order to use registers to detect outcome events. A good candidate for providing outcome information is therefore the register of causes of death. Other outcome events could be studied using hospital discharge data, if the event has a high hospitalisation rate, like acute myocardial infarction.
Concluding remarks

Registers differ markedly in their purpose, content and size, but the registration processes of registers have some points in common. These characteristics lead to some general advantages and disadvantages if registers are used in epidemiological research.

A distinctive feature of registers is their goal of recording consecutive (registration) events within a population. This provides a natural link between registers and studies trying to determine the frequency of conditions in the general population (single moment studies). The main restriction is found in the preference of registers to capture event-like phenomena rather than chronic diseases or functional states. The sensitivity of the registration process (completeness) becomes important once a potential register is available.

Researchers using registers in research other than single moment studies are faced with two problems. First, the recording of follow-up or historical information in registers is difficult, so this type of information is either not recorded, short-term or missing. All epidemiological research involving temporal relationships (serial studies) suffers from this lack of historical or prospective information. Record linkage can provide a solution by bringing information from different sources together or by reconstructing event-oriented registers into patient-oriented registers. Second, feasibility of data collection is an issue given the long-term perspective of registers. This means that register data is less detailed and precise compared with data from primary studies. This problem hampers studies needing clinical information to construct prognostic subgroups or to control for differences in case mix in comparative studies. On-site chart review to obtain additional information, to check key variables on validity, and to detect missing information is an important but time-consuming solution.

In this chapter we have discussed the role of registers in general, limiting ourselves to the main characteristics of registers and to a few types of research questions. Each register has its specific features and each research project its own requirements. Teamwork involving epidemiologists, physicians, registry personnel and statisticians is needed to avoid mistakes and to maximise the use of data from registers. Illustrations of how these potential problems turn out in specific situations can be found in chapter two through five. One disease or group of diseases takes centre stage and its impact is studied using data from one or two national registers (causes of death and hospital admissions). More information on the powerful tool for extending the role of registers, medical record linkage, can be found in chapter 6.
Chapter 7 illustrates the value of linked data sources. We used record linkage to identify the pattern of cardiovascular readmissions in a cohort of patients with acute myocardial infarction. Other recommendations on ways to improve the use of registers in epidemiological research are given in the general discussion at the end of this thesis.

REFERENCES


74. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998;36:3-5.


